

Vorbereitende Forschung für die zweite Evaluationsrunde Mindestlöhne

Erschließung neuer Datenquellen

Hans Verbeek, Michael Fertig (ISG)

Philipp vom Berge, Matthias Umkehrer, Stefan Bender (IAB)

Inhaltsverzeichnis

Tabellenverzeichnis	5
Abkürzungsverzeichnis	6
1 Hintergrund und Aufgabenstellung	10
2 Vorgehensweise	11
3 Anbieter von Daten für die Mindestlohnforschung	18
3.1 Institut für Arbeitsmarkt- und Berufsforschung/Bundesagentur für Arbeit	18
3.2 Statistische Ämter des Bundes und der Länder	20
3.3 Bundesinstitut für Berufsbildung (BIBB)	24
3.4 Berufsgenossenschaften und Sozialkassen	25
3.4.1 Berufsgenossenschaften	25
3.4.2 Sozialkassen	26
3.5 Bureau van Dijk (BvD)	28
3.6 Sonstige Datenanbieter	29
4 Daten auf Individualebene	31
4.1 Administrative Daten	31
4.1.1 Integrierte Erwerbsbiografien (IEB)	31
4.1.2 Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB)	43
4.1.3 Lohn- und Einkommenssteuerstatistik	46
4.1.4 Arbeitsunfallstatistik	48
4.1.5 Daten der Lohnausgleichskasse für das Dachdeckerhandwerk (LAK)	49
4.1.6 Datensatz der Sozialkassen des Baugewerbes (SOKA-Bau)	51
4.1.7 Daten der Sozialkasse des Gerüstbaugewerbes und der Zusatzversorgungskasse des Gerüstbaugewerbes VVaG (SOKA-Gerüst)	53
4.1.8 Auszubildendenstatistik	54
4.2 Amtliche Befragungsdaten	55
4.2.1 Verdienststrukturerhebung (VSE)	55
4.2.2 Mikrozensus	59
4.3 Nichtamtliche Daten	61
4.3.1 Beschäftigtenbefragung Bauhauptgewerbe der Forschungsgemeinschaft aus IAB, RWI und ISG	61
4.3.2 Linked Personnel Panel Beschäftigtenbefragung (LPP Beschäftigte)	63
4.3.3 Panel Arbeitsmarkt und soziale Sicherung (PASS)	64
4.3.4 BIBB/BAuA - Erwerbstätigenbefragung	67
4.3.5 Datenbank des Copenhagen Psychological Questionnaire (COPSOQ)	69
4.3.6 Sozio-Oekonomisches Panel (SOEP)	71
4.3.7 Befragung zur gewünschten und erlebten Arbeitsqualität	74
5 Daten auf Betriebs-/Unternehmensebene	75
5.1 Administrative Daten	76
5.1.1 Betriebs-Historik-Panel (BHP)	76
5.1.2 Umsatzsteuerstatistik	78
5.1.3 Unternehmensregister – System 95 (URS)	80
5.2 Amtliche Befragungsdaten	82
5.2.1 Jahrerhebung einschließlich der Investitionserhebung im Bauhaupt- sowie im Ausbaugewerbe	82

5.2.2	Pflegestatistik	83
5.3	Amtliche Firmendaten für Deutschland (AFiD)	85
5.3.1	AFiD-Panel Industriebetriebe	86
5.3.2	AFiD-Panel Dienstleistungen	89
5.4	Nichtamtliche Daten	91
5.4.1	IAB-Betriebspanel (IAB-BP)	91
5.4.2	Linked Personnel Panel Betriebsbefragung (LPP Betriebe)	94
5.4.3	BIBB-Betriebspanel zu Qualifizierung und Kompetenzentwicklung	96
5.4.4	Befragung des Instituts für Angewandte Wirtschaftsforschung e.V. (IAW) zum Mindestlohn im Elektrohandwerk	98
5.4.5	Befragung des Instituts für Angewandte Wirtschaftsforschung e.V. (IAW) zum Mindestlohn im Maler- und Lackiererhandwerk	99
5.4.6	Befragung zum Mindestlohn in der Pflegebranche durch T.I.P. Biehl & Partner	100
5.4.7	Befragung des Instituts Arbeit und Qualifikation (IAQ) zum Mindestlohn in der Branche Wäschereidienstleistungen	102
5.4.8	Dafne	103
5.4.9	Mannheimer Unternehmenspanel (MUP)	105
6	Daten auf aggregierter Ebene	107
6.1	Datenangebot der Statistik der Bundesagentur für Arbeit	107
6.2	Datenangebot der Statistischen Ämter des Bundes und der Länder	108
6.3	Elektronisches Verbandsinformations-, Recherche- und Analysesystem (ELVIRA) des Hauptverbands der Deutschen Bauindustrie e.V.	109
6.4	Tarifarchiv des Wirtschafts- und Sozialwissenschaftlichen Instituts der Hans-Böckler-Stiftung (WSI-Tarifarchiv)	111
7	Nicht verwendbare Datensätze	112
7.1	Arbeiten und Lernen im Wandel (ALWA)	112
7.2	IAB-Stellenerhebung	113
7.3	Berufliche Weiterbildung als Bestandteil lebenslangen Lernens (WeLL)	113
7.4	AFiD-Panel Industrieunternehmen	114
7.5	Einkommens- und Verbrauchsstichprobe (EVS)	114
7.6	Daten der Zusatzversorgungskasse des Steinmetz- und Steinbildhauerhandwerks VVaG	115
7.7	Datenbestand der Deutschen Rentenversicherung	115
7.8	Mikrodaten bei der Deutschen Bundesbank	116
7.9	Mindestlohndatenbank des Wirtschafts- und Sozialwissenschaftlichen Instituts der Hans-Böckler-Stiftung	116
7.10	Nationales Bildungspanel (NEPS)	117
8	Handlungsempfehlungen	117
	Literatur	125
	Anhang A.1: Leitfaden für die Expertengespräche	129

Tabelle 2.1:	Im Rahmen der Primärsuche direkt kontaktierte Institutionen	14
Tabelle 2.2:	Überblick der untersuchten Datensätze	16
noch Tabelle 2.3:	Überblick der untersuchten Datensätze	17
Tabelle 3.1:	Überblick des Datenangebots des IAB und der BA	20
Tabelle 3.2:	Standorte der FDZ der Statistischen Ämter sowie Zuständigkeit für die relevanten Mikrodaten.....	22
Tabelle 3.3:	Überblick über das Datenangebot des BIBB	25
Tabelle 3.4:	Aufteilung der Mindestlohnbranchen auf die neun BG	26
Tabelle 3.5:	Mindestlohnbranche und zuständige Sozialkassen.....	27
Tabelle 3.6:	Überblick des Datenangebots des Bureau van Dijk	28
Tabelle 3.7:	Datenangebot weiterer Datenanbieter	31
Tabelle E.1:	Systematik fehlender Werte.....	38
Tabelle E.2:	Arbeitsstunden nach Kategorien im Jahr 2010.....	41
Tabelle E.3:	Durchschnittlicher Bruttostundenverdienst im Jahr 2010	43
Tabelle 4.1:	Stichprobenzusammensetzung des SOEP Welle 2011.....	71
Tabelle 5.1:	Zeitliche Verfügbarkeit der Einzelerhebungen.....	88
Tabelle 6.1:	Zugangskosten zu ELVIRA.....	110
Tabelle 8.1:	Ergebnisindikatoren in den wichtigsten Mikrodatensätzen	121

Tabellenverzeichnis

Tabelle 2.1:	Im Rahmen der Primärsuche direkt kontaktierte Institutionen	14
Tabelle 2.2:	Überblick der untersuchten Datensätze	17
Tabelle 3.1:	Überblick des Datenangebots des IAB und der BA	20
Tabelle 3.2:	Standorte der FDZ der Statistischen Ämter sowie Zuständigkeit für die relevanten Mikrodaten	22
Tabelle 3.3:	Überblick über das Datenangebot des BIBB	25
Tabelle 3.4:	Aufteilung der Mindestlohnbranchen auf die neun BG	26
Tabelle 3.6:	Mindestlohnbranche und zuständige Sozialkassen	27
Tabelle 3.7:	Überblick des Datenangebots des Bureau van Dijk	28
Tabelle 3.7:	Datenangebot weiterer Datenanbieter	31
Tabelle E.1:	Systematik fehlender Werte	38
Tabelle E.2:	Arbeitsstunden nach Kategorien im Jahr 2010	41
Tabelle E.3:	Durchschnittlicher Bruttostundenverdienst im Jahr 2010	43
Tabelle 4.1:	Stichprobenzusammensetzung des SOEP Welle 2011	71
Tabelle 5.1:	Zeitliche Verfügbarkeit der Einzelerhebungen	88
Tabelle 6.1:	Zugangskosten zu ELVIRA	110
Tabelle 8.1:	Ergebnisindikatoren in den wichtigsten Mikrodatensätzen	121

Abkürzungsverzeichnis

ADM	Arbeitsgemeinschaft Deutscher Marktforschungsinstitute
AEntG	Arbeitnehmer-Entsendegesetz
AFiD	Amtliche Firmendaten für Deutschland
ALWA	Datensatz Arbeiten und Lernen im Wandel
BA	Bundesagentur für Arbeit
BAuA	Bundesanstalt für Arbeitsschutz und Arbeitsmedizin
BBiG	Berufsbildungsgesetz
BDSG	Bundesdatenschutzgesetz
BeH	Beschäftigten-Historik
BG	Berufsgenossenschaft
BHP	Betriebs-Historik-Panel
BIBB	Bundesinstitut für Berufsbildung
BMAS	Bundesministerium für Arbeit und Soziales
BMBF	Bundesministerium für Bildung und Forschung
BSI	Bundesamt für die Sicherheit in der Informationstechnik
BvD	Bureau van Dijk
CAPI	Computerassistierte persönliche Interviews
CATI	Computerassistierte Telefoninterviews
COPSOQ	Copenhagen Psychological Questionnaire
DAZUBI	Datensystem Auszubildende
DEÜV	Datenerfassungs- und Übermittlungsverordnung
DFG	Deutsche Forschungsgemeinschaft
DFV	Datenfernverarbeitung
DGUV	Deutsche Gesetzliche Unfallversicherung
DIW	Deutsches Institut für Wirtschaftsforschung
ELVIRA	Elektronisches Verbandsinformations-, Recherche- und Analysesystem des Hauptverbandes der Deutschen Bauindustrie e.V.
EVS	Einkommens- und Vermögensstichprobe
FDZ	Forschungsdatenzentrum
ffas	Freiburger Forschungsstelle Arbeits- und Sozialmedizin
FFP	Forschungszentrum Familienbewusste Personalpolitik
GLS	Gehalts- und Lohnstrukturerhebung
GRLC	German Record Linkage Center
GWA	Gastwissenschaftlerarbeitsplatz
HwO	Handwerksordnung
IAB	Institut für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit

IAB-ITM	IT- und Informationsmanagement des IAB
IAB-BP	IAB-Betriebspanel
IAQ	Institut Arbeit und Qualifikation
IAW	Institut für Angewandte Wirtschaftsforschung e.V.
IEB	Integrierte Erwerbsbiografien
ifo Institut	Leibniz-Institut für Wirtschaftsforschung an der Universität München e.V.
INBIL	Institut für bildungswissenschaftliche Längsschnittforschung (INBIL) an der Otto-Friedrich-Universität Bamberg
infas	Institut für angewandte Sozialwissenschaft GmbH
ISCO-88	Internationale Standardklassifikation der Berufe (International Standard Classification of Occupations) in der Fassung von 1988
ISG	Institut für Sozialforschung und Gesellschaftspolitik GmbH
LAK-Dach	Sozialkassen des Dachdeckerhandwerks
LIAB	Linked-Employer-Employee-Datensatz des IAB
LPP Beschäftigte	Linked Personnel Panel Beschäftigte
LPP Betriebe	Linked Personnel Panel Betriebe
MiDi	Mikrodatenbank Direktinvestitionen
MUP	Mannheimer Unternehmenspanel
NEPS	Nationales Bildungspanel (National Educational Panel Study)
PAPI	Persönliche Interviews (Paper and Pencil Interviews)
PASS	Panel Arbeitsmarkt und Soziale Sicherung
PUF	Public-Use-File
RatSWD	Rat für Sozial- und WirtschaftsDaten
RWI	Rheinisch-Westfälisches Institut für Wirtschaftsforschung
SGB	Sozialgesetzbuch
SIAB	Stichprobe der Integrierten Erwerbsbiografien
SOEP	Sozio-Oekonomisches Panel
SOKA-BAU	Gemeinsamer Name für die Urlaubs- und Lohnausgleichskasse der Bauwirtschaft und die Zusatzversorgungskasse des Baugewerbes AG
SOKA-Gerüst	Sozialkasse des Gerüstbaugewerbes und Zusatzversorgungskasse des Gerüstbaugewerbes VVaG
SUF	Scientific-Use-File
URS	Unternehmensregister – System 95
UVMG	Unfallversicherungsmodernisierungsgesetz
UVTöH	Unfallversicherungsträger der öffentlichen Hand
VSE	Verdienststrukturerhebung
WeLL	Panel „Berufliche Weiterbildung als Bestandteil Lebenslangen Lernens“

WSI	Wirtschafts- und Sozialwissenschaftliches Institut der Hans-Böckler-Stiftung
WZ03	Wirtschaftszweigklassifikation 2003
WZ08	Wirtschaftszweigklassifikation 2008
ZEW	Zentrum für Europäische Wirtschaftsforschung

Zusammenfassung

Dieser Methodenreport erfasst und untersucht systematisch den aktuell vorhandenen Datenbestand in Deutschland und seine Eignung für die Untersuchung von branchenspezifischen Mindestlohnregelungen nach dem Arbeitnehmer-Entsendegesetz (AEntG). Ziel ist es neue Datenquellen für die Evaluation zu erschließen, das Analysepotenzial von bereits für die Evaluation genutzten Datenquellen zu bewerten und mögliche Erweiterungs- bzw. Verbesserungsvorschläge herauszuarbeiten. Wir kommen zu dem Ergebnis, dass es zwar nicht einen allumfassenden Datensatz für die Evaluation von branchenspezifischen Mindestlöhnen in Deutschland gibt, die existierenden Datensätze insgesamt jedoch eine ausreichende Basis für eine Evaluation darstellen, sofern sie sinnvoll miteinander kombiniert werden (dürfen).

Abstract

This Methodenreport systematically compiles existing datasets in Germany and explores their suitability for evaluation of sectoral minimum wages according to the German „Arbeitnehmer-Entsendegesetz“ (posted workers act). It aims at making new data sources known, assessing the potential of data used already in previous evaluations and making suggestions with respect to enhancements and improvements. We conclude that while a comprehensive dataset suitable for evaluating German sectoral minimum wages does not yet exist, a combination of sources (if permitted) might prove satisfactory.

Keywords: Mindestlohn, Evaluation, branchenspezifisch, Deutschland.

Das Projektteam bedankt sich herzlichst bei Manfred Antoni, Konstantin Bätz, Marie-Christine Laible, Ulrike Schüller und Nadja Wehl für die Unterstützung bei diesem Forschungsprojekt.

Ferner möchten wir uns auch bei unseren Gesprächspartnerinnen und -partnern folgender Institutionen für die umfangreichen Auskünfte bedanken:

Forschungsdatenzentrum des Bundesinstituts für Berufsbildung, Deutsche Gesetzliche Unfallversicherung, Forschungsdatenzentrum der Deutschen Rentenversicherung, Forschungsdatenzentren der Statistischen Landesämter – Standort Berlin/Potsdam - im Amt für Statistik Berlin-Brandenburg, Forschungsdatenzentrum des Sozio-Oekonomischen Panels (SOEP) am Deutschen Institut für Wirtschaftsforschung (DIW), Freiburger Forschungsstelle Arbeits- und Sozialmedizin, Hauptverband der Deutschen Bauindustrie e.V., Lohnausgleichskasse für das Dachdeckerhandwerk, SOKA-BAU, Sozialkasse des Gerüstbaugewerbes/Zusatzversorgungskasse des Gerüstbaugewerbes VVaG, Zusatzversorgungskasse des Steinmetz- und Steinbildhauerhandwerks VVaG, Wirtschafts- und Sozialwissenschaftliches Institut der Hans-Böckler-Stiftung, Zentrum für Europäische Wirtschaftsforschung.

Dieses Gutachten wurde im Auftrag für das Bundesministerium für Arbeit und Soziales erstellt. Wir danken dem Bundesministerium für die freundliche Genehmigung, das Gutachten als Methodenreport zu veröffentlichen.

1 Hintergrund und Aufgabenstellung

In den Jahren 2010 und 2011 wurden im Auftrag des Bundesministeriums für Arbeit und Soziales (BMAS) die zu dem Zeitpunkt geltenden branchenspezifischen Mindestlohnregelungen nach dem Arbeitnehmer-Entsendegesetz (AEntG) evaluiert. Hauptaugenmerk der Evaluationsprojekte in den acht Mindestlohnbranchen waren dabei die Auswirkungen des Mindestlohns auf die Beschäftigung, den Arbeitnehmerschutz und zum Teil den Wettbewerb. Durch die Verteilung der Aufträge auf insgesamt vier Forschungsteams aus sieben Instituten¹ kam nicht nur ein sehr breites Spektrum an qualitativen und quantitativen Methoden zum Einsatz, sondern es wurden auch unterschiedliche Datenprodukte verwendet.

Dabei stellte sich heraus, dass die verfügbaren Datengrundlagen häufig nicht ausreichend waren, um belastbare Ergebnisse zu erhalten. Diese Erkenntnis wurde auf einem Expertenworkshop des BMAS im November 2012, an dem etwa 30 Wissenschaftler/innen aus der Arbeitsmarkt- und Sozialforschung sowie Vertreter/innen der Bundesministerien teilgenommen haben, bekräftigt (ISG 2012).

Für eine zweite Evaluationsrunde der Mindestlöhne, die für die Jahre 2014 und 2015 geplant ist, strebt das BMAS eine methodische und inhaltliche Erweiterung an (ISG 2012). Dies beinhaltet sowohl die Erfüllung der gesetzlichen Vorgaben hinsichtlich noch fehlender Ergebnisse zu den Auswirkungen branchenspezifischer Mindestlöhne auf die Beschäftigung sowie auf die Schaffung angemessener Mindestarbeitsbedingungen als auch eine thematische Erweiterung insbesondere auf die Aspekte fairer und funktionierender Wettbewerb sowie Tarifautonomie.

Im Vorfeld dieser zweiten Evaluationsrunde soll im Rahmen einer Vorstudie der aktuelle Datenbestand systematisch erfasst und auf die Verwendbarkeit für die Mindestlohnforschung untersucht werden. Die Ergebnisse dieser Vorstudie legen das Institut für Arbeitsmarkt- und Berufsforschung (IAB) und das Institut für Sozialforschung und Gesellschaftspolitik (ISG) in diesem Bericht vor.

Die Ergebnisindikatoren für die Mindestlohnevaluation sind teilweise individuelle Merkmale wie Einkommen, Qualifikation oder Arbeitszeit, zum Teil betriebsspezifische Merkmale wie Umsatz oder Investitionen und teilweise Merkmale auf der aggregierten Ebene der Branche

¹ Das Institut Arbeit und Qualifikation (IAQ) hatte den Evaluationsauftrag in den Branchen Gebäudereinigung (IAQ 2011a) und Wäschereidienstleistungen im Objektkundengeschäft (IAQ 2011b), das Institut für Angewandte Wirtschaftsforschung (IAW) in den Branchen Elektrohandwerk (IAW 2011a), Maler- und Lackiererhandwerk (IAW 2011b) und Pflege (IAW 2011c), die Forschungskooperation aus dem Zentrum für Europäische Wirtschaftsforschung (ZEW) und dem Institut für Angewandte Sozialwissenschaft (infias) in den Branchen Abfallwirtschaft (ZEW und infias 2011a) und Dachdeckerhandwerk (ZEW und infias 2011b) und der Forschungsverbund des Instituts für Arbeitsmarkt- und Berufsforschung (IAB), dem Rheinisch-Westfälischen Institut für Wirtschaftsforschung (RWI) und dem Institut für Sozialforschung und Gesellschaftspolitik (ISG) in der Branche Bauhauptgewerbe (IAB, RWI und ISG 2011).

oder der Region wie beispielsweise das Ausmaß der Schattenwirtschaft oder die Entwicklung der Steuereinnahmen.

Aus dieser Überlegung heraus bieten sich zwei Formen der Berichtsgliederung an: Eine thematische Gliederung nach den zu untersuchenden Ergebnisindikatoren sowie eine Gliederung nach Beobachtungseinheiten. Eine thematische Gliederung wird durch die Tatsache erschwert, dass die wichtigsten Datensätze mehrere Themenbereiche abdecken. Im Gegensatz dazu sind alle Datensätze eindeutig nach ihren Beobachtungseinheiten differenzierbar. Daher wurde diese Darstellungsform als die praktikablere gewählt. Um für eine thematische Mindestlohnevaluation die geeigneten Datensätze zu identifizieren, ist es notwendig, im Vorfeld zu definieren auf welchen Beobachtungseinheiten die Untersuchung durchgeführt werden soll. Für eine leichtere Identifikation der geeigneten Datensätze können die Übersichtstabellen im Excel-Anhang zu Rate gezogen werden.

Den Datensatzbeschreibungen vorangestellt sind eine kurze Beschreibung des gewählten Vorgehens in diesem Projekt (Abschnitt 2) sowie eine Kurzbeschreibung der Datenanbieter (Abschnitt 3). Auf diese beiden Abschnitte folgen die Datensatzbeschreibungen auf Individual- (Abschnitt 4), Betriebs- (Abschnitt 5) sowie auf aggregierter Ebene (Abschnitt 6).

Auf Individualebene werden die Datensätze noch nach administrativen Daten (Unterabschnitt 4.1), amtlichen Befragungen mit Auskunftspflicht (Unterabschnitt 4.2) sowie nicht-amtlichen Daten (Unterabschnitt 4.3) differenziert. Eine ähnliche Unterscheidung wird bei der betrieblichen Ebene gewählt. Allerdings wird dort nach den Unterabschnitten zu administrativen Daten (Unterabschnitt 5.1) und amtlichen Befragungsdaten (Unterabschnitt 5.2) ein zusätzlicher Unterabschnitt zu den Datenprodukten aus dem Projekt Amtliche Firmendaten für Deutschland (AFiD) eingefügt (Unterabschnitt 5.3). Diese Datenprodukte sind eine Zusammenstellung unterschiedlicher Datensätze, die zum Teil administrative Daten darstellen, teilweise als amtliche Befragungen erhoben wurden. Abschließend folgt der Abschnitt der nicht-amtlichen Daten (Unterabschnitt 5.4). Innerhalb der Unterabschnitte sind die Daten zunächst nach Datenanbietern sortiert und anschließend bei mehreren Datensätzen eines Anbieters alphabetisch.

Abschließend werden noch Datensätze kurz vorgestellt, die im Rahmen des Projekts untersucht wurden, aber für die Mindestlohnforschung als ungeeignet erachtet werden (Abschnitt 7). Die Handlungsempfehlungen fassen die Erkenntnisse aus der Bestandsaufnahme zusammen und geben Empfehlungen für die Erstellung einer geeigneten Datenbasis für die Mindestlohnforschung (Abschnitt 8).

2 Vorgehensweise

Die Erschließung neuer Datenquellen sowie die Auswertung des Analysepotenzials von bereits in der letzten Mindestlohnrunde genutzten Datenquellen wurden in vier Arbeitsschritten vorgenommen:

1. Primärsuche
2. Systematische Erschließung
3. Identifikation des Evaluationspotenzials
4. Möglichkeiten zur Verbesserung/Nutzbarmachung der Datensätze

Im Folgenden wird die Vorgehensweise in den einzelnen Arbeitspaketen genauer beschrieben.

2.1.1.1 Primärsuche nach Datensätzen

Ausgehend von den in der ersten Evaluationsrunde genutzten Daten² wurden zunächst die Datenanbieter dieser Daten kontaktiert. Anhand des Evaluationspotenzials dieser Datensätze sowie auf Hinweise der Ansprechpartnerinnen und -partner wurde Kontakt zu weiteren Datenanbietern aufgenommen. Um eine systematische und vollständige Recherche bei den einzelnen Datenanbietern gewährleisten zu können, wurde ein Leitfaden entwickelt, der sämtliche wichtige Aspekte der Datensätze erfasst. Der Leitfaden ist im Anhang A.1 dokumentiert.

Der Leitfaden orientiert sich dabei an den Grundvoraussetzungen für eine Kausalanalyse. Dies bedeutet, dass die Datensätze folgende Kriterien erfüllen mussten:

- Klar definierte Beobachtungseinheiten
- Geeignete Ergebnisvariablen
- Zeitliche Struktur des Datensatzes schließt sowohl Messungen vor als auch nach der Mindestlohneinführung ein.

Für die Einbeziehung der Datensätze waren folgende Beobachtungseinheiten von Interesse: Beschäftigte, Betriebe, Regionen oder Wirtschaftszweige. Auch Datensätze auf Unternehmensebene können verwendet werden, allerdings gilt dabei zu beachten, dass der Geltungsbereich der Mindestlöhne auf Betriebsebene definiert wird. Dies führt dazu, dass die Abgrenzung des Geltungsbereichs des Mindestlohns unscharf sein kann, wenn ein Unternehmen mehrere wirtschaftliche Schwerpunkte hat. Dieses Problem muss mit Rücksicht auf die Struktur innerhalb der Branche untersucht werden, bevor Datensätze auf Unternehmensebene verwendet werden können.

Für die Mindestlohnforschung wurden zudem folgende Merkmale als besonders relevant identifiziert:

- Auf Individualebene:
 - Einkommen als Stundenlohn oder Entgelt

² Siehe hierzu die acht Evaluationsberichte von IAB, RWI und ISG (2011), IAQ (2011a), IAQ (2011b), IAW (2011a), IAW (2011b), IAW (2011c), ZEW, infas (2011a) und ZEW, infas (2011b).

- Arbeitszeit, sowohl vertragliche als auch tatsächliche
- Qualifikation und Tätigkeitsbereiche
- Qualität von Beschäftigungsverhältnissen
- Arbeitsmotivation und Produktivität
- Arbeitsbeziehungen
- Bildungsentscheidungen junger Kohorten
- Regionale und branchenbezogene Mobilität
- Auf Betriebsebene:
 - Umsatz
 - Beschäftigungshöhe und -struktur (Qualifikation, Alter)
 - Marktein- und -austritte
 - Output und Investitionen
 - Preise
- Auf regionaler bzw. Wirtschaftszweigebene:
 - Konsumausgaben und Ersparnis
 - Schattenwirtschaft
 - Steuereinnahmen nach Steuerarten
 - Sozialversicherungsbeiträge

Insgesamt wurden auf diese Weise mit 14 Datenanbietern Gespräche über das jeweilige Datenangebot geführt, die teilweise über spezifische Datensätze, teilweise über das gesamte Datenangebot Auskunft gegeben haben. Tabelle 2.1 gibt einen Überblick zu den kontaktierten Institutionen sowie zu den behandelten Datensätzen.

Tabelle 2.1: Im Rahmen der Primärsuche direkt kontaktierte Institutionen

Institution	Datensätze
FDZ des Bundesinstituts für Berufsbildung	Datenangebot des FDZ des BIBB sowie des BIBB allgemein
Deutsche Gesetzliche Unfallversicherung	Arbeitsunfallstatistik
FDZ der Deutschen Rentenversicherung	Datenangebot des FDZ der Deutschen Rentenversicherung
FDZ der Statistischen Landesämter- Standort Berlin/Potsdam - im Amt für Statistik Berlin-Brandenburg	Datenangebot der FDZ der Statistischen Ämter des Bundes und der Länder
FDZ des Sozio-Oekonomischen Panels (SOEP) am Deutschen Institut für Wirtschaftsforschung (DIW)	Sozio-Oekonomisches Panel (SOEP)
Freiburger Forschungsstelle Arbeits- und Sozialmedizin	COPSOQ-Datenbank Projekt "Gewünschte und erlebte Arbeitsqualität"
Hauptverband der Deutschen Bauindustrie e.V.	Elektronisches Verbandsinformations-, Recherche- und Analysesystem (ELVIRA)
Lohnausgleichskasse für das Dachdeckerhandwerk	Daten der Lohnausgleichskasse des Dachdeckerhandwerks
SOKA-BAU	Daten der Sozialkassen des Baugewerbes
Sozialkasse des Gerüstbaugewerbes/ Zusatz-versorgungskasse des Gerüstbaugewerbes VVaG	Daten der Sozialkasse des Gerüstbaugewerbes und der Zusatzversorgungskasse des Gerüstbaugewerbes
Gemeinnützige Urlaubskasse für das Maler- und Lackiererhandwerk e.V./Zusatzversorgungskasse des Maler- und Lackiererhandwerks VVaG	Daten der Sozialkasse Maler- und Lackiererhandwerk
Zusatzversorgungskasse des Steinmetz- und Steinbildhauerhandwerks VVaG	Daten der Zusatzversorgungskasse des Steinmetz- und Steinbildhauerhandwerks VVaG
Wirtschafts- und Sozialwissenschaftliches Institut der Hans-Böckler-Stiftung	WSI-Tarifarchiv Mindestlohndatenbank
Zentrum für Europäische Wirtschaftsforschung	Mannheimer Unternehmenspanel

Quelle: Eigene Darstellung

Die Dauer der Gespräche war sehr unterschiedlich und einerseits abhängig vom Umfang des Datenmaterials, das für die Mindestlohnforschung genutzt werden kann, andererseits von der Bereitschaft über den Datenbestand und die Nutzungsmöglichkeiten zu reden. Im Schnitt dauerten die Gespräche etwa 30-60 Minuten und wurden mit einer Ausnahme³ telefonisch durchgeführt.

³ Aufgrund des umfangreichen Datenangebots der Statistischen Ämter wurde das Gespräch persönlich durchgeführt.

2.1.1.2 Systematische Erschließung der Datenbestände

Die im ersten Arbeitsschritt zusammengetragenen Informationen werden im nachfolgenden Arbeitsschritt systematisch aufbereitet. Dabei wurde für alle relevanten Datensätze jeweils ein Stammbblatt mit den zentralen Informationen zum Datensatz generiert. Die Stammbblätter zu den einzelnen Datensätzen befinden sich im Excel-Anhang zu diesem Bericht (http://doku.iab.de/fdz/reporte/2014/MR_03-14_Anhang.xlsx). Die Stammbblätter enthalten Informationen zum Datensatznamen sowie zum Anbieter des Datensatzes, zu seiner Verfügbarkeit für die wissenschaftliche Forschung, zur Art der Daten, zum Zeitraum, den die Daten abbilden und in welcher Frequenz, zum Anonymisierungskonzept sowie abschließend zu relevanten Merkmalen für die Verknüpfung des jeweiligen Datensatzes mit anderen Datensätzen.

Die Stammbblätter dienen als Grundlage für die Datensatzbeschreibungen (Abschnitte 4 bis 6) und als Vorlage für die Identifikation des Evaluationspotenzials sowie für die Identifikation von Möglichkeiten zur Verbesserung und Nutzbarmachung der Datensätze.

2.1.1.3 Identifikation des Evaluationspotenzials

Anhand der zusammengetragenen Informationen sowie der Datensatzbeschreibungen wird das Evaluationspotenzial jedes Datensatzes analysiert. Dabei werden die Datensätze zunächst anhand der verfügbaren Merkmale auf ihre inhaltliche Verwendbarkeit überprüft. In einem zweiten Schritt wird untersucht, ob die Zeitstruktur der Daten für eine Kausalanalyse von Mindestlöhnen geeignet ist. Weiterhin wird der Abdeckungsgrad der Daten berücksichtigt, d.h., ob sich der Datensatz beispielsweise nur auf bestimmte Branchen bezieht. Abschließend wird bei Stichproben untersucht, ob die Stichprobengröße ausreichend ist, um detaillierte Analysen zum Mindestlohn durchführen zu können. Insbesondere bei Befragungen, die alle Erwerbstätigen oder alle Betriebe als Grundgesamtheit haben, ist die Wahrscheinlichkeit groß, dass in den Mindestlohnbranchen nicht ausreichend Beobachtungen vorliegen.

2.1.1.4 Möglichkeiten zur Verbesserung/Nutzbarmachung der Datensätze

Die untersuchten Datensätze wurden zusätzlich hinsichtlich der Möglichkeiten die Struktur, die erhobenen Merkmale oder die Stichprobengröße zu verbessern untersucht. Dies umfasst sowohl die Option einer Erweiterung oder Neudefinition der Bezugsgröße, eine Anpassung der Merkmale sowie die Nutzbarmachung der Daten über Zusatzerhebungen zu verbessern.

Neben den beschriebenen Untersuchungsschritten, die ausschließlich den Datensatz selbst betreffen, wird ermittelt, inwieweit das Analysepotenzial durch eine Verknüpfung des untersuchten Datensatzes mit anderen Datensätzen gesteigert werden kann. Dabei wurde ausschließlich die technische Möglichkeit der Verknüpfung untersucht. Inwieweit datenschutzrechtliche Aspekte bzw. die Dateneigentümer den vorgeschlagenen Verknüpfungen zustimmen muss im Einzelfall geklärt werden, wenn eine Verknüpfung angestrebt wird.

Insgesamt wurden auf diese Art und Weise rund 40 Datensätze sowie die allgemeinen Datenbestände der Statistischen Ämter sowie der Statistik der Bundesagentur für Arbeit (BA) untersucht und auf ihre Verwendungsmöglichkeiten für die Mindestlohnforschung durchleuchtet. Tabelle 2.2 gibt einen Überblick über sämtliche in der Studie berücksichtigten Datensätze mit einem Hinweis auf die Nutzungsmöglichkeit für die Mindestlohnforschung sowie auf den Abschnitt der jeweiligen Datensatzbeschreibung. Die Nutzungsmöglichkeit wird dabei differenziert nach uneingeschränkt nutzbar („Ja“), mit Einschränkungen nutzbar („Bedingt“) sowie für die Mindestlohnforschung ungeeignet („Nein“). Die Einschränkungen der Nutzungsmöglichkeiten sind sehr heterogen und reichen von zu kleinen Stichproben über fehlende Merkmale, die zugespielt werden müssen, bis zu methodischen Aspekten bei der Datenerhebung. Diese Einschränkungen werden in den jeweiligen Datenkapiteln ausführlich diskutiert.

Tabelle 2.2: Überblick der untersuchten Datensätze

Datensatzname	Anbieter	Nutzbarkeit	Kapitel
AFiD Panel Industriebetriebe	FDZ der statistischen Ämter	Bedingt	5.3.1
AFiD Panel Industrieunternehmen	FDZ der statistischen Ämter	Nein	7.4
AFiD-Panel Dienstleistungen	FDZ der statistischen Ämter	Bedingt	5.3.2
Arbeiten und Lernen im Wandel	IAB	Nein	7.1
Arbeitsunfallstatistik	DGUV/BMAS	Bedingt	4.1.4
Auszubildendenstatistik	BIBB	Bedingt	4.1.8
Befragung Bauhauptgewerbe	IAB/BMAS	Ja	4.3.1
Befragung Elektrohandwerk	IAW/BMAS	Ja	5.4.4
Befragung zur gewünschten und erlebten Arbeitsqualität	FFAS/FFI/infas/BMAS	Bedingt	4.3.7
Befragung Maler und Lackierer	IAW/BMAS	Ja	5.4.5
Befragung Pflege	T.I.P. Biehl & Partner/BMAS	Ja	5.4.6
Befragung Wäscherei	IAQ/BMAS	Ja	5.4.7
Jahreserhebung einschließlich der Investitions- erhebung im Bauhaupt- sowie im Aus- baugewerbe	FDZ der statistischen Ämter	Bedingt	5.2.1
Betriebs-Historik-Panel	IAB	Ja	5.1.1
BIBB/BAuA-Erwerbstätigenbefragung	BIBB	Bedingt	4.3.2
BIBB-Betriebspanel zu Qualifizierung und Kompetenzentwicklung	BIBB	Bedingt	5.4.2
COPSOQ-Datenbank	FFAS	Bedingt	4.3.5
Dafne	Bureau van Dijk	Ja	5.4.8
Datenbestand der Deutschen Rentenversi- cherung	Deutsche Rentenversicherung	Nein	7.6
Einkommens- und Verbrauchsstichprobe	FDZ der statistischen Ämter	Nein	7.5
ELVIRA	Hauptverband der deutschen Bauindustrie	Ja	6.3

Quelle: Eigene Darstellung

noch Tabelle 2.3: Überblick der untersuchten Datensätze

Datensatzname	Anbieter	Nutzbarkeit	Kapitel
IAB-Betriebspanel	IAB	Ja	5.4.1
IAB-Stellenerhebung	IAB	Nein	7.1
Integrierte Erwerbsbiografien	Institut für Arbeitsmarkt- und Berufsforschung	Ja	4.1.1
Daten der Lohnausgleichskasse des Dachdeckerhandwerks	LAK	Ja	4.1.5
Lohn- und Einkommensteuerstatistik	FDZ der statistischen Ämter	Bedingt	4.1.3
Linked Personnel Panel Beschäftigtenbefragung	IAB	Bedingt	4.3.2
Linked Personnel Panel Betriebsbefragung	IAB	Bedingt	5.4.2
Mannheimer Unternehmenspanel	ZEW	Ja	5.4.9
Mikrodaten bei der Deutschen Bundesbank	Deutsche Bundesbank	Nein	7.8
Mikrozensus	FDZ der statistischen Ämter	Bedingt	Fehler! Verweisquelle konnte nicht gefunden werden.
Mindestlohndatenbank des WSI	WSI	Nein	7.8
Nationales Bildungspanel – Startkohorte Klasse 9	NEPS Datenzentrum Bamberg	Nein	7.10
Panel Arbeitsmarkt und soziale Sicherung	IAB	Bedingt	4.3.3
Pflegestatistik	FDZ der statistischen Ämter	Bedingt	5.2.2
Datensatz der Sozialkassen des Baugewerbes	SOKA-Bau	Ja	4.1.6
Daten der Sozialkasse des Gerüstbaugewerbes	SOKA-Gerüst	Ja	4.1.7
Sozio-Oekonomisches Panel	DIW	Bedingt	4.3.6
Statistiken der Bundesagentur für Arbeit	Bundesagentur für Arbeit	Ja	6.1
Statistiken der Statistischen Ämter	Statistische Ämter	Ja	6.2
Stichprobe Integrierter Arbeitsmarktbiografien	IAB	Bedingt	4.1.2
Umsatzsteuerstatistik	FDZ der statistischen Ämter	Ja	5.1.2
Unternehmensregister	FDZ der statistischen Ämter	Ja	5.1.3
Verdienststrukturerhebung	FDZ der statistischen Ämter	Bedingt	4.2.1
Berufliche Weiterbildung als Bestandteil Lebenslangen Lernens (WeLL)	IAB	Nein	7.3
WSI Tarifarchiv	WSI	Ja	6.4
Daten der Zusatzversorgungskasse des Steinmetz- und Steinbildhauer-	Zusatzversorgungskasse des Steinmetz- und Steinbild-	Nein	7.6

Datensatzname	Anbieter	Nutzbarkeit	Kapitel
handwerks VVaG	hauerhandwerks VVaG		

Quelle: Eigene Darstellung

3 Anbieter von Daten für die Mindestlohnforschung

3.1 Institut für Arbeitsmarkt- und Berufsforschung/Bundesagentur für Arbeit

Das Institut für Arbeitsmarkt- und Berufsforschung (IAB) verfügt über eine umfangreiche Datenbasis, die es im Rahmen seines gesetzlichen Auftrages für die Forschung nutzt. Sie umfasst Informationen aus dem Meldeverfahren zur Sozialversicherung, Prozessdaten der Bundesagentur für Arbeit (BA) sowie Befragungsdaten. Projektabhängig können auch externe Wissenschaftlerinnen und Wissenschaftler diese Daten für ihre Forschungsvorhaben verwenden. Der Zugangsweg führt dabei über das Forschungsdatenzentrum (FDZ) der Bundesagentur für Arbeit (BA) im IAB. Das FDZ bereitet aus der am IAB vorliegenden Datenbasis standardisierte Datenprodukte für wissenschaftliche Zwecke auf und stellt diese kostenlos der Forschungsgemeinschaft zur Verfügung. Es führt Aufbereitungen, Aktualisierungen und Prüfungen der Daten durch und dokumentiert diese unter Berücksichtigung rechtlicher Aspekte der Datensicherheit. Es berät außerdem über Datenzugang, Datenhandling und Analysemöglichkeiten sowie zur Reichweite und Gültigkeit der Daten.

Aus dem Angebot der umfangreichen Mikrodatensätze des FDZ werden im Rahmen dieses Berichts das IAB-Betriebspanel (IABB), die Stichprobe der integrierten Arbeitsmarktbiografien (SIAB) sowie das Betriebs-Historik-Panel (BHP) vorgestellt, da diese für die Mindestlohnforschung von besonderer Relevanz sind. Der Linked-Employer-Employee-Datensatz des IAB (LIAB) wird nicht separat besprochen, da er eine Kombination der genannten Datenprodukte darstellt. Informationen über weitere Datenprodukte, z.B. den Datensatz „Arbeiten und Lernen im Wandel“ (ALWA), das Panel Arbeitsmarkt und soziale Sicherung (PASS) sowie das Panel „Berufliche Weiterbildung als Bestandteil Lebenslangen Lernens“ (WeLL), sind auf der Internetpräsenz des FDZ unter <http://fdz.iab.de/> zu finden.

Bei den im FDZ angebotenen Daten handelt es sich fast ausschließlich um Sozialdaten und diesen gleichgestellte betriebsbezogene Daten, zu denen ein Zugang nur unter besonderen datenschutzrechtlichen Auflagen möglich ist. Das FDZ bietet daher lediglich Stichproben aus der Grundgesamtheit an und verwendet außerdem weitere Methoden der Anonymisierung in mehreren Stufen, von denen zwei für die Mindestlohnforschung relevant sind.

Bei schwach anonymisierten Daten werden Identifikatoren wie Namen, Adresse, Sozialversicherungsnummer oder Betriebsnummer gelöscht und die Ausprägungen einiger besonders sensibler Merkmale wie z.B. die Staatsangehörigkeit zusammengefasst. Um das Deanonymisierungsrisiko weiter zu senken, werden außerdem standardmäßig Wirtschaftszweige und Regionalangaben nur vergrößert weitergegeben.

Bei Scientific Use Files (SUF) handelt es sich um faktisch anonymisierte Mikrodaten, deren Informationsgehalt soweit reduziert wurde, dass eine Deanonymisierung nur mit unverhältnismäßig großem Aufwand an Zeit, Kosten und Arbeitskraft möglich ist. Bei der Erstellung der SUFs orientiert sich das FDZ dabei an den Empfehlungen von Müller et al. (1991).

Es bestehen drei verschiedene Datenzugangswege, die sich im Hinblick auf den Anonymisierungsgrad der bereitgestellten Daten und die Modalitäten der Datennutzung voneinander unterscheiden. Schwach anonymisierte Daten können über Gastaufenthalte oder per Datenfernverarbeitung genutzt werden. Während im ersten Fall der Nutzer an einem der FDZ Standorte interaktiv an seinen Programmen arbeiten kann, wird im zweiten Fall lediglich übermittelter Programmcode am FDZ gestartet. In beiden Fällen werden die Analyseergebnisse den Forschern und Forscherinnen erst nach einer Datenschutzprüfung durch die Mitarbeiter des FDZ zugesandt. SUFs werden den Nutzern über eine Datenaustauschplattform zum Download zur Verfügung gestellt und können entsprechend eines vereinbarten Datensicherheitskonzepts gespeichert und ausgewertet werden. Alle Datenzugangswege am FDZ haben gemeinsam, dass sie nur eine projektspezifische Nutzung entsprechend des Datennutzungsvertrags zulassen. Bei Gastaufenthalten bedarf es außerdem der Genehmigung durch das Bundesministerium für Arbeit und Soziales (BMAS), das insbesondere die Einhaltung der Bedingungen des § 75 SGB X prüft.

Reichen die vom FDZ zur Verfügung gestellten Standardprodukte nicht aus, können auf Antrag nach § 75 SGB X in der BA (Statistik der BA) oder am IAB Datenauszüge jeweils in enger Absprache mit dem Forschungsprojekt gemäß dessen spezifischem Bedarf aufbereitet werden. Für die Mindestlohnforschung könnte insbesondere eine Ziehung aus den Integrierten Erwerbsbiografien (IEB) des IAB relevant sein. Diese Datenlieferungen werden jedoch nicht am FDZ erstellt. Hierfür ist der Geschäftsbereich IT- und Informationsmanagement (ITM) zuständig. Aufgrund der nicht-standardisierten Natur des resultierenden Datenprodukts muss mit längeren Bearbeitungszeiten als bei den Standardprodukten des FDZ gerechnet werden, insbesondere da zunächst die Durchführbarkeit der Anfrage und die Einhaltung des Datenschutzes geprüft werden müssen. Eine Genehmigung durch das BMAS ist auch hier nötig. Darüber hinaus ist die Datenbereitstellung durch ITM kostenpflichtig. Es bestehen Fallpauschalen, die allerdings bei erheblichem Aufwand durch zusätzliche aufwandsorientierte Mehrkosten ergänzt werden können. Insgesamt liegen die Gebühren für eine Datenlieferung derzeit bei mindestens 10.000 Euro. Bei Forschungsvorhaben im Auftrag von Bundesministerien bestehen aber Rabattmöglichkeiten.

Aggregierte Statistiken für den Arbeitsmarkt können direkt über den Internetauftritt der BA bezogen werden. Hier werden vor allem quellenspezifische Querschnittsbetrachtungen angeboten. Für die Darstellung von Makroentwicklungen ist die Beschäftigungsstatistik von besonderer Bedeutung. Für weitere Datenanfragen steht das Datenzentrum der Statistik der BA zur Verfügung. Detailliertere Datenziehungen sind allerdings ebenfalls in Abhängigkeit vom Aufwand kostenpflichtig.

Tabelle 3.1 gibt einen Überblick über das für die Mindestlohnforschung relevante Datenangebot des Instituts für Arbeitsmarkt- und Berufsforschung sowie der Bundesagentur für Arbeit.

Tabelle 3.1: Überblick des Datenangebots des IAB und der BA

		Kapitel
Datenanbieter	Forschungsdatenzentrum des Instituts für Arbeitsmarkt- und Berufsforschung	
Homepage	http://fdz.iab.de/	
Datenangebot für die Mindestlohnforschung	Stichprobe der integrierten Arbeitsmarktbiografien (SIAB)	4.1.2
	Betriebs-Historik-Panel (BHP)	5.1.1
	Linked-Employer-Employee-Datensatz des IAB (LIAB)	4.1.2& 5.1.1
	IAB-Betriebspanel (IAB-BP)	5.4.1
	Linked Personnel Panel Beschäftigtenbefragung (LPP Beschäftigte)	4.3.2
	Linked Personnel Panel Betriebsbefragung (LPP Betriebe)	5.4.2
Datenanbieter	Institut für Arbeitsmarkt- und Berufsforschung - Geschäftsbereich IT- und Informationsmanagement (ITM)	
Homepage	http://www.iab.de/124/section.aspx/Bereichsnummer/20	
Datenangebot für die Mindestlohnforschung	Integrierte Erwerbsbiografien (IEB)	4.1.1
Datenanbieter	Statistikservice der Bundesagentur für Arbeit	
Homepage	http://statistik.arbeitsagentur.de/	
Datenangebot für die Mindestlohnforschung	Beschäftigtenstatistik der Bundesagentur für Arbeit	6.1

Quelle: Eigene Darstellung

3.2 Statistische Ämter des Bundes und der Länder

Das Statistische Bundesamt ist die zentrale Behörde für die Erhebung, Auswertung und Veröffentlichung amtlicher statistischer Informationen. Dabei kooperieren sie insbesondere bei der Datenerhebung sehr eng mit den 14 Statistischen Landesämtern.⁴ Das gemeinsame Datenangebot von Bund und Ländern wird sowohl im Rahmen von regelmäßigen Veröffentlichungen zu verschiedenen Themen aufbereitet. Zudem gibt es über Online-Datenbanken einen direkten Zugriff auf eine Vielzahl aggregierter statistischer Auswertungen. Das Datenangebot dieser Onlinedatenbanken ist kostenlos, allerdings kann über kostenpflichtige Zugänge auf ein erweitertes Angebot sowie unterstützende Funktionen zurückgegriffen werden.

⁴ Grundsätzlich gibt es in jedem Bundesland ein Statistisches Landesamt. Berlin und Brandenburg sowie Hamburg und Schleswig-Holstein haben jeweils ein gemeinsames Statistisches Landesamt.

Die Möglichkeiten und Grenzen der Onlinedatenbanken werden im Abschnitt 6.2 ausführlicher diskutiert.

Um eine Nutzung der amtlichen Daten auf Mikroebene für Forschungszwecke zu ermöglichen, wurde 2001 das Forschungsdatenzentrum des Statistischen Bundesamtes gegründet, gefolgt 2002 von der Gründung eines gemeinsamen Forschungsdatenzentrums der Statistischen Landesämter durch die Statistischen Landesämter. Das gemeinsame Ziel der beiden FDZ ist es die Nutzungsmöglichkeiten und den Zugang zu Mikrodaten insbesondere durch die Einrichtung unterschiedlicher Zugangswege zu verbessern. Dabei werden folgende drei Aspekte als zentral für die Zielerreichung gesehen:

- **Fachlich zentralisierte Datenhaltung:** Der überwiegende Teil der amtlichen Statistik wird dezentral in den Statistischen Landesämtern erhoben, aufbereitet und gespeichert. Um den Zugangsweg zu den Daten für Forschungszwecke zu erleichtern wurde eine fachlich zentralisierte Datenhaltung eingerichtet, die eine Bereitstellung der Mikrodaten für das gesamte Bundesgebiet länderübergreifend ermöglicht. Dabei werden die einzelnen Mikrodaten den regionalen Standorten thematisch zugeteilt, die dann sowohl für die Aufbereitung zu Forschungszwecken als auch für die Beratung der Datennutzer dieser Mikrodaten verantwortlich sind.
- **Regionalisierte Infrastruktur:** Zur Nutzung faktisch anonymisierter Mikrodaten werden an den einzelnen Standorten des FDZ des Statistischen Bundesamtes sowie des FDZ der Statistischen Landesämter Gastwissenschaftlerarbeitsplätze zur Verfügung gestellt.
- **Metadaten-system:** Um die Mikrodaten zielgerichtet nutzen sowie Ergebnisse interpretieren und analysieren zu können, sind umfangreiche Hintergrundinformationen zu den Daten notwendig. Diese werden im Rahmen eines Metadaten-systems zur Verfügung gestellt.

Das Datenangebot an den FDZ der Statistischen Ämter des Bundes und der Länder, im Folgenden als FDZ der Statistischen Ämter bezeichnet, umfasst über 100 unterschiedliche Datensätze aus folgenden Statistikbereichen:

- Sozialstatistiken
- Wirtschaftsstatistiken
- Finanz- und Steuerstatistiken
- Rechtspflegestatistiken
- Agrar-, Energie- und Umweltstatistiken

In diesem Bericht werden die für den Mindestlohn relevanten Mikrodaten vorgestellt. Dies gilt sowohl für Datenquellen, die auch nach einer gründlichen Überprüfung für die Mindestlohnforschung als geeignet identifiziert wurden, als auch für Datensätze, die sich nach dieser Überprüfung als ungeeignet herausgestellt haben. Die in diesem Bericht vorgestellten Da-

tensätze werden in Tabelle 3.2 den für die zentrale Datenhaltung verantwortlichen Standorten zugeordnet.

Tabelle 3.2: Standorte der FDZ der Statistischen Ämter sowie Zuständigkeit für die relevanten Mikrodaten

Standort der FDZ	Datensätze	Kapitel
Standorte des FDZ des Statistischen Bundesamtes		
Berlin Bonn Wiesbaden	Einkommens- und Verbrauchsstichprobe	7.5
	Lohn- und Einkommenssteuerstatistik	4.1.3
Standorte des FDZ der Statistischen Landesämter		
Bad Ems		
Berlin/Potsdam	AFiD-Panel Dienstleistungen	5.3.2
Bremen		
Düsseldorf	Mikrozensus	Fehler! Verweisquelle konnte nicht gefunden werden.
Erfurt		
Halle	AFiD-Panel Industriebetriebe	5.3.1
	AFiD-Panel Industrieunternehmen	7.4
Hannover	Unternehmensregister	5.1.3
Kamenz/Dresden	Pflegestatistik	5.2.2
Hamburg/Kiel		
München		
Saarbrücken		
Schwerin		
Stuttgart	Jahreserhebung einschließlich der Investitionserhebung im Bauhauptgewerbe	5.2.1
	Jahreserhebung einschließlich der Investitionserhebung im Ausbaugewerbe	
Wiesbaden	Umsatzsteuerstatistik	5.1.2
	Verdienststrukturerhebung	4.2.1

Quelle: Eigene Darstellung

Insgesamt bieten die FDZ der Statistischen Ämter vier unterschiedliche Zugangswege zu den Daten an, die sich in ihrer Verfügbarkeit, den Kosten und der Anonymisierung unterscheiden.

Der einfachste Datenzugriff erfolgt über Public Use Files (PUF). Diese sind sogar teilweise kostenlos als CAMPUS-Files von den Internetseiten der FDZ der Statistischen Ämter downloadbar. Andere PUF, die nicht als CAMPUS-Files zur Verfügung stehen, haben unterschiedlich hohe Entgelte. Die zur Verfügung gestellten Daten sind absolut anonymisiert, d.h. eine

Identifikation einzelner Beobachtungen ist technisch nicht möglich. In der Regel wird dies über eine Stichprobe des eigentlichen Datensatzes, eine Auswahl der Merkmale des eigentlichen Datensatzes sowie einer Vergrößerung bestimmter Merkmale erreicht.

Scientific-Use-Files (SUF) sind auch eine Off-Site-Nutzung der Daten, d.h. die Daten verlassen die Räumlichkeiten der FDZ der Statistischen Ämter und können am eigenen Arbeitsplatz ausgewertet werden. Im Unterschied zu den PUF ist hier keine absolute, sondern lediglich eine faktische Anonymisierung vorgeschrieben. Dies bedeutet, dass eine Deanonymisierung zwar theoretisch noch möglich wäre, aber lediglich unter einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitsaufwand. Im Unterschied zu den PUF ist der Nutzerkreis auf Hochschulen und unabhängige wissenschaftliche Forschungseinrichtungen eingeschränkt.

Neben den bereits beschriebenen Off-Site-Nutzungswegen stehen zwei unterschiedliche On-Site-Nutzungswege zur Verfügung. Im Unterschied zu den bisher beschriebenen Nutzungswegen verbleiben die Daten auf den Rechnern der FDZ der Statistischen Ämter. Ähnlich wie bei den SUF werden beim Zugang über einen Gastwissenschaftlerarbeitsplatz an den Standorten der FDZ der statistischen Ämter die Daten faktisch anonymisiert zur Verfügung gestellt. Auch dieser Zugangsweg ist ausschließlich Angehörigen von Hochschulen und unabhängigen wissenschaftlichen Forschungseinrichtungen möglich. Im Unterschied zu den SUF wird die faktische Anonymisierung nicht ausschließlich über die Anonymisierung der Daten erreicht sondern auch über die Regulierung des Datenzugangs. Zudem können neben den Standardangeboten auch projektspezifisch aufbereitete Daten zur Verfügung gestellt werden.

Über die Kontrollierte Datenfernverarbeitung kann auf formal anonymisierte Daten zurückgegriffen werden, d.h. dass Auswertungen und Analysen mit den Originaldaten vorgenommen werden. Allerdings existiert kein direkter Zugriff der Nutzer/innen auf die Daten, sondern die Auswertungsprogramme werden an die FDZ der Statistischen Ämter geschickt, dort ausgeführt und die Ergebnisse der Auswertung einer Geheimhaltungsprüfung unterzogen.

Mit Ausnahme der CAMPUS-Files muss für die Nutzung der Daten ein Datennutzungsantrag mit Angaben zu Projekt, Institution und den vorgesehenen Datennutzerinnen und -nutzern gestellt werden. Wenn projektspezifisch aufbereitete Daten beantragt werden, muss der Antrag durch eine Liste der benötigten Merkmale ergänzt werden. Da die Standardprodukte bereits aufbereitet inklusive abgestimmten Anonymisierungskonzepten zur Verfügung stehen, ist bei einer Verwendung der Standardprodukte mit einer kurzen Beantragungsdauer zu rechnen.

Projektspezifisch aufbereitete Daten können ausschließlich über die On-Site-Nutzungswege beantragt werden. Im Gegensatz zu den Standardprodukten existieren hierfür noch keine Anonymisierungskonzepte, die individuell anhand des Datenbedarfs der Nutzer/innen erstellt werden. Die Dauer der Beantragung ist abhängig von den beantragten Variablen. Bei sehr detaillierten Angaben ist die Erstellung und Überprüfung des Anonymisierungskonzepts sehr

aufwändig. Das erstellte Anonymisierungskonzept muss anschließend von den Dateneigentümern – in der Regel die 14 Statistischen Landesämter – genehmigt werden, bevor die Daten in der gewünschten Form zur Verfügung gestellt werden können.

Grundsätzlich gilt für die Zugangswege SUF sowie die On-Site-Nutzung, dass pro beantragte Statistik, pro Erhebungsjahr und pro Nutzungsweg 250 Euro berechnet werden. Die CAMPUS-Files sind grundsätzlich kostenlos, die PUF haben unterschiedliche Entgelthöhen. Um Daten der FDZ der Statistischen Ämter nutzen zu können, müssen sämtliche Nutzer/innen nach dem Bundesdatenschutzgesetz verpflichtet werden.

Wie bereits erwähnt, wird über eine regionalisierte Infrastruktur möglichst flächendeckend ein möglichst einfacher Datenzugang über alle Zugangswege ermöglicht. Dies wurde durch die Einrichtung von Gastwissenschaftlerarbeitsplätzen an den Standorten der FDZ der Statistischen Ämter erreicht. Tabelle 3.2 gibt einen Überblick über die einzelnen Standorte sowie deren Datenzuständigkeit, die für die Evaluation des Mindestlohns relevant sind.

3.3 Bundesinstitut für Berufsbildung (BIBB)

Das Bundesinstitut für Berufsbildung (BIBB) ist 1970 als rechtsfähige Anstalt des öffentlichen Rechts gegründet worden und wird als solche aus Haushaltsmitteln des Bundes finanziert. Es steht unter der Rechtsaufsicht des Bundesministeriums für Bildung und Forschung (BMBF). Die Aufgaben des Instituts sind im Berufsbildungsgesetz (BBiG) beschrieben.

Die Forschung des BIBB ist in fünf Themenschwerpunkte aufgeteilt:

- Ausbildungsmarkt und Beschäftigungssystem
- Modernisierung und Qualitätsentwicklung der beruflichen Bildung
- Lebensbegleitendes Lernen, Durchlässigkeit und Gleichwertigkeit der Bildungswege
- Berufliche Bildung für spezifische Zielgruppen
- Internationalität der beruflichen Bildung

Eine der Aufgaben des BIBB ist die Aufbereitung der Berufsbildungsstatistik der Statistischen Ämter des Bundes und der Länder sowie die Auswertung dieser Daten für den jährlichen Berufsbildungsbericht. Die aufbereiteten Daten werden in aggregierter Form in die Datenbank Datensystem Auszubildende (DAZUBI) auf den Seiten des BIBB gespeist.⁵ Die Auszubildendenstatistik, die die Grundlage für DAZUBI ist, wird im Unterabschnitt 4.1.8 näher vorgestellt.

Neben der Aufbereitung und Auswertung der Berufsbildungsstatistik führt das BIBB zu bestimmten Themenbereichen in seinen Themengebieten eigene Befragungen durch. Die Datensätze werden im FDZ des BIBB interessierten Nutzern und Nutzerinnen für die nicht-

⁵ Internetseite: www.bibb.de/dazubi [abgerufen am: 11.11.2013].

kommerzielle Forschung zur Verfügung gestellt. Ähnlich wie beim FDZ der BA am IAB sowie den FDZ der Statistischen Ämter bietet das FDZ des BIBB abhängig vom Datensatz die vier Zugangswege PUF, SUF, Gastwissenschaftlerarbeitsplatz und Datenfernverarbeitung an.

Die Bereitstellung der Daten erfolgt auf Basis eines Antrags, in dem die Namen der Nutzer/innen die beantragende Institution sowie eine Projektskizze enthalten sind. Die Daten werden innerhalb weniger Tage grundsätzlich kostenlos zur Verfügung gestellt. Bei einer Beantragung der Daten als SUF ist es möglich, dass für die Datenbereitstellung von GESIS - Leibniz-Institut für Sozialwissenschaften⁶ eine Bereitstellungsgebühr von 25 bis 50 Euro erhoben wird. Tabelle 3.3 gibt einen Überblick über den relevanten Datenbestand am BIBB und am FDZ des BIBB.

Tabelle 3.3: Überblick über das Datenangebot des BIBB

		Kapitel
Datenanbieter	Bundesinstitut für Berufsbildung	
Homepage	http://www.bibb.de/de/index.htm	
Datenangebot für die Mindestlohnforschung	Auszubildendenstatistik (Grundlage des Datensystems Auszubildende - DAZUBI)	4.1.8
Datenanbieter	FDZ des Bundesinstituts für Berufsbildung	
Homepage	http://www.bibb.de/de/50113.htm	
Datenangebot für die Mindestlohnforschung	BIBB/BAuA- Erwerbstätigenbefragung	4.3.2
	BIBB-Qualifizierungspanel	5.4.2

Quelle: Eigene Darstellung

3.4 Berufsgenossenschaften und Sozialkassen

Da eine Mitgliedschaft sowohl bei den Sozialkassen als auch bei den Berufsgenossenschaften für Unternehmen und Betriebe verpflichtend ist, bieten Datensätze, die bei diesen Institutionen erstellt werden eine gute Ausgangslage, da es sich in der Regel um Vollerhebungen handelt. Im Folgenden werden die Funktionsweisen der Berufsgenossenschaften (Unterabschnitt 3.4.1) sowie der Sozialkassen (Unterabschnitt 3.4.2) beschrieben.

3.4.1 Berufsgenossenschaften

Die Berufsgenossenschaften (BG) sind Anstalten des öffentlichen Rechts, die für die gesetzliche Unfallversicherung verantwortlich sind. Eine Mitgliedschaft in der BG ist für Unternehmen verpflichtend.

Insgesamt gibt es neun Berufsgenossenschaften, die unter dem Dach der Deutschen Gesetzlichen Unfallversicherung e.V. (DGUV) organisiert sind. Die BG sind einzelnen Wirtschaftsbereichen zugeteilt und sind für die Unfallversicherung von Unternehmen und Betrie-

⁶ Die Bereitstellung der SUF wird in der Regel durch das Datenarchiv für Sozialdaten des GESIS - Leibniz-Institut für Sozialwissenschaften vorgenommen.

ben in diesem Sektor verantwortlich. Das System ist umlagefinanziert, d.h. dass die BG keinen Gewinn erwirtschaften dürfen und dementsprechend die Einnahmen ihre Ausgaben decken müssen.

Im Rahmen ihrer Aufgaben sammeln die Berufsgenossenschaften Daten zu Arbeits- und Arbeitsweegeunfällen, arbeitsbedingten Gesundheitsgefahren und Berufskrankheiten. Diese werden im Abschnitt 4.1.4 näher beschrieben. Tabelle 3.4 gibt einen Überblick über die neun BG und für welche Mindestlohnbranchen diese zuständig sind.

Tabelle 3.4: Aufteilung der Mindestlohnbranchen auf die neun BG

Berufsgenossenschaft		Mindestlohnbranchen
Berufsgenossenschaft der Bauwirtschaft	BG BAU	Baugewerbe Dachdeckerhandwerk Gebäudereinigung Gerüstbauerhandwerk Maler- und Lackiererhandwerk Steinmetz- und Steinbildhauerhandwerk
Berufsgenossenschaft Energie Textil Elektro Medienerzeugnisse	BG ETEM	Elektrohandwerk
Berufsgenossenschaft Rohstoffe und Chemische Industrie	BG RCI	
Berufsgenossenschaft für Transport und Verkehrswesen	BG Verkehr	Abfallwirtschaft
Berufsgenossenschaft Holz und Metall	BGHM	
Berufsgenossenschaft Handel und Warendistribution	BGHW	
Berufsgenossenschaft Nahrungsmittel und Gastgewerbe	BGN	
Berufsgenossenschaft für Gesundheitsdienst und Wohlfahrtspflege	BGW	Pflegebranche
Verwaltungs-Berufsgenossenschaft	VBG	Aus- und Weiterbildung Sicherheitsdienstleistungen

Quelle: Eigene Darstellung

3.4.2 Sozialkassen

In einigen Wirtschaftsbereichen existieren allgemeinverbindliche Tarifverträge, die eine Teilnahme an einem Sozialkassenverfahren vorschreiben. Da die Sozialkassen für die Einhaltung und Umsetzung bestimmter tarifvertraglicher Regelungen verantwortlich sind, liegt die Trägerschaft der Sozialkassen immer bei den beiden Tarifparteien. Tabelle 3.5 gibt einen Überblick, in welchen Mindestlohnbranchen ein Sozialkassenverfahren existiert.

Die Übersicht zeigt, dass insbesondere die Mindestlohnbranchen im Bauhaupt- und Ausbaugewerbe ein Sozialkassenverfahren implementiert haben. Dies ist nicht weiter überraschend, da die Sozialkassen unter anderem dafür zuständig sind, saisonale Schwankungen auszugleichen.

Insgesamt sind die Aufgabenbereiche der Sozialkassen sehr unterschiedlich und reichen von betrieblichen Rentenbeihilfen, der Sicherung von Urlaubsansprüchen, über die Finanzierung der Berufsausbildung bis hin zur Gewährung eines Teils des 13. Monatseinkommens bzw. eines Ausfallgeldes beim witterungsbedingten Arbeitsausfall. In der Regel richten sich die Beiträge nach den Bruttolohnsummen der Betriebe.

Zur Wahrnehmung ihrer Aufgaben erheben die Sozialkassen Daten, die in der Regel die tatsächliche Arbeitszeit und die Bruttolöhne enthalten. Meist erfolgt die Erhebung auf Individualebene, so dass ein Datensatz mit detaillierten Arbeitszeit- und Lohninformationen für einen langen Zeitraum vorliegen. Diese Informationen sind insbesondere bei einer Verknüpfung mit anderen Datenquellen eine gute Methode valide und verlässliche Stundenlohninformationen zu erhalten. Diese Option wurde in der vergangenen Evaluationsrunde bereits für die Mindestlohnevaluation im Bauhauptgewerbe (IAB, ISG, RWI 2011) und im Dachdeckerhandwerk (ZEW, infas 2011b) genutzt.

Tabelle 3.5: Mindestlohnbranche und zuständige Sozialkassen

Mindestlohnbranche	Sozialkasse	Gewerkschaft	Arbeitgeberverband	Kapitel
Baugewerbe	SOKA BAU	Industriegewerkschaft Bauen-Agrar-Umwelt	Zentralverband des Deutschen Baugewerbes e.V. Hauptverband der Deutschen Bauindustrie e.V.	4.1.6
Dachdeckerhandwerk	Lohnausgleichskasse für das Dachdeckerhandwerk	Industriegewerkschaft Bauen-Agrar-Umwelt	Zentralverband des Deutschen Dachdeckerhandwerks	4.1.5
Gerüstbauerhandwerk	Sozialkasse des Gerüstbauergewerbes	Industriegewerkschaft Bauen-Agrar-Umwelt	Bundesverband Gerüstbau	4.1.7
Maler- und Lackiererhandwerk	Gemeinnützige Urlaubs- und Zusatzversorgungskasse für das Maler- und Lackiererhandwerk	Industriegewerkschaft Bauen-Agrar-Umwelt	Bundesverband Farbe und Gestaltung Landesverbände des Deutschen Maler- und Lackiererhandwerks	
Steinmetz- und Steinbildhauerhandwerk	Zusatzversorgungskasse des Steinmetz- und Steinbildhauerhandwerks VVaG	Industriegewerkschaft Bauen-Agrar-Umwelt	Bundesverband Deutscher Steinmetze	7.6

Quelle: Eigene Darstellung auf Basis BMAS (2013a)

Bisher ist der Datenbestand der Sozialkassen für Forschungszwecke kaum genutzt worden. Dementsprechend gibt es bei keiner Sozialkasse ein formalisiertes Verfahren der Datenbeantragung. Welche Daten zur Verfügung stehen und in welcher Form, ist dementsprechend stark abhängig von der Kooperationsbereitschaft der Sozialkasse. Diese ist nicht bei allen

Sozialkassen gegeben.⁷ Die potenziell verfügbaren und für die Mindestlohnforschung nutzbaren Datensätze der Sozialkassen werden in den Unterabschnitten 4.1.5 (Dachdecker), 4.1.6 (Bauhauptgewerbe) und 4.1.7 (Gerüstbauer) beschrieben. Die Daten der Zusatzversorgungskasse des Steinmetz- und Steinbildhauerhandwerks VVaG werden im Unterabschnitt 7.6 beschrieben, da sie für die Mindestlohnforschung nicht nutzbar sind.

3.5 Bureau van Dijk (BvD)

Das 1971 gegründete Bureau van Dijk (BvD) ist ein kommerzieller Anbieter von Firmen- und Finanzdaten, die weltweit angeboten werden. Die Kernkompetenzen des BvD liegen dabei in der Zusammenführung und Aufbereitung verschiedener Firmendaten unterschiedlicher Anbieter. Diese werden überwiegend von privaten Unternehmen genutzt, jedoch werden die Daten des BvD auch zunehmend für die Forschung erschlossen.

Derzeitig werden Informationen auf nationaler und internationaler Ebene in über 20 Datenbanken angeboten: Beispielsweise umfasst Orbis ca. 100 Millionen Unternehmen weltweit; Amadeus konzentriert sich auf europäische Firmen, Bankscope enthält detaillierte Informationen über Banken weltweit und die Datenbank Dafne führt Daten über Unternehmen in Deutschland. Für die Evaluationsforschung innerhalb Deutschlands ist dabei nur Dafne relevant, da sie das gleiche Merkmalspektrum wie Orbis aufweist. Die Möglichkeiten internationaler Vergleiche sollen hier nicht erörtert werden.

Der Datenzugang erfolgt generell über Jahreslizenzen, die die Recherche auf einer Softwareoberfläche ermöglichen. Dabei liegt der Fokus auf der Analyse einzelner Unternehmen. Für die Wissenschaft ist es möglich, Datenextrakte zu erhalten. Die Kosten sind abhängig vom beantragten Datenumfang. Das FDZ der BA im IAB hat einen solchen Datenextrakt für die Datenbank Dafne erworben, der Bilanz- und Finanzinformationen für über eine Millionen Unternehmen in Deutschland enthält. Aufgrund der Nutzungsbedingungen des BvD können momentan nur Mitarbeiter/innen des FDZ der BA im IAB diese Daten nutzen.

Tabelle 3.6: Überblick des Datenangebots des Bureau van Dijk

		Kapitel
Datenanbieter	Bureau van Dijk	
Homepage	http://www.bvdinfo.com/de-de/home	
Datenangebot für die Mindestlohnforschung	Dafne	5.4.8

Quelle: Eigene Darstellung

⁷ Die Gemeinnützige Urlaubs- und Zusatzversorgungskasse für das Maler- und Lackiererhandwerk hat sich nicht zu einem Interview bereit erklärt, da sie ausschließlich hinsichtlich Urlaubsansprüche auskunftspflichtig sind.

3.6 Sonstige Datenanbieter

Neben den in den vorangegangenen Abschnitten beschriebenen Datenanbietern, gibt es eine Reihe weiterer, bei denen eine Nutzung durch Dritte entweder noch nicht eindeutig geklärt ist oder deren Datenbestände und -angebote in den Abschnitten 4 bis 7 näher beschrieben werden. Im Folgenden werden die Datenanbieter kurz vorgestellt.

Von den derzeit 25 durch den Rat für Sozial- und Wirtschaftsdaten (RatSWD) akkreditierten Forschungsdatenzentren (FDZ) wurden oben bereits das FDZ der BA im IAB, das FDZ des Statistischen Bundesamtes, das FDZ der Statistischen Landesämter sowie das FDZ des BIBB näher beschrieben. Im Rahmen dieser Studie wurden Daten aus drei weiteren FDZ auf ihre Verwendbarkeit überprüft: Das FDZ der Rentenversicherung, das FDZ des Sozio-Oekonomischen Panels (SOEP) am Deutschen Institut für Wirtschaftsforschung in Berlin (DIW) sowie das FDZ des Nationalen Bildungspanels (National Educational Panel Study – NEPS).

Das FDZ der Rentenversicherung stellt ähnlich wie das FDZ der BA im IAB prozessproduzierte Daten aus dem DEÜV-Meldeverfahren für Forschungszwecke zur Verfügung. Dabei bietet es ebenfalls die beiden Off-Site Nutzungsmöglichkeiten (Public- und Scientific-Use-Files) als auch die beiden On-Site Alternativen Gastwissenschaftlerarbeitsplatz (GWA) und kontrollierte Datenfernverarbeitung (DFV). Allerdings erfolgt der Zugang über kontrollierte Datenfernverarbeitung ausschließlich nach einem Gastwissenschaftleraufenthalt an einem der beiden Standorte in Berlin oder Würzburg. Der Datenbestand des FDZ der Rentenversicherung und dessen Verwendbarkeit für die Mindestlohnforschung wird im Unterabschnitt 7.6 beschrieben.

Die beiden weiteren FDZ werden durch Wirtschaftsforschungsinstitute oder Universitäten unterhalten. Dementsprechend stellen sie keine amtlichen Daten zur Verfügung, sondern greifen auf eigene Erhebungen zurück. Das FDZ des SOEP am DIW betreut mit dem SOEP die größte Panelbefragung deutscher Haushalte (siehe Unterabschnitt 4.3.6). Neben der Bereitstellung der faktisch anonymisierten Daten über die gängigen Zugangswege (SUF, DFV und GWA), beinhaltet dies eine umfangreiche Dokumentation sowie eine Beratung bei der Nutzung der Daten. Das FDZ des NEPS ist am Institut für bildungswissenschaftliche Längsschnittforschung (INBIL) an der Otto-Friedrich-Universität Bamberg angesiedelt. Neben einer detaillierten Dokumentation der Daten des Nationalen Bildungspanels werden die Daten für Forschungszwecke über die drei Zugangswege SUF, DFV und GWA zur Verfügung gestellt. Der Datenbestand und das Nutzungspotenzial werden im Unterabschnitt 7.10 beschrieben.

Die Deutsche Bundesbank stellt ebenfalls Daten für Forschungszwecke zur Verfügung. Allerdings unterhält die Deutsche Bundesbank zurzeit kein durch den RatSWD akkreditiertes FDZ. Die Daten sind ausschließlich über einen Gastwissenschaftlerarbeitsplatz am Sitz der Deutschen Bundesbank in Frankfurt erhältlich. Datenbestand und Nutzungspotenzial werden im Unterabschnitt 7.8 kurz beschrieben

Weitere mögliche Datenanbieter sind die Gewerkschaften und Verbände der Mindestlohnbranchen bzw. deren Dachverbände und gemeinsamen Institutionen. Im Rahmen dieser Studie werden Daten des Hauptverbandes der Deutschen Bauindustrie e.V. (Unterabschnitt 6.3) sowie des Wirtschafts- und Sozialwissenschaftlichen Instituts der Hans-Böckler-Stiftung (WSI) behandelt. Zu den besprochenen Daten des WSI gehört zum einen das Tarifarchiv (Unterabschnitt 6.4), zum anderen die Mindestlohndatenbank (Unterabschnitt 7.9).

In einigen Instituten sind für die Mindestlohnforschung interessante Datensätze entstanden, die aufgrund datenschutzrechtlicher Auflagen ausschließlich in Kooperation mit den jeweiligen Instituten detailliert ausgewertet werden können. Die Freiburger Forschungsstelle Arbeits- und Sozialmedizin (ffas) hat umfangreiche Studien zur gewünschten und erlebten Arbeitsqualität anhand des Copenhagen Psychological Questionnaires (COPSOQ) durchgeführt. Ein kleiner Auszug der Daten ist im Internet kostenlos verfügbar. Ein Zugriff auf die gesamte Datenbank ist ausschließlich in Kooperation mit Mitarbeitern oder Mitarbeiterinnen der ffas möglich (siehe Unterabschnitt 4.3.5). Ähnlich ist die Struktur beim Mannheimer Unternehmenspanel (MUP) des Zentrums für Europäische Wirtschaftsforschung (ZEW). Das ZEW betreibt für die Bereitstellung eigener Forschungsdaten ein FDZ, das MUP steht aber aus Datenschutzgründen hier nicht zur Verfügung. Standardisierte Auswertungen können kostenpflichtig bestellt werden. Tiefergehende Analysen können ausschließlich in Kooperation mit Mitarbeiterinnen oder Mitarbeitern des ZEW vorgenommen werden (Unterabschnitt 5.4.9).

Abschließend sind im Auftrag des Bundesministeriums für Arbeit und Soziales (BMAS) einige Datensätze, die teilweise im Zusammenhang mit der ersten Runde der Mindestlohnevaluationen erhoben wurden. Diese Datensätze wurden in der Regel für den Forschungsauftrag ausgewertet, stehen aber für eine allgemeine Nutzung durch die Forschung nicht zur Verfügung. Hier ist insbesondere zu klären, wie die Daten zur Verfügung gestellt werden können. Dies betrifft vor allem die Beschäftigtenbefragung im Bauhauptgewerbe des Forschungsverbundes aus dem Institut für Arbeitsmarkt- und Berufsforschung (IAB), dem Rheinisch-Westfälischen Institut (RWI) und dem Institut für Sozialforschung und Gesellschaftspolitik (ISG) sowie die Betriebsbefragungen im Elektrohandwerk durch das Institut für Angewandte Wirtschaftsforschung (IAW), im Maler- und Lackiererhandwerk durch das IAW, der Pflegebranche durch T.I.P. Biehl & Partner sowie der Wäschereidienstleistungen durch das Institut Arbeit und Qualifikation (IAQ). Tabelle 3.7 gibt einen Überblick über die Daten der in diesem Abschnitt vorgestellten Datenanbieter.

Tabelle 3.7: Datenangebot weiterer Datenanbieter

Datensatz	Datenanbieter	Kapitel
Datenbestand der Deutschen Rentenversicherung	FDZ der Rentenversicherung	7.5
Sozio-Oekonomisches Panel (SOEP)	FDZ des SOEP	4.3.6
Nationales Bildungspanel (NEPS)	FDZ des NEPS	7.10
Mikrodaten aus der Deutschen Bundesbank	Deutsche Bundesbank	7.8
Elektronisches Verbandsinformations-, Recherche- und Analysesystem (ELVIRA)	Hauptverband der Deutschen Bauindustrie e.V.	6.3
WSI-Tarifarchiv	WSI	6.4
Mindestlohn Datenbank des WSI	WSI	7.8
Mannheimer Unternehmenspanel (MUP)	ZEW	5.4.9
Datenbank des Copenhagen Psychological Questionnaire (COPSOQ)	ffas	4.3.5
Befragung zum Mindestlohn im Elektrohandwerk	IAW & BMAS	5.4.4
Befragung zum Mindestlohn im Maler- und Lackiererhandwerk	IAW & BMAS	5.4.5
Befragung zum Mindestlohn in der Pflegebranche	T.I.P. Biehl & Partner & BMAS	5.4.6
Befragung zum Mindestlohn in der Branche Wäschereidienstleistungen im Objektkundengeschäft	IAQ & BMAS	5.4.7
Befragung zur gewünschten und erlebten Arbeitsqualität	ffas/FFP/infas & BMAS	4.3.7
Beschäftigtenbefragung Bauhauptgewerbe	IAB/RWI/ISG & BMAS	4.3.1

Quelle: Eigene Darstellung

4 Daten auf Individualebene

4.1 Administrative Daten

4.1.1 Integrierte Erwerbsbiografien (IEB)

4.1.1.1 Allgemeine Datensatzinformationen

Bei den Integrierten Erwerbsbiografien (IEB) handelt es sich um einen administrativen Datensatz, der vom IAB aufbereitet wird. Die IEB enthalten Informationen zu Beschäftigungszeiten, Leistungsempfangszeiten, Zeiten der Arbeitslosigkeit und Arbeitsuche sowie Maßnahmeteilnahmen von Personen und ermöglichen es somit, Erwerbsverläufe nachzuvollziehen.

Da in der IEB verschiedene Quelldatensätze zusammengeführt werden, unterscheidet sich die Grundgesamtheit nach der jeweiligen Quelle. Zunächst sind Daten aus dem „Gemeinsamen Meldeverfahren zur gesetzlichen Kranken-, Pflege-, Renten- und Arbeitslosenversicherung“ enthalten. Diese liegen in Westdeutschland seit 1975 und in Ostdeutschland seit 1991 vor. Neben den versicherungspflichtig Beschäftigten sind außerdem seit 1999 Angaben zu geringfügig Beschäftigten enthalten.

Des Weiteren beinhalten die IEB Meldungen aus mehreren Fachverfahren der BA. Dazu gehören ab 1975 Angaben über Leistungsbezug nach SGB III für Bezieher/innen von Arbeitslosengeld, Arbeitslosenhilfe und Unterhaltsgeld. Seit den gesetzlichen Neuregelungen

im Zuge der Arbeitsmarktreformen sind außerdem seit 2005 Informationen zur Leistungsgewährung nach SGB II für Personen und Bedarfsgemeinschaften aus einem anderen Verfahren enthalten.

Darüber hinaus sind Daten zu Maßnahmen, Maßnahmeteilnahmen und -trägern der aktiven Arbeitsmarktpolitik nach SGB II und III enthalten. Diese sind seit 2000 erfasst. Ebenfalls seit 2000 umfassen die IEB Daten zur Arbeitssuche, sofern diese über ein BA-Verfahren erfasst wurde.

Ab 2005 konnten einige Kommunen Aufgaben im Bereich des SGB II eigenständig verwalten (sog. „Optionskommunen“). Daten von diesen Trägern aus den Bereichen der Grundsicherung und Arbeitssuche sind ebenfalls in der IEB enthalten. Allerdings ist die Vergleichbarkeit dieser Angaben mit denen aus Prozessen der BA zum Teil eingeschränkt.

Da nicht in allen Quellen eine Versicherungsnummer oder BA-Kundennummer zur Verfügung steht, werden Informationen über das Konzept der Einheitlichen Statistischen Person zusammengeführt. Die IEB enthält Informationen zu über 80 Millionen Personenkonten mit insgesamt beinahe 2 Milliarden einzelnen Sätzen. Sie stellt damit die Grundgesamtheit der Personen dar, die von den benannten Meldegründen betroffen sind.

4.1.1.2 Verfügbarkeit

Die IEB wird durch das IAB angeboten. Innerhalb des IAB ist der Geschäftsbereich IT- und Informationsmanagement für die Erstellung, Aktualisierung und Dokumentation der IEB zuständig und übernimmt auch die Datenziehung. Die Beantragung folgt gemäß den im Unterabschnitt 3.1 beschriebenen Voraussetzungen. Dauer und Kosten der Beantragung sind abhängig vom genauen Datenbedarf und der Komplexität der Datenziehung.

Die IEB wird jährlich aktualisiert und ist derzeit für den Zeitraum 1975 bis 2011 verfügbar. Da Meldungen zur Sozialversicherung zum Teil erst mit einer deutlichen Verzögerung eingehen, wird die IEB nicht näher am aktuellen Rand gefüllt. Metadaten zur IEB unterliegen der statistischen Geheimhaltung. Daher sind Datenreporte und Auszählungen nicht frei zugänglich. Sie können erst im Rahmen eines genehmigten Projekts zur Verfügung gestellt werden.

4.1.1.3 Merkmale zur Mindestlohnevaluation

Die IEB enthält eine Reihe von Merkmalen, die für die Mindestlohnforschung von hoher Relevanz sind. Die folgenden Ausführungen erfolgen auf Basis der Annahme, dass das volle Analysepotenzial der IEB genutzt werden kann. Es ist jedoch jeweils zu klären, inwieweit eine Bereitstellung mehrerer sehr detaillierter Informationen aus Datenschutzgründen überhaupt möglich ist.

Die IEB enthält tagesgenaue Angaben zu sozialversicherungspflichtigen Beschäftigungszeiten sowie das im jeweiligen Zeitraum erzielte Bruttoentgelt. Das Entgelt wird nur bis zur Beitragsbemessungsgrenze erfasst, was in Mindestlohnstudien aber keine so bedeutende Rolle

spielt wie bei manchen anderen Fragestellungen. Andererseits sind Sonderzahlungen mit im Entgelt enthalten und es liegen bis 2009 keine Angaben zu Arbeitsstunden vor (abgesehen von einem Indikator für Voll- und Teilzeit). Dies erschwert die Abgrenzung von durch den Mindestlohn betroffenen Arbeitnehmerinnen und Arbeitnehmern, vor allem wenn sie nicht Vollzeit beschäftigt sind. In manchen Branchen ist es notwendig, bezüglich der Mindestlohn-betroffenheit Arbeiter und Angestellte zu trennen. Dies ist über die Angabe zur Stellung im Beruf möglich.⁸

Übergänge von und in Beschäftigung können mit der IEB untersucht werden, da Zeiten in Arbeitslosigkeit oder Leistungsbezug ebenfalls tagesgenau abgebildet sind. Wechsel des Arbeitgebers sind über einen Betriebsidentifikator zu erkennen. Beschäftigungszeiten als Beamte oder Selbstständige sind allerdings nicht enthalten und erscheinen im Datensatz - ebenso wie Zeiten der Nicht-Partizipation am Arbeitsmarkt – als Lücken in der Erwerbsbiografie.

Die IEB enthält außerdem eine Reihe von Angaben über persönliche Charakteristika, die bei statistischen Analysen als weitere Variablen Einsatz finden können, etwa zum Alter, zur Staatsangehörigkeit, Bildungsabschlüssen, dem ausgeübten Beruf und dem Geschlecht.

Auf Betriebsebene sind neben einem Regionalschlüssel bis auf Gemeindeebene auch Angaben zum Wirtschaftszweig bis hin zur 5-Steller-Ebene vorhanden. Diese Definition änderte sich im Zeitablauf mehrmals. Vor 1998 sind nur Informationen zur dreistelligen Wirtschaftsklasse 73 vorhanden. Ab 1999 gibt es auch für Beschäftigungsmeldungen eine gemeindscharfe Angabe zum Wohnort.

Bei den Angaben zum Arbeitsort ist insbesondere auch die Logik der Betriebsnummernvergabe zu beachten. Hierbei können mehrere Arbeitsstätten des Arbeitgebers zu einem Betrieb zusammengefasst werden, wenn sie sich in derselben Gemeinde befinden und dieselbe wirtschaftsfachliche Zuordnung aufweisen. Dadurch werden Beschäftigte demselben Betrieb zugeordnet, auch wenn sie an unterschiedlichen Standorten innerhalb einer Gemeinde tätig sind.

4.1.1.4 Merkmale zur Verknüpfung

Die IEB ist über entsprechende Identifikatoren mit anderen Datensätzen des IAB (insbesondere dem IAB-Betriebspanel) verknüpfbar. Darüber hinaus können auch Informationen aus anderen Datenbeständen der BA genutzt werden. Beispielsweise könnten Daten aus der Förderstatistik genutzt werden, um Betriebe in der Aus- und Weiterbildungsbranche zu identifizieren, die Maßnahmen nach dem SGB II und III durchführen. Dies könnte bei der Analyse von Mindestlohnwirkungen in dieser Branche hilfreich sein. Vorläufige Auswertungen von IAB-ITM zeigen, dass für einen überwiegenden Teil der Teilnahmen an geförderten Maß-

⁸ Diese Angabe fiel mit der Umstellung des Tätigkeitsschlüssels weg (siehe Exkurs). Arbeiter müssen nun über den Beruf identifiziert werden.

nahmen ein entsprechender Förderbetrieb identifiziert werden kann. Die Trefferquoten liegen etwa im Bereich Förderung beruflicher Weiterbildung bei ca. 92%, bei Trainingsmaßnahmen bei ca. 94% und bei durch den Europäischen Sozialfonds mitfinanzierten Qualifizierungsmaßnahmen im Rahmen des Kurzarbeitergelds sogar bei fast 100%. Den so identifizierten Betrieben könnten dann die dort Beschäftigten aus der IEB zugespielt werden.⁹

Außerdem ist es technisch möglich, weitere Identifikatoren und Adressdaten zu Personen und Betrieben zuzuspielen, die an sich nicht in der IEB enthalten sind. Dies gilt auch für einige weitere Originalschlüssel wie etwa die Versicherungsnummer. Dadurch wird eine direkte Verknüpfung oder ein Record Linkage mit anderen Datenquellen möglich. Hierzu zählen beispielsweise die Daten der Deutschen Rentenversicherung oder der Sozialkassen. Datenschutzrechtliche und technische Einschränkungen sollten allerdings möglichst frühzeitig geklärt werden.

Neben einer individuellen Verknüpfung ermöglichen der Regionalschlüssel sowie der Wirtschaftszweig eine Verknüpfung der Daten auf diesen aggregierten Ebenen. Dadurch können eine Vielzahl anderer Statistiken zugespielt und auch Makroanalysen durchgeführt werden.

4.1.1.5 Identifikation des Evaluationspotenzials

Die IEB ist für die Mindestlohnforschung aus einer Reihe von Gründen sehr gut geeignet. Da sie die Grundgesamtheit abbildet, sind auch sehr detaillierte Analysen und Auswertungen für Subpopulationen durch ausreichende Fallzahlen gesichert. Die Herkunft aus administrativen Quellen sichert zumindest bei einigen zentralen Angaben eine sehr hohe Datenqualität. Es lässt sich über lange Zeiträume ein tagesgenaues Panel auf individueller (oder aggregierter) Ebene erstellen, ohne dass Erinnerungslücken von befragten Personen ein Problem darstellen.

Eine Einschränkung der Nutzbarkeit besteht darin, dass Stundenlöhne nicht präzise zu berechnen sind und nur unter relativ restriktiven Annahmen abgeschätzt werden können. Hier bieten andere Datensätze (SOEP/VSE/Mikrozensus) mehr Möglichkeiten. Auch wenn die Daten auf Betriebsebene aggregiert werden können, sind den Analysemöglichkeiten aus Arbeitgebersicht (etwa Wettbewerb) aufgrund des quellenbedingt eingeschränkten Informationsgehalts gewisse Grenzen gesetzt. Gerade hier wäre eine genauere Prüfung der Verknüpfbarkeit mit anderen Datenbeständen, die detailliertere Unternehmensangaben beinhalten, wünschenswert.

⁹ Es ist jedoch zu bedenken, dass die Förderstatistik keine Auskunft gibt, ob ein Betrieb „überwiegend“ diese Leistungen erbringt. Da der momentan geltende Branchenmindestlohn nur in diesem Fall greift, müssten dennoch weitere umfangreiche Arbeiten bzw. Befragungen durchgeführt werden. Es ist bislang nicht geklärt, ob die Beschäftigungsstruktur eines Betriebs oder Daten aus anderen Quellen bei dieser Identifizierung helfen können. Außerdem fehlen Daten zu zugelassenen kommunalen Trägern in der Förderstatistik.

Gerade für die Evaluation von Beschäftigungs- und Lohneffekten von Mindestlöhnen ist die IEB trotz dieser Einschränkungen im Vergleich zu anderen Datenprodukten sehr gut geeignet. Aufgrund der starken Zugangsbeschränkungen ist allerdings immer zu prüfen, ob für eine bestimmte Fragestellung nicht auch andere Datenprodukte, etwa die SIAB oder das BHP, ausreichend sind.

Exkurs: Die Umstellung des Meldeverfahrens zur Sozialversicherung

Um für die bereits verwendeten Daten etwaige Änderungen zu untersuchen, die das Analysepotenzial beeinflussen könnten, sollen in diesem Abschnitt die Folgen der Einführung eines neuen Tätigkeitsschlüssels in das DEÜV-Meldeverfahren sowie die Qualität der mit der Erweiterung des DEÜV-Meldeverfahrens in den IEB erfassten Arbeitsstunden analysiert werden.

a) Aktualisierung des Meldeverfahrens zur Sozialversicherung

Seit dem 1. Dezember 2011 findet ein neuer Tätigkeitsschlüssel im Meldeverfahren zur Sozialversicherung Anwendung. Zentrale Neuerung ist die Übermittlung von Angaben zur ausgeübten Tätigkeit auf Grundlage eines 9-stelligen anstelle eines zuvor 5-stelligen Schlüssels (vgl. Bundesagentur für Arbeit, 2011). Da letzterer bereits 1992 ausgegeben und lediglich im Jahr 2003 aktualisiert wurde, war insbesondere eine Abbildung der aktuellen Berufsstrukturen auf dem deutschen Arbeitsmarkt nicht mehr adäquat möglich. Daher wurde eine neue „Klassifikation der Berufe 2010“ (KldB 2010) entwickelt, welche die 3-stellige Erfassung von Berufsfeldern (nach der bisherigen KldB 1988) ablöst und durch eine 5-stellige Erfassung ersetzt.¹⁰

Einerseits gestattet die neue Gliederungsstruktur eine differenziertere Abbildung des ausgeübten Berufes. Außerdem erfolgt eine Untergliederung von Teilzeitbeschäftigten nun nach der Stellung im Beruf, was zuvor nur für Vollzeitbeschäftigte möglich war. Auch eine genauere Differenzierung der Variable „(Aus-)Bildung“ (z.B. *schulisch* getrennt von *beruflich*) sowie Analysen zu den Merkmalen „Arbeitnehmerüberlassung“ und „Befristung“ sind nach dem neuen Verfahren möglich. Dies eröffnet eine Vielzahl neuer Analysemöglichkeiten.

Andererseits ist eine konsistente Vergleichbarkeit von Angaben auf Grundlage der unterschiedlichen Systematiken nur eingeschränkt möglich. Erste Analysen haben ergeben, „dass die Informationen nach neuer und alter Struktur, auch bei eigentlich gleichen Sachverhalten, unterschiedliche Ergebnisse liefern [können]“ (Dundler/Frank, 2012, S.4). Darüber hinaus scheint es im Zuge der Anpassung zu Aktualisierungsprozessen seitens der Arbeitgeber zu kommen, die zwar einen Zeitvergleich erschweren aber gleichzeitig die

¹⁰ Eine detaillierte Beschreibung der KldB 2010 findet sich in Paulus/Matthes (2013).

Datenqualität erhöhen.¹¹ Des Weiteren war die Einführung des neuen Tätigkeitsschlüssels mit einer Umstellungsphase von Dezember 2011 bis Mai 2013 verbunden. Während dieses Zeitraums werden Meldungen entweder auf Grundlage des alten oder des neuen Tätigkeitsschlüssels erfasst, wobei der Anteil der Meldungen nach der neuen Struktur sukzessive ansteigt. Jedoch fehlt bei beinahe 17 % der Meldungen in diesem Zeitraum die Angabe zur Tätigkeit vollständig. Dies betreffe insbesondere Auszählungen zu den Merkmalen „Beruf“, „Bildung“ und „Stellung im Beruf“, falls das Fehlen einer Angabe zum Tätigkeitsschlüssel systematisch erfolgt. Dies soll im folgenden Abschnitt untersucht werden.

b) Systematik fehlender Werte im Tätigkeitsschlüssel während der Übergangsphase von Dezember 2011 bis Mai 2013

Gültige Statistiken auf Grundlage der Merkmale „Beruf“, „Bildung“ und „Stellung im Beruf“ sind nur dann im Übergangszeitraum Dezember 2011 bis Mai 2013 konstruierbar, wenn die zuvor beschriebenen fehlenden Werte in den jeweiligen Variablen unsystematisch, d.h. zufällig verteilt sind. Eine notwendige Bedingung hierfür ist, dass der Anteil fehlender Werte über die Ausprägungen einer Variablen hinweg konstant ist. Um dies zu untersuchen, wurde der Personenkreis auf die Personen in der Beschäftigtenhistorik (BEH) beschränkt, die zu Beginn 2011 im selben Betrieb wie Ende 2010 gemeldet waren. Dies ist für etwa 61 % der 2011 gemeldeten Personen möglich. Für knapp 18 % dieser Personen fehlt die Angabe zum Tätigkeitsschlüssel. Unter der Annahme, dass sich der Tätigkeitsschlüssel zwischen Ende 2010 und Anfang 2011 nicht geändert hat, lassen sich die fehlenden Werte durch die Angaben aus dem Vorjahr ergänzen. Anhand der 82 % der Personen, für die in beiden Jahren eine Angabe zum Tätigkeitsschlüssel vorliegt, lässt sich die Wahrscheinlichkeit abschätzen, mit der die getroffene Annahme verletzt ist: Für die Variable Berufsordnung, welche in der BEH anhand eines Schwerpunktumstiegs von der KldB 2010 auf die KldB 1988 rückschlüsselt, weicht die Angabe 2010 in 28,3 % der Fälle von der Angabe 2011 ab. Der Schwerpunktumstieg ist hierbei nicht unproblematisch, da die den beiden Berufsklassifikationen zugrunde liegende Systematik voneinander abweicht. Deshalb sollte dieser Wert als „Obergrenze“ gesehen werden. Die Abweichungen scheinen jedoch zu häufig um alleine auf den Schwerpunktumstieg zurückgeführt werden zu können. Für die Variable „Bildung“ ist die Abweichungswahrscheinlichkeit mit 26,7 % etwas geringer, für die Variable „Stellung im Beruf“ ist sie mit 67,7 % jedoch deutlich höher.

Vor diesem Hintergrund kann nun für die genannten Variablen untersucht werden, wie sich die fehlenden Werte 2011 auf die einzelnen Ausprägungen verteilen. Tabelle 1 enthält in der Spalte „Anteil 2010 (%)“ für die jeweilige Variable den prozentualen Anteil der

¹¹ Dundler/Frank (2012) zeigen, dass der Anteil der sozialversicherungspflichtig Teilzeitbeschäftigten vom Stichtag 30. Juni 2011 bis zum Stichtag 30. September 2011 von 20 auf 21,7 Prozent gestiegen ist. Im Vorquartal betrug der Anstieg lediglich 0,3 Prozentpunkte. Da die Variable „Teilzeit“ konsistent über die Klassifikationen fortgeschrieben werden kann, deutet dies auf Aktualisierungseffekte hin, die zwar merklich aber nicht gravierend sind.

Beobachtungen, den die einzelnen Ausprägungen im Jahr 2010 an allen Beobachtungen einnehmen. Dies charakterisiert die Verteilung der Variablen in einem Zustand ohne fehlende Angaben zum Tätigkeitsschlüssel. Die Spalte „Anteil missing 2011 (%)“ zeigt diese Verteilung für die Untergruppe der Beobachtungen, für die im Jahr 2011 die eigentliche Angabe zum Tätigkeitsschlüssel fehlt, für die aber eine Übertragung aus dem Vorjahr möglich war. Die Spalte „Abweichung“ enthält die Differenz beider Werte.

Für die Variable Berufsordnung zeigt sich beispielsweise, dass die Gruppe von Personen, die zu Beginn 2011 im selben Betrieb wie Ende 2010 gemeldet waren, 11,6 % Bürofachkräfte enthält, was zugleich die häufigste Ausprägung ist. Allerdings entfällt deren Anteil mit 9,2 % in der Gruppe von Personen ohne gültige Angaben zum Tätigkeitsschlüssel deutlich geringer aus – es ergibt sich eine Abweichung von 2,3 Prozentpunkten. Dies zeigt, dass Bürofachkräfte unter den fehlenden Werten eine um 2,3 Prozentpunkte geringere Auftretenswahrscheinlichkeit aufweisen als in der Gesamtpopulation und somit deutlich unterrepräsentiert sind. Kraftfahrzeugführer hingegen treten in der Gesamtpopulation 2010 mit einer Wahrscheinlichkeit von 3,2 % auf. Ihr Anteil in der Verteilung fehlender Werte 2011 hingegen ist mit 4,1 % deutlich höher. Folglich ist die Gruppe der Kraftfahrzeugführer unter den fehlenden Werten überrepräsentiert.

Hierbei zeigt sich ein generelles Muster: Meldungen über Berufe im Dienstleistungssektor weisen eine geringere Wahrscheinlichkeit auf, im Jahr 2011 eine fehlende Angabe zum Tätigkeitsschlüssel zu enthalten (hohe negative Abweichungen) als Meldungen über handwerkliche Berufe o.Ä. Dies lässt sich auch bei Betrachtung der Variablen „Bildung“ und „Stellung im Beruf“ erkennen: Die Ausprägungen „Abitur“ und „Hochschulabschluss“ weisen deutlich seltener fehlende Angaben auf als die Ausprägung „Hauptschulabschluss“. Gleiches gilt für „Angestellte“ und „Arbeiter“. Insgesamt muss somit davon ausgegangen werden, dass das Fehlen von Angaben im Tätigkeitsschlüssel abhängig von den Eigenschaften der meldenden Betriebe ist. Ignoriert man dies bei Auszählungen der untersuchten Variablen erhält man Statistiken, die keinen Rückschluss auf die Gesamtpopulation erlauben, sondern lediglich Aussagekraft über die Verteilungen der Variablen in der Population mit gültigen Angaben zum Tätigkeitsschlüssel. Letztere ist tendenziell besser gebildet und häufiger in Dienstleistungsberufen als Angestellte tätig.

Ab Mai 2013 sollten alle Meldungen nach der neuen Erhebungsstruktur erfolgen. Bis dahin sind retrospektive Analysen auf Basis der betroffenen Merkmale nur auf einer Aggregationsebene möglich, auf der konsistent zwischen der alten und der neuen Erfassungssystematik fortgeschrieben werden kann. Gegenwärtig wird an einer derartigen Übertragung gearbeitet. Endgültige Ergebnisse liegen allerdings noch nicht vor.

Tabelle E.1: Systematik fehlender Werte

	Anteil fehlender Werte		Abweichung
	2010	2011	
Berufsordnung			
Bürofachkräfte	11,57%	9,23%	-2,34
Verkäufer	6,61%	4,87%	-1,74
Kellner/Stewards	2,46%	0,97%	-1,49
Raum-/Hausratreiniger	4,42%	3,43%	-0,99
...			
Kunststoffverarbeiter	0,44%	1,12%	0,68
Metallarbeiter o.n.a.	0,67%	1,47%	0,8
Kraftfahrzeugführer	3,24%	4,13%	0,89
Arbeiter mit Haushaltsscheckverf.	0,63%	2,70%	2,07
Bildung			
Unbekannt	30,50%	20,54%	-9,96
Abitur (mit BA)	3,83%	2,66%	-1,17
Abitur (ohne BA)	2,66%	1,60%	-1,06
Hochschule/Uni	5,04%	4,21%	-0,83
FH-abschluss	2,82%	2,14%	-0,68
ohne Ausbildung	0,02%	0,03%	0,01
Hauptschule (ohne BA)	12,65%	19,18%	6,53
Hauptschule (mit BA)	42,48%	49,64%	7,16
Stellung im Beruf			
Arbeitszeit < 18h/Woche	24,51%	6,62%	-17,89
Angestellter (nicht Meister)	28,64%	25,45%	-3,19
Heimarbeiter	0,07%	0,03%	-0,04
Keine Angabe	0,01%	0,05%	0,04
Meister/Polier	0,69%	0,84%	0,15
Fehlender Wert	0,15%	0,49%	0,34
Arbeitszeit > 18h/Woche Teilzeit	14,27%	15,47%	1,2
Auszubildende	5,25%	8,96%	3,71
Facharbeiter	11,99%	19,19%	7,2
Arbeiter	14,41%	22,89%	8,48

Quelle: IAB-Berechnungen auf Grundlage der Beschäftigtenhistorik

c) Erfassung der Arbeitsstunden in der Beschäftigtenhistorik

Seit dem 1. Januar 2009 sind die von den Arbeitnehmern und -nehmerinnen geleisteten Arbeitsstunden durch ihre Arbeitgeber gegenüber den Einzugsstellen der Krankenversicherung anzeigepflichtig.¹² In diesem Abschnitt soll zuerst die Definition der Arbeitsstunden sowie deren Erfassung erläutert werden. Anschließend erfolgt ein Abgleich mit den in der Verdienststrukturerhebung (VSE) der Statistischen Ämter erfragten Stunden, um eine erste Einschätzung der Datenqualität zu ermöglichen. In einem dritten Schritt soll die Eignung der Arbeitsstunden zur Berechnung von Stundenlöhnen und deren Potenzial zur Erweiterung des Analyserahmens von Mindestlohnuntersuchungen diskutiert werden.

¹² Da die Betriebsprüfung zur Unfallversicherung im Jahr 2010 von den Berufsgenossenschaften und Unfallkassen auf die Rentenversicherung übergegangen ist, war auch das DEÜV-Meldeverfahren um Angaben der gesetzlichen Unfallversicherung zu erweitern. Dies geschah im Zuge des Unfallversicherungsmodernisierungsgesetzes (UVMG).

Definition und Erfassung

Anzugeben sind die tatsächlich geleisteten vollen Arbeitsstunden für den jeweiligen Meldezeitraum. Ausfallzeiten wegen Krankheit, Urlaub oder Feiertagen sind hierbei nicht einzubeziehen. Weiterhin wird keine gesonderte Datenerfassung erforderlich: Informationen können aus bereits verfügbaren Quellen entnommen werden. Stehen diese nicht zur Verfügung, kann auf arbeitsvertragliche Sollwerte, also die vertraglich oder tariflich geschuldete Arbeitszeit abgestellt werden. Alternativ können die von der Deutschen Gesetzlichen Unfallversicherung e.V. (DGUV) jährlich festgelegten Vollarbeiterrichtwerte angegeben werden.¹³ Letztendlich sind auch qualifizierte Schätzungen zulässig.¹⁴

Grundsätzlich ist ein Vergleich von Stundeninformationen aus unterschiedlichen Datenquellen aufgrund abweichender Definitionen schwierig. Beispielhaft soll dies anhand eines Vergleichs mit den Konzepten im Mikrozensus, im Sozioökonomischen Panel (SOEP), im Panel „Arbeitsmarkt und soziale Sicherung“ (PASS), im Panel „Berufliche Weiterbildung als Bestandteil Lebenslangen Lernens“ (WeLL) und der Verdienststrukturerhebung (VSE) verdeutlicht werden:

- Im Mikrozensus wird die normalerweise geleistete Arbeitszeit je Woche erfragt. Dies schließt Überstunden mit ein (FDZ der Statistischen Ämter 2011).
- Im Sozioökonomischen Panel unterscheidet sich die Erhebung der Arbeitsstunden für die Jahre 2010 und 2011: Im Jahr 2010 wurde erfasst, wie viele Stunden an einem durchschnittlichen Werktag gearbeitet werden. 2011 wurde dies verfeinert, indem Stunden separat für Werktage, Samstage und Sonntage erfragt wurden. Außerdem wurde nach den durchschnittlichen Arbeitsstunden pro Werktag sowie den durchschnittlichen Werktagen pro Woche gefragt (vgl. TNS Infratest Sozialforschung, 2012).
- Im Panel „Arbeitsmarkt und soziale Sicherung“ sowie im Panel „Berufliche Weiterbildung als Bestandteil Lebenslangen Lernens“ wird die wöchentliche vertragliche Arbeitszeit erfasst. Zusätzlich existiert eine generierte Variable, die die wöchentliche Arbeitszeit inklusive Überstunden enthält (vgl. Berg et al., 2012, und Huber/Schmucker, 2012).

¹³ Die Vollarbeiterrichtwerte betragen 1.610 Stunden für das Jahr 2010 und 1.570 Stunden für das Jahr 2011. Ist die Beschäftigung nicht ganzjährig oder ganztätig ist dies anteilig zu melden.

¹⁴ In folgenden Fällen wird keine Angabe der Arbeitsstunden notwendig: Meldungen aus Abgangsgründen der Kategorien „Meldung des Unterschiedsbetrags bei Entgeltersatzleistungen während Altersteilzeit“, „Entgeltmeldung für unständig Beschäftigte“, „Jahresmeldung für freigestellte Arbeitnehmer“ und „Entgeltmeldung zum rechtlichen Ende der Beschäftigung“, sowie für Seelotsen, unständig Beschäftigte, mit Haushaltsscheckverfahren gemeldete geringfügig entlohnte Beschäftigte sowie für diverse Beschäftigungsgruppen, die in der Beschäftigtenhistorik generell nicht erfasst werden. Gleiches gilt, wenn der zuständige Unfallversicherungsträger der öffentlichen Hand angehört oder einer landwirtschaftlichen Berufsgenossenschaft entspricht, sowie für Mitarbeiter/innen der Unfallversicherungsträger selbst.

- In der Verdienststrukturerhebung wird die regelmäßige, wöchentliche Arbeitszeit als die vertraglich vereinbarte Arbeitszeit erhoben. Richtet sich die Entlohnung eines Arbeitnehmers bzw. einer Arbeitnehmerin nach den Arbeitsstunden, so sind auch die bezahlten Stunden anzugeben. Bezahlte Überstunden werden separat erfasst (vgl. Statistisches Bundesamt, 2013d).

Somit wird in Befragungen tendenziell die normalerweise zu leistende oder vertraglich festgelegte Arbeitszeit erhoben. Demgegenüber spiegeln die administrativ erhobenen Arbeitsstunden in der BEH eher die tatsächlich geleisteten Stunden wider. Dies ist gerade für die Analysen von Mindestlöhnen von großer Bedeutung, weshalb auf diesen Punkt gesondert eingegangen wird.

Abgleich mit Befragungsdaten

Eigene Auszählungen haben ergeben, dass etwa 16 % der Meldungen in der BEH keine gültige Stundenangabe aufweisen. Für die übrigen 84 % der Meldungen wird der Anteil pauschal gemeldeter Stunden auf etwa 15 % geschätzt.

Um die Qualität der gültigen Angaben näher zu untersuchen, sollen die administrativ erfassten Stundenangaben mit den Arbeitsstunden verglichen werden, die in der Verdienststrukturerhebung 2010 (VSE10) abgefragt wurden. Hierzu werden die durchschnittlichen Wochenarbeitsstunden, untergliedert nach Wirtschaftszweigen, gegenübergestellt.

Weiterhin erfolgt der Vergleich innerhalb von Gruppen, die einerseits dadurch bestimmt sind, dass das beschäftigende Unternehmen seinen Sitz in West- oder in Ostdeutschland hat. Andererseits werden die durchschnittlichen Wochenstunden von Voll- und Teilzeitkräften verglichen. Innerhalb der Vollzeitbeschäftigten wird nach dem Geschlecht untergliedert.

Durchschnittlich liegen die Angaben zu den wöchentlichen Arbeitsstunden in der BEH 5,6 Stunden unterhalb der in der VSE10 erfragten Stunden. Dies lässt sich durch die abweichenden Definitionen erklären, die dem jeweiligen Konzept zugrunde liegen: Die Nicht-Erfassung von Ausfallzeiten wegen Krankheit, Urlaub oder Feiertagen in den administrativen Daten führt zu signifikant geringeren Durchschnittswerten bei der wöchentlichen Arbeitszeit. In absoluten Werten sind diese Abweichungen für vollzeitbeschäftigte Frauen mit durchschnittlich 7,5 Stunden über alle Wirtschaftszweige hinweg am deutlichsten. Für in Betrieben mit Sitz in Ostdeutschland beschäftigte Personen beträgt die Abweichung durchschnittlich 5,3 Stunden und für alle Beobachtungen insgesamt liegt sie bei 6,2 Stunden. Mit Ausnahme der Wirtschaftszweige „Verarbeitendes Gewerbe“, „Baugewerbe“ und „Kunst, Unterhaltung und Erholung“ arbeiten Männer gemäß der VSE10 in beinahe allen Industrien im Schnitt länger als Frauen. Dies gilt laut der BEH in allen Sektoren, wobei die Unterschiede hier noch deutlich ausgeprägter sind. Im Vergleich von Voll- und Teilzeitbeschäftigten sind die Unterschiede in den erfassten Wochenstunden der VSE10 zufolge mit durchschnittlich 15,6 Stunden wiederum signifikant höher als die in der BEH mit durchschnittlich 10,8 Stunden. Beschäftigte in den neuen Bundesländern weisen laut beiden

Tabelle E.2: Arbeitsstunden nach Kategorien im Jahr 2010

	alle Kategorien			Vollzeit		Teilzeit	
	Gesamt	West	Ost	Gesamt	Männer	Frauen	Gesamt
Verdienststrukturerhebung: durchschnittlich bezahlte Wochenarbeitsstunden							
Bergbau und Gewinnung von Steinen und Erden	37,18	37,34	36,45	39,27	39,30	38,80	23,02
Verarbeitendes Gewerbe	34,93	34,63	37,13	37,70	37,68	37,80	24,16
Energieversorgung	35,42	35,30	36,14	38,40	38,40	38,39	24,36
Wasserver- und Abfallentsorgung	36,17	35,76	37,57	39,52	39,56	39,22	23,87
Baugewerbe	35,87	35,47	37,48	39,00	38,99	39,06	21,52
Handel; Instandhaltung von Kraftfahrzeugen	31,70	31,59	32,70	39,14	39,28	38,81	23,58
Verkehr und Lagerei	33,40	33,18	34,57	40,21	40,49	38,91	22,43
Gastgewerbe	28,27	27,74	31,09	39,61	39,74	39,48	24,08
Information und Kommunikation	34,71	34,69	34,92	39,11	39,21	38,84	23,53
Finanz- und Versicherungsdienstleistungen	34,14	33,99	36,37	38,94	38,96	38,91	23,15
Grundstücks- und Wohnungswesen	32,12	31,67	34,15	38,75	38,84	38,63	23,67
Freiberufl./wissenschaftl./techn. Dienstleistungen	33,87	33,82	34,33	39,26	39,28	39,21	22,88
Sonstige wirtschaftliche Dienstleistungen	28,41	27,78	31,68	37,92	37,95	37,85	23,26
Öffentliche Verwaltung/Verteidigung/Sozialvers.	35,68	35,51	36,47	39,13	39,14	39,12	24,52
Erziehung und Unterricht	31,27	31,09	32,32	39,04	39,25	38,88	23,11
Gesundheits- und Sozialwesen	30,13	29,49	33,80	39,11	39,28	39,01	24,55
Kunst, Unterhaltung und Erholung	29,89	29,26	32,98	39,55	39,49	39,64	22,45
Sonstige Dienstleistungen	30,16	29,72	32,86	39,10	39,25	38,97	23,49
Beschäftigtenhistorik: durchschnittlich geleistete Wochenarbeitsstunden							
Bergbau und Gewinnung von Steinen und Erden	30,71	30,78	30,43	31,96	32,14	29,82	21,02
Verarbeitendes Gewerbe	29,88	29,45	32,63	32,16	32,39	31,31	21,55
Energieversorgung	29,74	29,59	30,39	31,07	31,17	30,65	20,38
Wasserver- und Abfallentsorgung	31,80	31,19	33,17	34,51	34,88	32,29	22,71
Baugewerbe	30,69	30,21	32,34	34,56	35,06	29,48	19,34
Handel; Instandhaltung von Kraftfahrzeugen	26,10	25,88	27,31	33,38	34,29	31,93	21,93
Verkehr und Lagerei	28,91	28,37	31,23	35,55	35,92	33,86	21,87
Gastgewerbe	21,93	21,17	25,39	32,25	32,92	31,63	24,12
Information und Kommunikation	29,08	29,01	29,47	33,50	33,93	32,47	22,64
Finanz- und Versicherungsdienstleistungen	28,21	28,02	29,62	32,04	32,80	31,10	20,44
Grundstücks- und Wohnungswesen	20,05	18,85	24,71	31,64	32,46	30,65	19,00
Freiberufl./wissenschaftl./techn. Dienstleistungen	27,31	27,05	28,67	32,19	33,11	31,04	20,03
Sonstige wirtschaftliche Dienstleistungen	26,03	25,50	28,21	34,60	35,74	32,18	22,40
Öffentliche Verwaltung/Verteidigung/Sozialvers.	25,62	24,69	28,71	29,87	30,05	29,69	20,09
Erziehung und Unterricht	25,01	23,66	29,14	31,34	32,18	30,87	21,19
Gesundheits- und Sozialwesen	24,89	24,15	28,00	31,44	32,41	31,09	21,44
Kunst, Unterhaltung und Erholung	23,55	21,55	32,15	33,33	33,41	33,25	31,37
Sonstige Dienstleistungen	22,72	21,77	26,81	31,95	32,95	31,28	20,18

Quelle: Statistisches Bundesamt (2013d); IAB-Berechnungen auf Grundlage der Beschäftigtenhistorik.

Datenquellen im Schnitt eine höhere wöchentliche Arbeitszeit auf als Beschäftigte in den alten Bundesländern.

Durchschnittlich liegen die Angaben zu den wöchentlichen Arbeitsstunden in der BEH 5,6 Stunden unterhalb der in der VSE10 erfragten Stunden. Dies lässt sich durch die abweichenden Definitionen erklären, die dem jeweiligen Konzept zugrunde liegen: Die Nicht-Erfassung von Ausfallzeiten wegen Krankheit, Urlaub oder Feiertagen in den administrativen Daten führt zu signifikant geringeren Durchschnittswerten bei der wöchentlichen Arbeitszeit. In absoluten Werten sind diese Abweichungen für vollzeitbeschäftigte Frauen mit durchschnittlich 7,5 Stunden über alle Wirtschaftszweige hinweg am deutlichsten. Für in Betrieben mit Sitz in Ostdeutschland beschäftigte Personen beträgt die Abweichung

durchschnittlich 5,3 Stunden und für alle Beobachtungen insgesamt liegt sie bei 6,2 Stunden. Mit Ausnahme der Wirtschaftszweige „Verarbeitendes Gewerbe“, „Baugewerbe“ und „Kunst, Unterhaltung und Erholung“ arbeiten Männer gemäß der VSE10 in beinahe allen Industrien im Schnitt länger als Frauen. Dies gilt laut der BEH in allen Sektoren, wobei die Unterschiede hier noch deutlich ausgeprägter sind. Im Vergleich von Voll- und Teilzeitbeschäftigten sind die Unterschiede in den erfassten Wochenstunden der VSE10 zufolge mit durchschnittlich 15,6 Stunden wiederum signifikant höher als die in der BEH mit durchschnittlich 10,8 Stunden. Beschäftigte in den neuen Bundesländern weisen laut beiden Datenquellen im Schnitt eine höhere wöchentliche Arbeitszeit auf als Beschäftigte in den alten Bundesländern.

Aufgrund der abweichenden Definitionen zwischen den administrativ erhobenen Stundenangaben und den üblicherweise in Befragungen erfassten Stunden, ist eine abschließende Beurteilung der Qualität der in der BEH enthaltenen Arbeitsstunden auf diesem Wege nicht möglich. Zwar ergeben sich bei einem Vergleich der durchschnittlichen wöchentlichen Arbeitszeit diverser Arbeitsmarktgruppen zwischen den Angaben in der BEH und der VSE10 qualitativ kaum Unterschiede, die in den administrativen Daten erfasste Arbeitszeit liegt jedoch regelmäßig deutlich unterhalb der in der VSE10 erfassten Arbeitszeit.

Stundenlohnberechnung

Im vorherigen Abschnitt wurde gezeigt, dass die Arbeitsstunden in der BEH generell unterhalb den in der VSE10 erfragten Arbeitsstunden liegen. Dies ist in erster Linie darauf zurückzuführen, dass in Befragungen tendenziell die „normalerweise zu leistende“ oder „vertraglich festgelegte“ Arbeitszeit anstelle der „tatsächlich geleisteten“ Arbeitszeit erhoben wird. Somit ist davon auszugehen, dass diese Abweichung nach unten auch im Vergleich mit Befragungsdaten aus anderen Quellen vorliegt.¹⁵

Errechnet man den Bruttostundenverdienst innerhalb eines Zeitraums als den Quotienten zwischen der Lohnsumme und der Summe gearbeiteter Stunden, bewirkt die vergleichsweise geringe Stundenzahl in der BEH, selbst bei identischer Definition der Brutto-Lohnsumme, einen Stundenlohn, der im Schnitt höher ausfällt als der in der VSE10 erhobene Stundenlohn. Dies ist, untergliedert nach Voll- und Teilzeit sowie nach Geschlecht, in Tabelle 3 illustriert: Die durchschnittlichen Bruttostundenverdienste sind gemäß der VSE10 für alle Gruppen geringer als die in der BEH erfassten Stundenlöhne. Ebenfalls zu berücksichtigen ist, dass die in der BEH abgebildete Lohnsumme auch unregelmäßige Sonderzahlungen enthalten kann, die in dem „laufenden Arbeitslohn“, wie er in der VSE10 erhoben wird, explizit nicht enthalten sind. Dies trägt ebenfalls zu einer Abweichung der Stundenlöhne in der BEH nach oben bei.

¹⁵ Die Einbeziehung von Überstunden gemäß der Arbeitszeitdefinition in der BEH, welche in Befragungen in der Regel gesondert erfasst werden, wirkt dem entgegen. Im Schnitt wiegt diese jedoch weniger schwer als die Nicht-Erfassung von Ausfallzeiten wegen Krankheit, Urlaub oder Feiertagen.

Grundsätzlich hängt die Eignung der auf diesen unterschiedlichen Erfassungsweisen beruhenden Daten für die Beantwortung einer spezifischen Fragestellung von der vorliegenden Forschungsfrage selbst ab. Speziell für die Frage, wie hoch der Anteil an Personen in einer Branche mit verbindlichem Lohnminimum ist, die eine Entlohnung entsprechend oder gar unterhalb dieses Mindestlohns aufweisen, sollte auf Datenquellen zurückgegriffen werden, die die Lohnsumme in etwa als „laufenden Arbeitslohn“ (gemäß der Definition des Statistischen Bundesamts, 2013d) und die gearbeiteten Stunden als „normalerweise geleistete Arbeitszeit“ erfassen. Zieht man die Stundenlöhne aus der BEH zur Beantwortung dieser Frage heran, so sollte sich lediglich eine Untergrenze für die Schätzung dieses Anteils ergeben. Letztere Stundenlohnangaben sind demnach eher für Untersuchungen der *tatsächlichen* Mindestlohnbetreffenheit geeignet.

Grundsätzlich ist es natürlich auch möglich, über gewisse Annahmen die Arbeitsstunden in der BEH entsprechend der vorliegenden Forschungsfrage zu modifizieren, z.B. indem man einen branchenspezifischen Urlaubsanspruch abzieht. Generell ist jedoch davon auszugehen, dass sich ein umfassenderes Bild durch die Kombination von Stundenlohnangaben aus verschiedenen Datenquellen und beruhend auf unterschiedlichen Definitionen ergibt.

Tabelle E.3: Durchschnittlicher Bruttostundenverdienst im Jahr 2010

	Verdienststrukturerhebung			Beschäftigtenhistorik		
	Gesamt	Männer	Frauen	Gesamt	Männer	Frauen
Vollzeit	18,98	20,12	16,36	20,54	22,02	17,62
Teilzeit	14,96	15,49	14,86	17,13	16,32	17,27

Quelle: Statistisches Bundesamt (2013d); IAB-Berechnungen auf Grundlage der Beschäftigtenhistorik

4.1.2 Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB)

4.1.2.1 Allgemeine Datensatzinformationen

Die Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB) ist eine 2%-Stichprobe aus der Grundgesamtheit der Integrierten Erwerbsbiografien (IEB) des IAB. Mit Ausnahme der Teilnahme an arbeitsmarktpolitischen Maßnahmen sind alle Datenquellen der IEB auch in der SIAB enthalten.

Die SIAB ist eine 2%-Stichprobe der Personen, für die ein Eintrag in den IEB vorliegt. Dadurch ist die SIAB repräsentativ für die Personenkreise aller Datenquellen der IEB, nicht nur für die Beschäftigten. Daher lassen sich auf Basis der Stichprobe mit Hochrechnungsfaktor 50 Fallzahlen der Grundgesamtheit schätzen. Die SIAB liegen in „episodengesplitteter“ Form vor. Dies bedeutet, dass überlappende Sätze derart ersetzt werden, dass nur vollständig parallele oder vollständig überschneidungsfreie Zeiträume entstehen.

Das Design der SIAB als Zufallsstichprobe auf Personenebene führt dazu, dass eine Aggregation der Personeninformationen auf Betriebsebene nicht zu sinnvollen Angaben über den

jeweiligen Betrieb führen. Um dieses Problem auszugleichen werden alle betriebsspezifischen Informationen in der SIAB aus dem BHP zugespielt. Sie basieren daher auf Angaben zu allen sozialversicherungspflichtig Beschäftigten des jeweiligen Betriebs. Aufgrund der verschiedenen Ziehungslogik kommt es dabei zu einer Unterscheidung in der Zeitstruktur zwischen personenbezogenen und betriebsbezogenen Angaben. Während die personenbezogenen Angaben wie in der IEB tagesaktuell Zeiträume von Beschäftigung, Leistungsbezug und Arbeitssuche abbilden, beruhen die betriebsbezogenen Angaben auf einer Querschnittsauszahlung zum 30. Juni des jeweiligen Jahres.¹⁶

Der Datensatz umfasst in der aktuellen Version (SIAB 7510) die Erwerbsverläufe von ca. 1,6 Millionen Individuen, deren Erwerbsbiografien in insgesamt über 45 Millionen Datenzeilen dokumentiert sind. Zugehörige Betriebsinformationen liegen für etwa 1,4 Millionen Betriebe vor.

4.1.2.2 Verfügbarkeit

Die SIAB wird durch das FDZ der BA im IAB angeboten. Das FDZ ist auch für die Erstellung, Aktualisierung und Dokumentation zuständig und übernimmt die Nutzerbetreuung. Der Datensatz ist sowohl als schwach anonymisierte Version über einen Gastaufenthalt oder als SUF nutzbar. Die Beantragung folgt gemäß den im Unterabschnitt 3.1 beschriebenen Voraussetzungen.

Die Dauer der Beantragung ist abhängig vom Datennutzungsweg und dauert bei Gastaufenthalten aufgrund des zusätzlichen Genehmigungsverfahrens etwas länger. Sicherheitshalber sollten 2 bis 4 Wochen veranschlagt werden. Die Datennutzung selbst ist kostenlos.

Die SIAB wird alle zwei Jahre aktualisiert und ist derzeit für den Zeitraum 1975 bis 2010 verfügbar. Das nächste Update ist für Anfang 2015 geplant. Der abgedeckte Zeitraum soll dann bis 2012 erweitert werden. Umfangreiche Dokumentationen und Auszahlungen sind auf der Internetpräsenz des FDZ verfügbar.

4.1.2.3 Merkmale zur Mindestlohnevaluation

Da die SIAB auf den IEB und dem BHP aufbaut, ist das Evaluationspotenzial des Datensatzes zunächst identisch mit dem seiner Quellen. In diesem Abschnitt sollen daher lediglich Einschränkungen angesprochen werden, die gegenüber diesen Datensätzen zu beachten sind. Weitere Informationen sind in den Unterabschnitten 4.1.1 und 5.1.1 zu finden.

Da die SIAB eine Stichprobe aus der IEB ist, enthält sie Informationen über deutlich weniger Personen als der Ausgangsdatsatz. Sehr detaillierte Analysen und Auswertungen für sehr kleine Subpopulationen könnten dadurch an zu geringer Beobachtungszahl scheitern. Außerdem enthält die SIAB im Gegensatz zur IEB derzeit keine Informationen zu Maßnahmen,

¹⁶ Siehe hierzu auch die Abschnitte 4.1.1 (IEB) und 5.1.1 (BHP)

die im Kontext der Mindestlohnevaluation für die Analyse der Effekte auf Wiedereingliederungen relevant sein könnten. Gegenüber dem BHP ist zu beachten, dass die Betriebsinformationen nur für die in der Stichprobe ausgewählten Beschäftigten vorliegen. Diese Daten sind daher auf Betriebsebene nicht repräsentativ gezogen.

Die weitere Anonymisierung bringt es auch mit sich, dass einige Merkmalsausprägungen in der SIAB nur in vergrößerter Form vorliegen. So ist die Geburtstagsinformation auf das Geburtsjahr eingeschränkt und Regionalinformationen liegen lediglich auf Kreis- statt Gemeindeebene vor. Anders als die IEB ist die SIAB allerdings über alle Quellen hinweg auf einen einheitlichen Gebietsstand bereinigt.

Im SUF der SIAB sind die Einschränkungen deutlich stärker ausgeprägt. Hier können Personen nicht mehr zu einzelnen Betrieben zugeordnet werden. Bei den Angaben zum Arbeitsort mussten zum Teil Kreise aggregiert werden. Der Wohnort ist nicht mehr angegeben. Auch andere sensible Merkmale wie die Angaben zum Wirtschaftszweig, dem Beruf und der Nationalität sind zum Teil stark vergrößert.

4.1.2.4 Merkmale zur Verknüpfung

Die SIAB enthält in der schwach anonymisierten Version Personen- und Betriebsidentifikatoren, die über Schlüsseltabellen dieselben Verknüpfungen zulassen wie die IEB. Daher sei hier auf Unterabschnitt 4.1.1 verwiesen. Die Vertragsbestimmungen des FDZ lassen allerdings eine Verknüpfung der SIAB mit anderen Mikrodaten nicht zu. Im SUF sind keine verknüpfbaren individuellen Identifikatoren enthalten.

In beiden Fällen ermöglichen der Regionalschlüssel sowie der Wirtschaftszweig eine Verknüpfung der Daten auf aggregierter Ebene.

4.1.2.5 Identifikation des Evaluationspotenzials

Die SIAB ist für die Mindestlohnforschung vor allem dann geeignet, wenn Untersuchungen auf individueller Ebene durchgeführt werden sollen und die Nachteile der Nutzung der IEB (Kosten, Zugangsvoraussetzungen) die Vorteile gegenüber der SIAB (Grundgesamtheit, detailliertere Angaben zu einigen Variablen) überwiegen. Gerade wenn eine zu untersuchende Branche relativ groß ist, kann bei den Analysen auf die SIAB zurückgegriffen werden.

Abgesehen davon gelten für die SIAB die gleichen Einschränkungen wie für die IEB, insbesondere was das Fehlen einer Stundenlohnangabe und die daraus folgende Schwierigkeit der Abgrenzung der vom Mindestlohn direkt Betroffenen angeht.

4.1.3 Lohn- und Einkommenssteuerstatistik

4.1.3.1 Allgemeine Datensatzinformationen

Die Lohn- und Einkommenssteuerstatistik wird seit 1950 alle drei Jahre aus dem Datenbestand der Landesfinanzverwaltungen generiert. Es handelt sich um eine Vollerhebung administrativer Informationen, die im Rahmen des Lohn- und Einkommenssteuerverfahrens erhoben werden. Somit sind alle unbeschränkt und beschränkt Einkommenssteuerpflichtige im erhobenen Veranlagungsjahr im Datensatz vertreten. Allerdings gilt zu beachten, dass dies nicht gleichzusetzen ist mit Individuen, da bei einer gemeinsamen Veranlagung die Personen in der Lohn- und Einkommenssteuerstatistik zusammen aufgeführt sind (Statistisches Bundesamt 2012b).

Die Angaben, die aus den Lohnsteuerkarten gewonnen werden, sind zeitlich nur eingeschränkt vergleichbar, da 2004 erstmals elektronisch übermittelte Lohnsteuerbescheinigungen der Arbeitgeber vorlagen. 2007 ist die elektronische Übermittlung der Lohnsteuerbescheinigungen für die Arbeitgeber verpflichtend. Dementsprechend steigt die Anzahl der Beobachtungen mit ausschließlich Angaben aus dem Lohnsteuermeldeverfahren im Datensatz stark an (Statistisches Bundesamt 2013a). Insgesamt ist bei einem zeitlichen Vergleich zu berücksichtigen, dass Änderungen in der Steuergesetzgebung die erhobenen Daten und Merkmale beeinflussen können und dadurch die Vergleichbarkeit eingeschränkt wird (Statistisches Bundesamt 2012b).

Aufgrund der Veranlagungsdauer stehen die Daten den Statistischen Ämtern erst $2\frac{3}{4}$ Jahre nach dem Bezugsjahr zur Verfügung. Aufgrund der Komplexität der Daten und der aufwändigen Bereinigungen, die vorgenommen werden müssen, dauert es ein weiteres $\frac{3}{4}$ Jahr bis die Daten durch das Statistische Bundesamt veröffentlicht werden können (Statistisches Bundesamt 2012b). Teilweise kommt es aus unterschiedlichen Gründen zu weiteren Verzögerungen (Statistisches Bundesamt 2005), weshalb mit einer Zeitspanne von mindestens $3\frac{1}{2}$ Jahren zwischen Berichtszeitpunkt und Veröffentlichung der Daten zu rechnen ist. Derzeit sind dementsprechend die Daten bis 2007 erhältlich. Im Sommer 2014 ist mit der Erhebungswelle 2010 zu rechnen.

4.1.3.2 Verfügbarkeit

Die Lohn- und Einkommenssteuerstatistik wird ab der Welle 1992 über die FDZ der Statistischen Ämter des Bundes und der Länder angeboten. Daher folgt die Beantragung gemäß den im Unterabschnitt 3.2 beschriebenen Voraussetzungen.

Die Daten können sowohl als CAMPUS-Files für die Jahre 1998 und 2001, als SUF ab 1998 sowie über die On-Site Nutzungswege Gastwissenschaftlerarbeitsplatz und Datenfernverarbeitung ab 1992 genutzt werden. Dabei ist zu beachten, dass die Nutzungswege unterschiedliche Anforderungen an das Anonymisierungskonzept stellen und somit die Nutzungswege nicht die gleichen Möglichkeiten haben (siehe auch Unterabschnitt 3.2).

Unter anderem werden aufgrund der Anonymisierungsvoraussetzungen bei den CAMPUS-Files eine 1 %-Stichprobe (Statistisches Bundesamt 2008) und bei den SUF eine 10 %-Stichprobe (Statistisches Bundesamt 2013a) der Lohn- und Einkommensstatistik zur Verfügung gestellt. Über die Kontrollierte Datenfernverarbeitung oder einen Gastwissenschaftlerarbeitsplatz ist ein Zugriff auf die Vollerhebung möglich.

4.1.3.3 Merkmale zur Mindestlohnevaluation

Die Angaben in der Lohn- und Einkommenssteuerstatistik stammen einerseits aus den Lohnsteuerbescheinigungen, andererseits aus den Einkommenssteuererklärungen. Dementsprechend beziehen sich die Angaben hauptsächlich auf Einkünfte und steuerrechtlich relevante Ausgaben. Da keine Arbeitszeitinformationen vorliegen, ist eine Identifikation der vom Mindestlohn betroffenen Beschäftigten nicht möglich. Allerdings können anhand der Wirtschaftszweigklassifikation die Beschäftigten einer Mindestlohnbranche identifiziert werden.

Im Unterschied zu anderen Datensätzen, kann anhand der Lohn- und Einkommenssteuerstatistik das tatsächlich verfügbare Einkommen ermittelt werden. Zwar ist bei einer gemeinsamen Veranlagung das verfügbare Einkommen der beiden veranlagten Personen zusammengerechnet, dennoch können Auswirkungen des Mindestlohns auf das verfügbare Einkommen untersucht werden.

Weitere relevante Merkmale in der Lohn- und Einkommenssteuerstatistik sind vor allem soziodemografische Merkmale wie Geburtsdatum oder der Wohnort.

4.1.3.4 Merkmale zur Verknüpfung

Die Daten stammen aus dem Lohn- und Einkommenssteuerverfahren. Dadurch sind sie über die Steuernummer bzw. über den Klarnamen mit Wohnort und Geburtsdatum theoretisch mit weiteren administrativen Daten, wie z.B. den IEB (Unterabschnitt 4.1.1), auf Individualebene verknüpfbar. Über den Gemeindegeschlüssel ist eine Verknüpfung der Lohn- und Einkommenssteuerstatistik auf regionaler Ebene möglich.

4.1.3.5 Identifikation des Evaluationspotenzials

Die detaillierten Informationen über das verfügbare Einkommen ermöglichen im Unterschied zu anderen Datenquellen, die lediglich den Bruttolohn erfassen, eine Analyse der Auswirkungen des Mindestlohns auf die finanzielle Situation der betroffenen Beschäftigten. Außerhalb der Angaben zu Einkommen und steuerrelevanten Ausgaben beinhaltet die Lohn- und Einkommenssteuerstatistik wenige zusätzliche Informationen, die für Analysezwecke hinsichtlich des Mindestlohns genutzt werden können. Außerdem liegen keine Angaben über die Arbeitszeit vor, was eine Identifikation der direkt vom Mindestlohn betroffenen Beschäftigten unmöglich macht.

Die Lohn- und Einkommenssteuerstatistik könnte als sinnvolle Ergänzung zu anderen Datensätzen dienen, wenn eine Verknüpfung nicht nur wie beschrieben technisch, sondern

auch datenschutzrechtlich möglich ist, da dann die detaillierten Informationen zum verfügbaren Einkommen mit einer umfangreicheren Merkmalsliste anderer Datensätze verbunden werden kann.

4.1.4 Arbeitsunfallstatistik

4.1.4.1 Allgemeine Datensatzinformationen

Die Deutsche Gesetzliche Unfallversicherung e.V. (DGUV) stellt als Dachverband der Berufsgenossenschaften (BG) und der Unfallversicherungsträger der öffentlichen Hand (UVTöH) die jährliche Statistik des Unfallgeschehens am Arbeitsplatz zusammen, veröffentlicht die Ergebnisse (DGUV 2013) und übermittelt die Daten an das BMAS.

Grundlage der Statistik sind die Unfallanzeigen bei den BG und dem UVTöH, in denen alle meldepflichtigen Arbeitsunfälle¹⁷ dokumentiert werden. Aus dieser Datenbasis wird über das Geburtstagsverfahren eine Stichprobe gezogen. Aus den Unfallanzeigen der BG und der UVTöH werden sämtliche Arbeitsunfälle von Unfallverletzten, die an einem zehnten oder elften eines Monats Geburtstag haben, gezogen. Dies entspricht einer Stichprobengröße von etwa 6,6 %. Zusätzlich werden aus den Unfallanzeigen des UVTöH alle Arbeitsunfälle von Unfallverletzten gezogen, die am zwölften eines Monats Geburtstag haben. Dies erhöht die Stichprobengröße beim UVTöH auf etwa 9,9 %.

Da Arbeitnehmer/innen nur in der Unfallanzeige auftauchen, wenn ein Arbeitsunfall vorliegt handelt es sich im strengen Sinne nicht um einen Personen- sondern um einen Episodendatensatz, da Personen mit zwei Arbeitsunfällen in einem Jahr auch zweimal in der Statistik auftauchen, während Personen ohne einen Arbeitsunfall gar nicht vertreten sind. Dementsprechend handelt es sich bei der Arbeitsunfallstatistik um einen Querschnittsdatsatz.

4.1.4.2 Verfügbarkeit

Bisher wurde die Arbeitsunfallstatistik für die Forschung nicht genutzt und dies ist derzeit auch nicht vorgesehen. Da die DGUV die Daten im Auftrag des BMAS auswertet und anschließend dem BMAS übermittelt, liegt die Bereitstellung der Daten für Forschungszwecke nach Ansicht des DGUV beim BMAS.

4.1.4.3 Merkmale zur Mindestlohnevaluation

Die Arbeitsunfallstatistik enthält eine Reihe von soziodemografischen Merkmalen wie Alter, Geschlecht, Staatsangehörigkeit und der ausgeübte Beruf. Über die Arbeitgeber ist die Betriebsgrößenklasse und der Wirtschaftszweig, derzeit auf Zweistellerebene,¹⁸ bekannt. Zum

¹⁷ Ein Arbeitsunfall ist meldepflichtig, wenn der Unfallverletzte mindestens drei Tage völlig oder teilweise arbeitsunfähig ist oder es sich um einen Arbeitsunfall mit Todesfolge handelt.

¹⁸ Eine Aufschlüsselung auf Ebene der Viersteller ist nach Auskunft des DGUV in Arbeit.

Arbeitsunfall selbst sind die Unfallart, das -geschehen sowie Angaben zur Verletzung enthalten.

4.1.4.4 Merkmale zur Verknüpfung

Die DGUV bekommt von den BG und dem UVTöH die Daten anonymisiert geliefert. Allerdings sollten die Kontaktdaten von Unfallverletzten und zugehörigen Betrieben bei diesen gespeichert sein, was eine Verknüpfung über Record Linkage mit entsprechenden Datenquellen wie den IEB auf Individualebene bzw. dem BHP oder dem URS auf Betriebsebene möglich macht. Zudem ist eine Verknüpfung auf Zweistellerebene der Wirtschaftszweige möglich.

4.1.4.5 Identifikation des Evaluationspotenzials

Das Evaluationspotenzial der Arbeitsunfallstatistik allein genommen ist sehr stark eingeschränkt. Für die meisten Mindestlohnbranchen ist eine Unterteilung der Branchen weder auf der Zweisteller- noch auf der Vierstellerebene ausreichend, da sich diese nur über die Fünfstellerebene abgrenzen lassen. Lohnangaben sind nicht vorhanden.

Eine Verknüpfung der Arbeitsunfallstatistik mit den IEB würde allerdings eine Untersuchung der Auswirkungen des Mindestlohns auf die Arbeitsunfälle ermöglichen. Hier ist allerdings zu klären, ob dafür die Stichprobe der DGUV ausreicht oder ob auf die Vollerhebung der Unfallanzeigen bei den einzelnen BG zurückgegriffen werden sollte.

4.1.5 Daten der Lohnausgleichskasse für das Dachdeckerhandwerk (LAK)

4.1.5.1 Allgemeine Datensatzinformationen

Betriebe, die sich im Geltungsbereich des Tarifvertrages des Dachdeckerhandwerkes befinden, sind zu einer Teilnahme am Sozialkassenverfahren verpflichtet. Die zu entrichtenden Beiträge richten sich nach der Bruttolohnsumme. Daher sind alle Betriebe im räumlichen Geltungsbereich der Tarifverträge für das Dachdeckerhandwerk¹⁹ zu einer Meldung von Informationen zu ihren gewerblich Beschäftigten an die Lohnausgleichskasse des Dachdeckerhandwerks (LAK) verpflichtet. Aus diesem administrativen Prozess liegen Informationen zu allen gewerblich Beschäftigten im Dachdeckerhandwerk vor.²⁰ Nicht erfasst werden Dachdecker/innen, die in Deutschland arbeiten, aber bei einem Betrieb mit Betriebssitz im Ausland angestellt sind.

¹⁹ Dies entspricht allen Betrieben mit Betriebssitz in Deutschland.

²⁰ Dies gilt unter der Voraussetzung, dass alle Betriebe ordnungsgemäß ihre gewerblich Beschäftigten bei der Lohnausgleichskasse anmelden. Dies ist nach Auskunft der LAK nicht in allen Fällen gegeben. Dennoch sollte die Differenz vernachlässigbar klein sein, so dass von einer Vollerhebung gesprochen werden kann.

Die Grundgesamtheit umfasst alle gewerblichen Arbeitnehmer/innen von Betrieben des Dachdeckerhandwerks (im Sinne der Tarifverträge, d. h. mehr als 50 % der Arbeitszeit des Betriebes entfällt auf Bauleistungen i. S. dieser Tarifverträge). Diese Definition stimmt nicht vollständig mit der Klassifikation der Wirtschaftszweige des Statistischen Bundesamtes überein. Nach ZEW und infas (2011b, S. 93) wird der Geltungsbereich allerdings über den Wirtschaftszweig 45.22.1 (WZ03) bzw. 43.91.1 (WZ08) „Dachdeckerei und Bauspenglerei“ fast vollständig erfasst. Allerdings beinhaltet der Wirtschaftszweig ebenfalls die Bauspenglereien, die nicht im tariflichen Geltungsbereich des Dachdeckerhandwerks liegen. Diese machen allerdings nur 1,5 % der Betriebe in dem Wirtschaftszweig aus (ZEW und infas 2011b, S. 93f).

Jeweils zum 15. eines Monats sind die Betriebe zu einer Meldung ihrer Beschäftigten im Vormonat verpflichtet. Die Meldungen werden anschließend in die Datenbank eingepflegt und stehen ab dem Zeitpunkt zur Verfügung. Somit liegt der aktuelle Rand – im Unterschied zu anderen administrativen Datenquellen – lediglich etwa einen Monat in der Vergangenheit. Entsprechend der Meldepflichten der Betriebe liegen die Daten als Monatspanel vor. Auf Beschäftigtenebene sammelt die LAK seit 1984 Daten, allerdings sind die Daten vor 2000 archiviert, was einen Zugriff erschwert.

4.1.5.2 Verfügbarkeit

Die Daten werden von der LAK gepflegt und angeboten. Es gibt kein standardisiertes Vorgehen, wie die Daten für wissenschaftliche Zwecke zur Verfügung gestellt werden. Die Erfahrungen mit der letzten Evaluation des Mindestlohns (ZEW und infas 2011b) setzen im Bereitstellungsprozess der Daten eine sehr enge Zusammenarbeit mit der LAK voraus.

In einem Auftaktgespräch werden der Datenbedarf sowie die Voraussetzungen für den Datenzugang besprochen. Auf dieser Grundlage wird eine Datenschutzvereinbarung getroffen, die den Umfang, die Speicherung der Daten an der Forschungseinrichtung sowie die Form der Datenübertragung regelt. Die Daten werden ausschließlich anonymisiert zur Verfügung gestellt. Die Aufbereitung der Daten müssen dabei von der Datennutzerin oder dem -nutzer vorgenommen werden. Dies setzt umfangreiche IT-Kenntnisse voraus.

4.1.5.3 Merkmale zur Mindestlohnevaluation

Das Meldeverfahren beinhaltet eine Auflistung aller gewerblich Beschäftigten eines Betriebes mit Angaben zum Beginn der Beschäftigung sowie dem Zeitpunkt zu dem ein Arbeitsverhältnis gekündigt wurde, Stunden mit Lohnanspruch und der Bruttolohnsumme. Mit Hilfe dieser detaillierten Angaben sind Berechnungen der Stundenlöhne möglich, die dann als Grundlage für die Mindestlohnforschung genutzt werden können. Dabei ist allerdings zu berücksichtigen, dass die Bruttolohnsumme auch Zulagen, beispielsweise für Überstunden oder Erschwernis, und Sonderzahlungen wie Urlaubsgeld enthält. Der berechnete Stundenlohn ist dementsprechend ein Effektivstundenlohn, der aber den für den Mindestlohn relevanten Stundenlohn überschätzt.

4.1.5.4 Merkmale zur Verknüpfung

Die LAK identifiziert gewerblich Beschäftigte im Meldeverfahren anhand der durch die Betriebe angegebenen Sozialversicherungsnummer oder durch den Klarnamen sowie das Geburtsdatum. Die Betriebe sind über einen durch die LAK vergebenen Identifikator eindeutig identifizierbar. Beide Merkmale – Sozialversicherungsnummer und Betriebsinformationen – können technisch für eine Verknüpfung mit anderen Daten auf Individual- (IEB) oder Betriebsebene (BHP, URS) mit anderen Datenquellen verknüpft werden. Inwieweit dies tatsächlich durchgeführt werden kann, muss allerdings in einer gesonderten datenschutzrechtlichen Prüfung untersucht werden.

4.1.5.5 Identifikation des Evaluationspotenzials

Das Evaluationspotenzial der Daten isoliert betrachtet ist eingeschränkt, da nur wenige Merkmale vorliegen. Zusätzlich enthält der Datensatz ausschließlich Informationen zu Beschäftigten, die in den Geltungsbereich des Mindestlohns fallen, was einige Kontrollgruppendesigns unmöglich macht. Allerdings ermöglichen die Daten die Betroffenheit vom Mindestlohn und die Compliance der Betriebe im Dachdeckerhandwerk bezüglich eines Mindestlohns abzuschätzen. Gleichzeitig können die Daten für mögliche Imputationsverfahren zur Ergänzung der Arbeitszeitangaben in anderen Datensätzen verwendet werden.

In Verbindung mit anderen administrativen Daten wie etwa den IEB eröffnet der Datensatz eine deutliche Verbesserung des Evaluationspotenzials, da die Daten der LAK Angaben zur tatsächlichen Arbeitszeit und zu Bruttolöhnen auf Monatsebene enthalten. Inwieweit eine derartige Verknüpfung rechtlich möglich ist, ist aber zunächst zu klären.

4.1.6 Datensatz der Sozialkassen des Baugewerbes (SOKA-Bau)

4.1.6.1 Allgemeine Datensatzinformationen

Betriebe, die sich im Geltungsbereich des Tarifvertrages des Baugewerbes befinden, sind zu einer Teilnahme am Sozialkassenverfahren verpflichtet. Die zu entrichtenden Beiträge richten sich nach der Bruttolohnsumme. Daher sind inländische Betriebe zu einer Meldung von Beschäftigteninformationen verpflichtet. Aus diesem administrativen Prozess liegen Informationen zu allen Beschäftigten im inländischen Baugewerbe vor. Außerdem sind Betriebe im Rahmen des Arbeitnehmer-Entsendegesetzes zur Meldung von ausländischen Beschäftigten verpflichtet.

Die Grundgesamtheit umfasst alle Angestellten und gewerblichen Arbeitnehmer von Unternehmen der Bauwirtschaft (im Sinne der Tarifverträge, d. h. mehr als 50 % der Arbeitszeit des Unternehmens entfällt auf Bauleistungen i. S. dieser Tarifverträge). Diese Definition stimmt nicht mit der Klassifikation der Wirtschaftszweige des Statistischen Bundesamtes überein.

Der Datensatz ist eine Vollerhebung und beinhaltet im Inlandsverfahren im Jahr 2012 Angaben zu ca. 70.000 Betrieben mit 620.000 Beschäftigten und im Entsendeverfahren zu ca. 4.000 Betrieben mit 75.000 Arbeitnehmern.

4.1.6.2 Verfügbarkeit

Die Daten werden von der SOKA-Bau gepflegt und angeboten. Sie sind ausschließlich in anonymisierter Form zu beziehen. Die Beantragung erfolgt formlos nach Beauftragung durch das BMAS. Die Dauer der Beantragung liegt bei etwa 4 Wochen. Die Kosten belaufen sich je nach Datensatz auf ca. 4.000 Euro für die Entsendedaten und ca. 5.000 Euro für die Inlandsdaten.

Die Daten werden regelmäßig nach der Meldung der Betriebe gepflegt. Daher haben die Daten eine sehr hohe Aktualität. Auswertungen am aktuellen Rand sind in der Regel mit einer Zeitverschiebung von nur wenigen Monaten möglich. Individualangaben aus dem Entsendeverfahren sind seit 1997 vorhanden, aus dem inländischen Meldeverfahren seit 2000. Stundeninformationen werden seit Oktober 2007 erhoben. Aggregierte Angaben zu den Betrieben reichen weiter zurück.

4.1.6.3 Merkmale zur Mindestlohnevaluation

Das Meldeverfahren beinhaltet eine Auflistung aller Beschäftigten mit Angaben zur Tätigkeit, der Bruttolohnsumme der gewerblichen Mitarbeiter und deren Arbeitszeit in Stunden. Über eine intern vergebene Arbeitnehmernummer sind die Beschäftigten eindeutig in den Daten identifizierbar. Auf Betriebsebene liegen Informationen zu Betriebstätigkeit, Zahl der Angestellten und Bruttolohnsumme des Betriebs vor. Im Entsendeverfahren sind außerdem Angaben zu den Beschäftigungstagen des Arbeitnehmers, dem Herkunftsland und dem Ort der Baustelle (untergliedert nach West/Ost/Berlin) enthalten.

4.1.6.4 Merkmale zur Verknüpfung

Der Datensatz enthält keine Identifikatoren, mit denen eine Verknüpfung mit anderen Datenquellen möglich ist. Die Arbeitnehmernummer wird von der SOKA-Bau selbst vergeben und ist in anderen Datensätzen nicht vertreten. Die BA-Betriebsnummer wird erst seit 2013 für neu erfasste Betriebe erhoben. Über die vorliegenden Klarnamen sowie das Geburtsdatum und die Adresse ist ein Record Linkage mit anderen Datensätzen technisch möglich.

4.1.6.5 Identifikation des Evaluationspotenzials

Das Evaluationspotenzial der Daten isoliert betrachtet ist eingeschränkt, da nur wenige Merkmale vorliegen. Detaillierte Angaben reichen nicht bis zur Einführung des Mindestlohns im Bauhauptgewerbe zurück, so dass höchstens Erhöhungen untersucht werden können. Zusätzlich enthält der Datensatz ausschließlich Informationen zu Beschäftigten, die in den Geltungsbereich des Mindestlohns fallen, was einige Kontrollgruppendesigns unmöglich

macht. Allerdings ermöglichen die Daten die Betroffenheit vom Mindestlohn und die Compliance der Betriebe im Baugewerbe bezüglich eines Mindestlohns abzuschätzen.

In Verbindung mit anderen administrativen Daten wie etwa den IEB eröffnet der Datensatz eine deutliche Verbesserung des Evaluationspotenzials, da die Daten der SOKA-Bau Angaben zur tatsächlichen Arbeitszeit und zu Bruttolöhnen auf Monatsebene enthalten. Inwieweit eine Verknüpfung überhaupt möglich ist, ist aber zunächst zu klären.

4.1.7 Daten der Sozialkasse des Gerüstbaugewerbes und der Zusatzversorgungskasse des Gerüstbaugewerbes VVaG (SOKA-Gerüst)

4.1.7.1 Allgemeine Datensatzinformationen

Betriebe, die sich im Geltungsbereich des Tarifvertrages des Gerüstbauergewerbes befinden, sind zu einer Teilnahme am Sozialkassenverfahren verpflichtet. Die zu entrichtenden Beiträge richten sich nach der Bruttolohnsumme. Daher sind die Betriebe zu einer Meldung von Beschäftigteninformationen verpflichtet. Aus diesem administrativen Prozess liegen Informationen zu allen Beschäftigten von Betrieben mit Betriebssitz in Deutschland im Gerüstbauergewerbe vor.

Die Daten werden regelmäßig nach der Meldung der Betriebe gepflegt. Daher haben die Daten eine sehr hohe Aktualität. Allerdings muss berücksichtigt werden, dass die Meldungen der Betriebe nicht immer pünktlich erfolgen und somit am aktuellen Rand noch nicht zu allen Betrieben die Informationen vorliegen.

Das detaillierte Meldeverfahren liegt erst seit dem 1. Januar 2012 vor. Davor wurden die Daten lediglich pauschal erhoben. Dementsprechend liegen Mikrodaten erst ab 2012 vor.

4.1.7.2 Verfügbarkeit

Bisher hat die SOKA-Gerüst noch keine Erfahrungen in der Bereitstellung ihrer Daten zu Forschungszwecken gemacht. Aus diesen Gründen gibt es kein standardisiertes Vorgehen zur Beantragung und Nutzung der Daten. Nach Auskunft der SOKA-Gerüst ist eine grundsätzliche Bereitschaft zur Bereitstellung der Daten für die Mindestlohnforschung vorhanden. Voraussetzung ist eine Unterstützung des Anliegens durch das BMAS in Form eines Unterstützungsschreibens.

Da es bisher noch keine Erfahrungswerte gibt, ist die Dauer von der Beantragung bis zur Bereitstellung der Daten schwierig einzuschätzen. Ebenfalls liegen keine Informationen zu möglicherweise aufgrund von Zusatzaufbereitungen anfallenden Kosten vor. Aufgrund der grundsätzlichen Kooperationsbereitschaft der SOKA-Gerüst dürfte die Bereitstellung der Daten jedoch relativ zügig möglich sein.

4.1.7.3 Merkmale zur Mindestlohnevaluation

Das Meldeverfahren beinhaltet eine Auflistung aller Beschäftigten mit Angaben zum Geburtsdatum, Geschlecht und ob es gewerbliche oder kaufmännische Beschäftigte sind. Für gewerbliche Beschäftigte liegen weiterhin Informationen zur tatsächlichen Arbeitszeit und dem Arbeitsentgelt – sowohl als Bruttolohnsumme inklusive Lohnersatzleistungen als auch als Stundenlohn – vor. Über eine intern vergebene Arbeitnehmernummer sind die Beschäftigten eindeutig in den Daten identifizierbar.

4.1.7.4 Merkmale zur Verknüpfung

Der Datensatz enthält keine Identifikatoren, mit denen eine Verknüpfung mit anderen Datenquellen möglich ist. Die Arbeitnehmernummer wird von der SOKA-Gerüst selbst vergeben und ist in anderen Datensätzen nicht vertreten. Sehr selten liegt die Sozialversicherungsnummer vor, was keine verlässliche Verknüpfung ermöglicht. Über die vorliegenden Klarnamen sowie das Geburtsdatum und die Postleitzahl ist eine Verknüpfung mit den IEB technisch möglich. Offen ist allerdings, ob, und wenn ja, unter welchen Bedingungen die SOKA-Gerüst diese Informationen zur Verfügung stellt.

4.1.7.5 Identifikation des Evaluationspotenzials

Das Evaluationspotenzial der Daten isoliert betrachtet ist sehr stark eingeschränkt, da nur wenige Merkmale vorliegen. Zusätzlich enthält der Datensatz ausschließlich Informationen zu Beschäftigten, die in den Geltungsbereich des Mindestlohns fallen. In Verbindung mit den IEB eröffnet der Datensatz eine deutliche Verbesserung des Evaluationspotenzials der IEB, da die Daten der SOKA-Gerüst sowohl Angaben zur tatsächlichen Arbeitszeit als auch zu Stundenlöhnen enthalten. Beide Informationen sind für die Evaluation des Mindestlohns von großer Bedeutung, können aber in den IEB lediglich imputiert werden, da Stundenlöhne gar nicht und die Arbeitszeit lediglich sehr grob vorliegen.²¹

4.1.8 Auszubildendenstatistik

4.1.8.1 Allgemeine Datensatzinformationen

Die Auszubildendenstatistik des BIBB ist die offizielle Statistik des Ausbildungsgeschehens in den dualen Ausbildungsberufen nach dem Berufsbildungsgesetz (BBiG) und der Handwerksordnung (HwO). Ähnlich wie bei der Unfallstatistik (siehe Unterabschnitt 4.1.4) handelt es sich bei der Auszubildendenstatistik nicht um einen Personendatensatz sondern um einen Ereignisdatensatz. Als Ereignis gelten hier Vertragsabschlüsse und -auflösungen sowie Prüfungsdaten.

²¹ Durch die Umstellung des Meldeverfahrens, könnte sich dieses Problem in den IEB etwas entschärfen (siehe Unterabschnitt 4.1.1). Allerdings ist die Qualität der Umstellung noch nicht bekannt. Eine Verknüpfung könnte auch hier Plausibilitätschecks ermöglichen.

Die Daten werden von den Kammern an die Statistischen Landesämter gemeldet, welche an das BIBB zu Auswertungszwecken weitergegeben werden. Bis einschließlich 2006 erfolgte diese Weitergabe auf aggregierter Ebene. Seit 2007 liegen die Daten als Mikrodaten vor. Die Informationen werden zum Stichtag 31. Dezember eines Jahres erhoben. Nach einer Aufbereitungszeit von etwa elf Monaten werden sie durch das BIBB veröffentlicht. Derzeit reichen die Daten bis 2011 (Uhly et al. 2012).

4.1.8.2 Verfügbarkeit

Das BIBB stellt über das Datensystem Auszubildende (DAZUBI) aggregierte Auswertungen kostenlos zur Verfügung.²² Über das Projektteam DAZUBI können Sonderauswertungen durchgeführt werden. Aus datenschutzrechtlichen Gründen gibt es allerdings keine Möglichkeit der eigenen Auswertung der Daten.

4.1.8.3 Merkmale zur Mindestlohnevaluation

Die Auszubildendenstatistik kann anhand der üblichen soziodemografischen Merkmale Alter, Geschlecht, Staatsangehörigkeit und Schulabschluss ausgewertet werden. Als Ergebnisindikatoren dienen die Ereignisse, d.h. die Vertragsabschlüsse, die Vertragsauflösungen sowie Ergebnisse aus den Prüfungen.

4.1.8.4 Merkmale zur Verknüpfung

Da die Daten weder beim BIBB noch bei den Statistischen Ämtern mit einer Personnummer oder persönlichen Angaben vorliegen ist eine Verknüpfung weder über einen Identifikator noch über Record Linkage möglich. Über die Statistischen Ämter ist eine Verknüpfung der Auszubildendenstatistik auf Kreisebene möglich.

4.1.8.5 Identifikation des Evaluationspotenzials

Die Auszubildendenstatistik ist eine stark fachbezogene Statistik. Dies ermöglicht detaillierte deskriptive Auswertungen der Ausbildungssituation in Fachberufen, die im Geltungsbereich des Mindestlohns liegen. Für weitergehende Analysen wäre aber eine Verknüpfung der Daten auf Individual- oder Branchenebene notwendig, die technisch nicht möglich ist.

4.2 Amtliche Befragungsdaten

4.2.1 Verdienststrukturerhebung (VSE)

4.2.1.1 Allgemeine Datensatzinformationen

Die „Erhebung der Struktur der Arbeitsverdienste nach § 4 Verdienststatistikgesetz“, kurz Verdienststrukturerhebung (VSE), ist eine amtliche Befragung, die seit 2006 dezentral von

²² Homepage: www.bibb.de/dazubi [abgerufen am: 14.11.2013].

den statistischen Ämtern der Länder alle vier Jahre durchgeführt wird (Statistisches Bundesamt 2013d). Sie ist das Nachfolgeprodukt der Gehalts- und Lohnstrukturerhebung (GLS), die in unregelmäßigen Abständen von 1990 bis 2001 erhoben wurde.²³

Aufgrund der Umstellung der Statistik zwischen 2001 und 2006 ist eine Vergleichbarkeit der Daten ab 2006 mit vorangegangenen Erhebungsjahren nur eingeschränkt möglich. Insbesondere die Tatsache, dass die Daten in Ost- und Westdeutschland nicht immer zeitgleich erhoben wurden (IAW 2011b, S. 31f) und eine unterschiedliche Abdeckung der Wirtschaftszweige vorliegt (Statistisches Bundesamt 2013d), erschwert die Vergleichbarkeit.

Grundgesamtheit der VSE sind Beschäftigte in Betrieben mit zehn oder mehr sozialversicherungspflichtig Beschäftigten des produzierenden und Dienstleistungsgewerbes, was den Abschnitten B bis S der Klassifikation der Wirtschaftszweige 2008 (WZ 2008) entspricht. Als Ziehungsgrundlage dient das Unternehmensregister, welches im Abschnitt 5.1.3 näher beschrieben wird. Entsprechend der bereits beschriebenen Grundgesamtheit wurden vor der Stichprobenziehung sämtliche Betriebe mit weniger als zehn sozialversicherungspflichtig Beschäftigten aus der Ziehungsgrundlage entfernt.

Das tatsächliche Auswahlverfahren wurde danach in zwei Stufen vorgenommen. Die Schichtung berücksichtigte die 16 Bundesländer, die 81 Abteilungen der Wirtschaftszweigklassifikation sowie eine Aufteilung der Betriebe in sechs Größenklassen. Für die VSE 2010 betrug der Auswahlatz 9,0 % der Grundgesamtheit, was 33.333 Betrieben entspricht. Der Auswahlatz unterscheidet sich allerdings deutlich zwischen den Größenklassen der Betriebe. So werden alle Betriebe mit 1.000 und mehr Beschäftigten erfasst.

In Abhängigkeit von den Größenklassen wurde ein Auswahlatz für die Ziehung der Beschäftigten definiert. Bei Betrieben mit zehn bis unter 50 sozialversicherungspflichtig Beschäftigten liegt dieser bei 100 %, während in Betrieben mit 1.000 und mehr Beschäftigten 6,67 % der sozialversicherungspflichtigen Beschäftigten befragt wurden. Die Auswahl auf der zweiten Stufe wurde entweder durch die Betriebe selbst oder auf Wunsch der Betriebe durch das zuständige statistische Landesamt durchgeführt. Insgesamt wurde 2010 ein Stichprobenumfang von 1,6 Mio. sozialversicherungspflichtig Beschäftigten realisiert.²⁴

4.2.1.2 Verfügbarkeit

Sowohl die VSE als auch der Vorgänger die GLS werden über die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder angeboten. Daher folgt die Beantragung gemäß den im Unterabschnitt 3.2 beschriebenen Voraussetzungen. Derzeit sind zwei Wellen der VSE (2006 und 2010) sowie vier Wellen der GLS (1990, 1992, 1995 und 2001) sowohl

²³ Insgesamt gibt es vier Wellen der GLS: 1990, 1992, 1995 und 2001.

²⁴ Die Stichprobenziehung erfolgte in der Wirtschaftsabteilung O84 und in den Wirtschaftsgruppen P85.1, P85.2, P85.3 und P85.4 der WZ 2008 auf Basis der Personalstandstatistik (Statistisches Bundesamt 2013f). Da es sich dabei um Beschäftigte des öffentlichen Dienstes handelt, ist eine Beschreibung der Stichprobenziehung hier irrelevant.

über Datenfernverarbeitung als auch über einen Gastwissenschaftlerarbeitsplatz verfügbar. Zudem sind alle Datensätze ab 2001 zusätzlich als Public-Use-File und als Scientific-Use-File erhältlich. Eine neue Erhebung der VSE wird 2014 durchgeführt und wird etwa 19 Monate später (Juli 2016) veröffentlicht.

4.2.1.3 Merkmale zur Mindestlohnevaluation

Die VSE enthält diverse Merkmale, die für die Mindestlohnforschung von Interesse sind. Da es sich bei dem Datensatz um einen Linked-Employer-Employee-Datensatz handelt, sind sowohl Daten auf Betriebs- als auch auf Individualebene verfügbar. Im Folgenden werden die wesentlichen Merkmale beschrieben. Da für die Off-Site-Produkte²⁵ höhere Datenschutzbestimmungen gelten und dementsprechend diese stärker anonymisiert sind, beziehen sich folgende Angaben auf die On-Site-Nutzung²⁶ um das komplette Analysepotenzial der VSE beleuchten zu können.

Auf Betriebsebene sind neben dem Regionalschlüssel bis auf Gemeindeebene und dem fünfstelligen Wirtschaftszweig vor allem Angaben zur Anzahl der Beschäftigten sowohl auf Unternehmens- als auch auf Betriebsebene von Interesse. Für tarifgebundene Betriebe sind bis zu fünf Tarifverträge anhand eines elfstelligen Tarifvertragsschlüssels dokumentiert. Der Tarifvertragsschlüssel identifiziert die Beschäftigtengruppe, die Art der Tarifregelung, den Wirtschaftszweig, die laufende Nummer der Eingliederungsvereinbarung des jeweiligen Landes sowie die Art des Tarifvertrags (FDZ der Statistischen Ämter der Länder 2013a, S. 25).²⁷ 2010 wurde die Tarifbindung genauer erfasst, was eine Unterscheidung zwischen einer tatsächlichen Tarifbindung über Branchen- bzw. Firmentarifverträgen sowie einer Orientierung oder Anlehnung an Tarifverträgen differenziert. Seit 2010 ist zudem erfasst, ob der Betrieb einem Mindestlohn nach dem Arbeitnehmer-Entsendegesetz (AEntG) unterliegt. Da dieses Merkmal 2010 das erste Mal erfasst wurde ist die Qualität des Merkmals zu überprüfen, um eine Aussage über die Verwendbarkeit tätigen zu können. Dies ist allerdings nicht Bestandteil dieses Gutachtens, da die Autoren davon ausgehen, dass dies in dem vom BMAS an das Statistische Bundesamt vergeben Gutachten behandelt wird.

Auf Individualebene sind insbesondere die detaillierten Angaben zur Vergütung und zur Arbeitszeit für die Mindestlohnevaluation von Interesse. So sind neben der vertraglichen Arbeitszeit auch die bezahlten Überstunden aufgeführt. Die Verdienste werden sowohl auf Monats- und Jahresbasis als Bruttoverdienste aufgeführt als auch auf Stundenbasis als Bruttostundenverdienst. Zudem existieren detaillierte Informationen zu Zulagen und Abzügen sowie, falls vorhanden, zur tariflichen Lohngruppe.

²⁵ Public- und Scientific-Use-Files.

²⁶ Datenfernverarbeitung und Gastwissenschaftlerarbeitsplatz.

²⁷ Da dieses Merkmal einerseits sehr aufwändig in der Aufbereitung ist und zudem eine Identifizierung einzelner Unternehmen und Betriebe sehr leicht ermöglicht, sind die Hürden der Beantragung sehr hoch.

4.2.1.4 Merkmale zur Verknüpfung

Die Stichprobe der VSE wurde aus dem Unternehmensregister – System 95 (URS) gezogen. Dadurch ist eine Verknüpfung der VSE mit dem URS technisch möglich. Da das URS eine Reihe weiterer Identifikationsmerkmale enthält, ist die VSE über den Umweg des URS mit einer Reihe weiterer Datensätze verknüpfbar, die detaillierter im Unterabschnitt 5.1.3 beschrieben werden.

Neben einer Verknüpfung über die Identifikationsnummer des URS ermöglichen der Regionalschlüssel, der eine Differenzierung bis auf Gemeindeebene ermöglicht, sowie der Wirtschaftszweig eine Verknüpfung der Daten auf diesen aggregierten Ebenen.

Auf Individualebene ist keine direkte Verknüpfung möglich. Über die Angaben zur ausgeübten Tätigkeit, der Stellung im Beruf sowie der Ausbildung aus dem Sozialversicherungsmeldev erfahren ist eine Verknüpfung anhand dieser Angaben möglich. Eine weitere Differenzierung nach Geschlecht und Geburtsjahr ist ebenfalls technisch möglich. Diese Merkmale ermöglichen es technisch, Informationen aus den VSE per Imputationsverfahren anderen Individualdatensätzen wie beispielsweise den IEB hinzuzuspielen.

4.2.1.5 Identifikation des Evaluationspotenzials

Die VSE bietet einige Vorteile für die Mindestlohnforschung gegenüber anderen Datensätzen, allerdings auch einige Nachteile. Aufgrund des sehr stark auf die Erhebung von Verdiensten eingeschränkten Untersuchungsgegenstands, beschränkt sich die Verwendbarkeit der VSE sehr stark auf quantitative Analysen mit dem Arbeitsverdienst oder der Arbeitszeit als Ergebnisindikatoren. Unter anderem Rattenhuber (2011) hat anhand des Bauhauptgewerbes die Auswirkungen des Mindestlohns mit der VSE untersucht.

Trotz dieser eingeschränkten Nutzungsmöglichkeit für Wirkungsanalysen, bietet die VSE umfangreiches Analysepotenzial für vorbereitende Schritte. Im Rahmen der letzten Evaluationsrunde wurden in einigen Branchen die ausführlichen Tarifvertragsinformationen genutzt, um den Geltungsbereich des Mindestlohns anhand der Wirtschaftszweige trennschärfer abgrenzen zu können (IAW 2011a, 2011b & 2011c). Dies ist mit der verbesserten Erfassung der Tarifbindung sowie der Erfassung des Geltungsbereichs des Mindestlohns ab 2010 einfacher möglich.

Aufgrund der Stichprobengröße ist die VSE auch für relativ feingliedrige Imputationsverfahren geeignet, da rund 1,6 Mio. sozialversicherungspflichtig Beschäftigte zur Verfügung stehen. Dadurch ist die Wahrscheinlichkeit gegeben, dass auch in Branchen mit relativ wenigen Beschäftigten eine Imputation möglich ist.

Neben den genannten Vorteilen, hat die VSE aber auch einen entscheidenden Nachteil mit Auswirkungen auf die Repräsentativität der Ergebnisse. Da ausschließlich Betriebe mit zehn oder mehr sozialversicherungspflichtig Beschäftigten in der VSE vertreten sind, können sämtliche durchgeführte Analysen nicht auf Kleinstbetriebe mit weniger als zehn sozialversi-

cherungspflichtig Beschäftigten übertragen werden. Dies ist insbesondere in Handwerksberufen problematisch, da der überwiegende Teil der Branche aus Kleinst- und Kleinbetrieben besteht.²⁸

Insgesamt bietet die VSE aufgrund der abgefragten Merkmale sowie der Stichprobengröße ein gutes Analysepotenzial, was sich allerdings inhaltlich sehr stark auf Fragestellungen zu Verdienst und Arbeitszeit beschränkt. Zudem ist sie durch das Fehlen von Kleinstbetrieben für viele Mindestlohnbranchen aufgrund fehlender Repräsentativität nur sehr eingeschränkt oder überhaupt nicht verwendbar.

4.2.2 Mikrozensus

4.2.2.1 Allgemeine Datensatzinformationen

Der Mikrozensus ist eine amtliche Befragung, die seit 1957 jährlich von den Statistischen Ämtern der Länder durchgeführt wird (Statistisches Bundesamt 2013b). Rechtsgrundlage ist insbesondere das Mikrozensusgesetz. Die Umfrage liefert wichtige Strukturdaten über die Bevölkerung, und umfasst unter anderem Fragen zum Familien- und Haushaltszusammenhang, zur Gesundheit und den Wohnverhältnissen sowie zur Erwerbstätigkeit, zum Einkommen und zur schulischen und beruflichen Ausbildung.

Grundgesamtheit des Mikrozensus bildet die gesamte Wohnbevölkerung in Deutschland mit Ausnahme der Angehörigen ausländischer Streitkräfte oder diplomatischer Vertretungen. Die Erhebungseinheiten sind Personen, Haushalte und Wohnungen. Für einen Großteil der erhobenen Merkmale besteht Auskunftspflicht.

Die Befragung findet bei einer einprozentigen, geklumpten Zufallsstichprobe der Haushalte in Deutschland statt. Dazu werden mit Hilfe von Ergebnissen aus der Volkszählung und der Bautätigkeitsstatistik Wohnungen entsprechend ihrer Größenklasse in Auswahlbezirke zusammengefasst. Regionale Repräsentativität wird durch eine Schichtung auf Basis von insgesamt 201 Raumeinheiten sichergestellt. Jährlich wird ein Viertel der Auswahlbezirke ersetzt, so dass jeder befragte Haushalt insgesamt vier Jahre in Folge an der Umfrage teilnimmt. Der Berichtszeitraum ist seit 2005 gleitend und gleichmäßig über das gesamte Jahr verteilt. Die Befragung der Haushalte erfolgt mittels Computer Assisted Personal Interviews (CAPI), schriftlich oder telefonisch. Proxy-Interviews, bei denen ein erwachsenes Haushaltsmitglied für andere antwortet, sind zulässig.

Die Vergleichbarkeit der Daten ist generell hoch. Auf räumlicher Ebene ergibt sich aber 2005 ein Bruch mit der Hinzurechnung von Westberlin zu den neuen Ländern statt zum früheren Bundesgebiet. Ebenfalls 2005 erfolgten die Umstellung von einer festen Berichtswoche im Frühjahr zur kontinuierlichen Erhebung sowie eine Überarbeitung der Fragen zum Erwerbs-

²⁸ In einigen Mindestlohnbranchen liegt der Anteil der Betriebe mit mindestens einem und maximal neun sozialversicherungspflichtig Beschäftigten bei etwa zwei Dritteln (bspw. Maler und Lackierer (IAW 2011b, S. 83)).

leben. Diese und weitere Änderungen müssen bei Vergleichen über die Zeit beachtet werden (Statistisches Bundesamt 2013b).

4.2.2.2 Verfügbarkeit

Der Mikrozensus wird über die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder angeboten. Daher folgt die Beantragung gemäß den im Unterabschnitt 3.2 beschriebenen Voraussetzungen. Da der Mikrozensus mit einer umfangreichen Version als Standarddatenprodukt vorliegt, ist die Dauer der Beantragung relativ kurz. Wenn Merkmale benötigt werden, die entweder nicht Bestandteil des Standardprodukts sind oder in einer detaillierteren Form benötigt werden, kann der Prozess der Beantragung deutlich länger dauern, da dann ein individuelles Anonymisierungskonzept erstellt und mit den 14 Statistischen Landesämtern abgestimmt werden muss.

Derzeit sind 32 Wellen des Mikrozensus zwischen 1973 und 2011 sowohl über Datenfernverarbeitung als auch über einen Gastwissenschaftlerarbeitsplatz verfügbar. Seit 1985 liegen sie jährlich vor. Zudem ist eine Auswahl der Datensätze zusätzlich als Public-Use-File und als Scientific-Use-File erhältlich. Erste Ergebnisse zum Berichtsjahr stehen in der Regel im Juli des Folgejahres zur Verfügung.

4.2.2.3 Merkmale zur Mindestlohnevaluation

Der Mikrozensus fragt nach einer Reihe von Merkmalen, die für die Mindestlohnforschung von Bedeutung sind. Hierzu zählen die aktuelle und durchschnittliche Wochenarbeitszeit, die gewünschte Arbeitszeit und detaillierte Fragen zum aktuellen oder letzten Arbeitsplatz, der Beschäftigungsart und zur Arbeitssuche. Außerdem wird nach dem Nettoeinkommen der Person und des Haushalts gefragt. Bildungsabschlüsse und berufliche Weiterbildung sind ebenfalls Teil des Fragenkanons.

4.2.2.4 Merkmale zur Verknüpfung

Auf der Individual- bzw. Haushaltsebene ist keine Verknüpfung mit anderen Datenprodukten möglich, da das Adressmaterial aus dem Mikrozensus nicht aufbewahrt wird, nachdem ein Haushalt aus der Rotation fällt.

Da eine Reihe von Merkmalen wie beispielsweise die Regionalangabe, der Wirtschaftszweig, das Alter und die Tätigkeit nur relativ grob abgefragt werden, sind die Möglichkeiten für Imputationsverfahren bzw. die Gewinnung von Aggregaten zur Zuspierung zu anderen Datensätzen begrenzt und müssen von Fall zu Fall geprüft werden.

4.2.2.5 Identifikation des Evaluationspotenzials

Das Evaluationspotenzial des Mikrozensus ist in allererster Linie dadurch eingeschränkt, dass die Haushaltsmitglieder zwar nach ihrem gesamten Nettoeinkommen, nicht aber nach ihrem Bruttolohn oder -gehalt aus einzelner Beschäftigung gefragt werden. Zudem liegt das

Nettoeinkommen ausschließlich in sehr groben Einkommensklassen vor, was eine Berechnung von Stundenlöhnen – trotz der genauen Fragen zur Arbeitszeit – nahezu unmöglich macht. Damit ist die Eingrenzung der Mindestlohn Betroffenheit kaum möglich. Da auch die Frage nach dem Wirtschaftszweig sehr grob ist, ist der Mikrozensus auch nicht für Kontrollgruppenansätze auf Branchenebene geeignet. Regionalansätze sind problematisch, da der Mikrozensus bei kleiner regionaler Einteilung nicht mehr repräsentativ ist. Aufgrund der mangelnden Zuspieldmöglichkeiten können diese Probleme auch nicht umgangen werden. Insgesamt ist der Mikrozensus in der Mindestlohnforschung also lediglich für Hilfsanalysen geeignet. Bei der Evaluation im Bauhauptgewerbe wurde er etwa genutzt, um wöchentliche Arbeitszeiten zu approximieren.

4.3 Nichtamtliche Daten

4.3.1 Beschäftigtenbefragung Bauhauptgewerbe der Forschungsgemeinschaft aus IAB, RWI und ISG

4.3.1.1 Allgemeine Datensatzinformationen

Im Rahmen der Evaluation bestehender gesetzlicher Mindestlohnregelungen wurde für das Bauhauptgewerbe im Auftrag des Forschungskonsortiums aus ISG, RWI und IAB eine Arbeitnehmerbefragung durchgeführt. Die Befragung fand zwischen dem 15. Februar und dem 30. März 2011 statt.

Die Grundlage für die Stichprobe bildeten die Betriebe im Bauhauptgewerbe, die 2009 im IAB-Betriebspanel befragt wurden und unter die Mindestlohnregelung im Baugewerbe fielen. Innerhalb dieser Gruppe wurden vollzeitbeschäftigte Arbeiter, die zum 31. Dezember 2009 beschäftigt waren, nach einem speziellen Schichtungsverfahren aus den IEB des IAB gezogen. Angestellte, Auszubildende, Rentner und Beschäftigte in Altersteilzeit wurden von der Befragung ausgeschlossen.

Das Auswahlverfahren wurde so angelegt, dass die im Betriebspanel vorherrschende Unterrepräsentation von Kleinstbetrieben in der Befragung etwas ausgeglichen wurde. Die Auswahlwahrscheinlichkeit war bei Betrieben mit einem sozialversicherungspflichtig Beschäftigten 100 Prozent, bei zwei Beschäftigten 50 Prozent, bei drei 33 Prozent, usw. Die Wahrscheinlichkeit in die Stichprobe zu gelangen war aber – außer bei Großbetrieben mit über 250 Beschäftigten – nie kleiner als 10 Prozent. Am Ende wurden Adressen von 6.139 Beschäftigten im Bauhauptgewerbe an das Befragungsinstitut geliefert.

Die Befragung wurde vom SOKO Institut Bielefeld als computerunterstützte telefonische Befragung (CATI) durchgeführt. Die Einwilligung zur Verknüpfung mit Prozessdaten der BA wurde telefonisch eingeholt. Insgesamt wurden 1.521 telefonische Interviews realisiert. Von den befragten Personen stimmten 1.399 einer Zuspieldung der administrativen Daten zur Befragung zu. Dies entspricht einer Zustimmungquote von 92 Prozent.

4.3.1.2 Verfügbarkeit

Der Datensatz wurde im Rahmen eines Forschungsprojekts erstellt. Bislang gibt es noch keinen standardisierten Datenzugang bzw. eine Datendokumentation für externe Nutzer. Bei Interesse sollte daher das FDZ der BA im IAB kontaktiert werden, um die Möglichkeiten einer Nutzung abzustimmen. Weitere Details zum Datensatz sind im Projektbericht (IAB/RWI/ISG, 2011) zu finden.

4.3.1.3 Merkmale zur Mindestlohnevaluation

Da der Datensatz explizit für die Mindestlohnforschung entwickelt wurde, beinhaltet er eine Vielzahl von Fragen zu Themenkomplexen, die mit dem Mindestlohn im Bauhauptgewerbe zusammenhängen oder für eine Evaluation weiterhin bedeutend sind. So wurden beispielsweise direkte Fragen nach der Relevanz des Mindestlohns im eigenen Betrieb oder für einen selbst gestellt. Auch nach der persönlichen Meinung zur Wirkung von Mindestlöhnen wurde gefragt. Neben Angaben zum Bruttostundenlohn und Zuschlägen wurde mit einem Kalenderium die Beschäftigungssituation zwischen Dezember 2009 und dem Befragungszeitpunkt im Frühjahr 2011 genau erfasst. Außerdem wurden unter anderem Fragen zur Stellung im Beruf, Befristung, Arbeitszeitregelungen und Arbeitszeit, Gründen der Beendigung von Beschäftigung und dem beschäftigenden Betrieb gestellt.

Weitere Themenkomplexe umfassten Arbeitszufriedenheit, Weiterbildung, Tarifbindung, Einkommen, Soziodemografie und Schwarzarbeit.

4.3.1.4 Merkmale zur Verknüpfung

Da während der Befragung um Zustimmung zur Zuspiegelung von Prozessdaten gebeten wurde, können für diejenigen Befragten, die ihre Einwilligung erteilt haben, Daten aus IEB, BHP und IAB-BP verknüpft werden.

4.3.1.5 Identifikation des Evaluationspotenzials

Die Beschäftigungsbefragung im Bauhauptgewerbe kann als das bislang umfassendste Stimmungsbild zum Mindestlohn im Bauhauptgewerbe angesehen werden. Ihr Vorteil besteht in dem breiten Fragenkatalog mit vielen mindestlohnrelevanten Fragestellungen und der Stichprobenziehung, die einen Fokus auf die tatsächlich betroffenen Beschäftigten zulässt. Die Verknüpfung mit BA-Prozessdaten erlaubt den Aufbau einer sehr umfangreichen Datenbasis.

Von Nachteil ist der Umstand, dass die Befragung nicht für Wiederholungen ausgelegt wurde und bislang keine weitere Welle geplant ist. Analysemethoden, die auf Paneldaten oder wiederholten Querschnitten aufbauen, können daher nicht angewandt werden. Zu vielen Fragestellungen, die mit den Daten zu untersuchen sind, wurden bereits Ergebnisse im Rahmen des Projektberichts veröffentlicht.

4.3.2 Linked Personnel Panel Beschäftigtenbefragung (LPP Beschäftigte)

4.3.2.1 Allgemeine Datensatzinformationen

Im Rahmen des Forschungsprojekts „Arbeitsqualität und wirtschaftlicher Erfolg“ wird derzeit das Linked Personnel Panel (LPP) aufgebaut. Der Datensatz entsteht in Kooperation zwischen IAB, ZEW, BMAS und der Universität Köln. Ziel des Projekts ist der Aufbau eines Linked Employer-Employee Datensatzes, mit dem sich die Auswirkung des Personalmanagements auf den Unternehmenserfolg sowie Arbeitszufriedenheit, Gesundheit und Unternehmensbindung der Mitarbeiter untersuchen lassen. Der Datensatz ist als Panel geplant und besteht aus einem Betriebsbefragungsteil (LPP Betriebe) und einem Beschäftigtenbefragungsteil (LPP Beschäftigte). In diesem Abschnitt wird die Beschäftigtenbefragung besprochen.

Ausgangspunkt für die Beschäftigtenbefragung ist wie beim IAB-BP die Betriebsdatei der BA (siehe Abschnitt 5.4.1) bzw. die in diesen Betrieben sozialversicherungspflichtig Beschäftigten. In die Auswahl gelangen aber nur Beschäftigte, die zum 31.12. des Vorjahres in einem Betrieb beschäftigt waren, der im Rahmen des LPP Betriebe befragt wurde (siehe Abschnitt 5.4.2). Als weitere Einschränkung der Grundgesamtheit werden nur die Beschäftigten berücksichtigt, die in panelbereiten Betrieben mit (laut den administrativen Daten) mindestens 50 sozialversicherungspflichtigen Beschäftigten arbeiten. Auszubildende, Mitarbeiter in Elternzeit oder Altersteilzeit werden nicht berücksichtigt.

Der Datensatz basiert auf einer nach vier Betriebsgrößenklassen geschichteten Stichprobe. Um repräsentative Aussagen über die Grundgesamtheit treffen zu können, müssen daher bei deskriptiven Auswertungen entsprechende Gewichtungsfaktoren verwendet werden. Diese sollen sowohl die Disproportionalität der Stichprobe für das LPP als auch Antwortausfälle während der Befragung ausgleichen. Die Ergebnisse der Befragung sind repräsentativ für sozialversicherungspflichtig Beschäftigte der befragten Betriebe. Über andere Beschäftigungsgruppen und andere Betriebe können keine validen Aussagen getroffen werden.

Die Befragung findet statt, nachdem die Befragung für das LPP Betriebe abgeschlossen ist. Sie wird unabhängig von der Betriebsbefragung vom Institut für angewandte Sozialwissenschaft GmbH (infas) durchgeführt und erfolgt telefonisch.

In der ersten Welle der Befragung wurde bei einer Bruttostichprobe von 30.703 Personen ein Rücklauf von 7.508 Personen aus 861 Betrieben realisiert. Die zweite Welle ist für 2014 geplant.

4.3.2.2 Verfügbarkeit

Das LPP wird derzeit aufgebaut und ist bislang noch nicht für externe Forscher zugänglich. Später soll der Datensatz durch das FDZ der BA im IAB angeboten werden. Ein Veröffentlichungstermin steht noch nicht fest.

4.3.2.3 Merkmale zur Mindestlohnevaluation

Das LPP Beschäftigte enthält unter anderem Angaben der Beschäftigten zu Arbeitsbedingungen und Vergütung, Personalentwicklung, Unternehmenskultur und Gesundheit. Für die Mindestlohnevaluation ist dabei vor allem hilfreich, dass über die Fragen zur Entlohnung und Arbeitszeit ein Stundenlohn berechnet werden kann.

Einzelne Themen betreffen zum Beispiel Fragen zu geleisteten Überstunden, beruflicher Weiterbildung, Arbeitsbelastung, Bindung an den Betrieb und Arbeitszufriedenheit und dem Gesundheitszustand.

4.3.2.4 Merkmale zur Verknüpfung

Das LPP Beschäftigte enthält Identifikatoren, die sowohl eine Verknüpfung mit dem LPP Betriebe als auch mit den anderen IAB-Datenprodukten, also dem IAB-BP, der IEB und dem BHP zulässt, sofern die Einwilligung der befragten Personen vorliegt. Dadurch soll am Ende ein Linked Employer-Employee Datensatz entstehen.

4.3.2.5 Identifikation des Evaluationspotenzials

Für sich allein genommen ist das Evaluationspotential des LPP Beschäftigte eingeschränkt. Zwar kann ein Zusammenhang zwischen dem Stundenlohn und den Dimensionen der Befragung hergestellt werden, andererseits fehlen viele relevante Informationen zum Entscheidungsverhalten und der Entwicklung der Betriebe als auch zu den Erwerbshistoriken der Beschäftigten. Die Kombination mit anderen IAB-Produkten zum LPP ergibt deutlich verbesserte Analysemöglichkeiten. Somit ist die Untersuchung der Zusammenhänge zwischen Mindestlöhnen und Arbeitsqualität bzw. Unternehmenskultur möglich.

Einschränkend ist anzumerken, dass sich das Panel erst im Aufbau befindet und noch nicht genutzt werden kann. Außerdem fand die erste Welle im Jahr 2012 statt, so dass bisherige Mindestlohneinführungen bzw. -erhöhungen nicht untersucht werden können. Für Branchenanalysen ist zu beachten, dass die sehr grobe Schichtungsmatrix hier kaum repräsentative Untersuchungen zulässt. Da bisher keine Finanzierung über die zweite Welle hinaus gesichert ist, ist das Langfristpotenzial der Daten ungewiss.

4.3.3 Panel Arbeitsmarkt und soziale Sicherung (PASS)

4.3.3.1 Allgemeine Datensatzinformationen

Das Panel Arbeitsmarkt und soziale Sicherung (PASS) ist eine jährliche Haushaltsbefragung, die seit 2006 im Auftrag des IAB durchgeführt wird. PASS ist primär konzipiert, um Forschung im Überschneidungsbereich zwischen Beschäftigung und Grundsicherungs- bzw. Sozialleistungsbezug durchzuführen, kann aber aufgrund seines umfangreichen Fragenkatalogs auch für andere Themengebiete verwendet werden.

Der Datensatz besteht aus zwei Teilpopulationen. Die Grundgesamtheit des ersten Teils bilden Haushalte, in denen mindestens eine Person Leistungen nach dem SGB II bezieht. Grundgesamtheit des zweiten Teils sind alle Privathaushalte in Deutschland. Durch dieses Design wird die Untersuchung von Übergängen in den bzw. aus dem Leistungsbezug heraus möglich.

Die Stichprobenziehung erfolgte in zwei Stufen. Zunächst wurden 300 Postleitzahlenbezirke zufällig ausgewählt, wobei die Auswahlwahrscheinlichkeit proportional zur Einwohnerzahl gesetzt wurde. Im zweiten Schritt wurden innerhalb dieser Auswahlbezirke Bedarfsgemeinschaften aus den Prozessdaten der BA und eine Gebäuestichprobe aus der Datenbank eines kommerziellen Anbieters gezogen. Die Größe der Bruttostichprobe der Bedarfsgemeinschaften war dabei vom Leistungsempfängeranteil abhängig. Die Haushalte aus der gesamten Wohnbevölkerung wurden geschichtet gezogen, so dass Haushalte mit niedrigem Status eine höhere Auswahlwahrscheinlichkeit hatten. In jedem Haushalt wurde zunächst ein Interview mit dem Haushaltsvorstand durchgeführt. Danach sollten, wenn möglich, weitere Interviews mit allen anderen Haushaltsmitgliedern geführt werden, die das 15. Lebensjahr vollendet hatten.

Die Befragung findet jährlich statt und wird seit der vierten Welle zwischen Februar und September von infas durchgeführt. Davor fand sie zwischen Dezember und Juli/August statt. Das Befragungsinstitut war bis dahin TNS Infratest Sozialforschung.

Zu Beginn des Panels wurden aus beiden Teilpopulationen etwa 10.000 Personen befragt. Dem Rückgang an Befragten durch das Ausscheiden von Bedarfsgemeinschaften aus der ersten Teilstichprobe wurde jährlich durch die Ziehung einer zusätzlichen Zugangsstichprobe entgegengewirkt. In der fünften Welle wurde auch die zweite Teilstichprobe aufgestockt. Hier konnten knapp 15.600 Personen in über 10.000 Haushalten befragt werden, darunter etwa 9.600 Personen, mit denen bereits zum wiederholten Mal ein Interview im Rahmen von PASS geführt wurde.

4.3.3.2 Verfügbarkeit

Das PASS wird durch das FDZ der BA im IAB angeboten. Innerhalb des IAB ist der Forschungsbereich Panel Arbeitsmarkt und soziale Sicherung für die Erstellung, Aktualisierung und Dokumentation zuständig. Die Nutzerbetreuung übernimmt das FDZ. Der Datensatz ist als SUF verfügbar. Die Beantragung folgt gemäß den im Unterabschnitt 3.1 beschriebenen Voraussetzungen.

Nach Eingang eines genehmigungsfähigen Datenantrags erfolgt die Freischaltung zur Datennutzung in der Regel innerhalb weniger Tage. Sicherheitshalber sollten 2 Wochen veranschlagt werden. Die Datennutzung selbst ist kostenlos.

Das PASS wird jedes Jahr aktualisiert und ist derzeit für den Zeitraum 2006/07 bis 2011 verfügbar. Das nächste Update ist für Ende 2013 bzw. Anfang 2014 geplant. Der abgedeckte

Zeitraum soll dann bis 2012 erweitert werden. Umfangreiche Dokumentationen und Auszählungen sind auf der Internetpräsenz des FDZ verfügbar.

4.3.3.3 Merkmale zur Mindestlohnevaluation

Das PASS enthält einen umfangreichen Fragenkatalog zu soziodemographischen Merkmalen, beschäftigungsbezogenen Merkmalen, Leistungsbezug, aber auch zu subjektiven Einschätzungen. Für die Mindestlohnforschung ist hilfreich, dass neben dem Bruttoarbeitseinkommen auch tatsächliche und vertraglich vereinbarte Wochenarbeitszeiten abgefragt werden. Außerdem werden Informationen zur Art der Tätigkeit und dem Wirtschaftszweig, dem Arbeitgeber und der Arbeitssuche erhoben. Auch Bildung, Haushaltseinkommen und Wohnumfeld werden berücksichtigt. Die subjektiven Einschätzungen umfassen unter anderem Fragen zur Erwerbsorientierung, zu Arbeitszeitwünschen und die Bewertung des eigenen Gesundheitszustands.

4.3.3.4 Merkmale zur Verknüpfung

Das PASS wird derzeit nur als SUF in faktisch anonymisierter Form angeboten. Da die Stichprobe der Bedarfsgemeinschaften aus Prozessdaten der BA gezogen ist, ist eine Verknüpfung mit administrativen Daten technisch möglich. Bei der Bevölkerungsstichprobe kann ein Linkage über Adressmaterial durchgeführt werden. Sofern die Befragten ihre Panel- bzw. Zuspieldbereitschaft erklärt haben, ist somit eine Erweiterung von PASS möglich. Das FDZ arbeitet derzeit an einer Verknüpfung, die voraussichtlich Anfang 2014 für externe Nutzer verfügbar sein wird.

Das PASS enthält ferner eine regionale Gliederung nach Bundesländern und eine Branchengliederung auf 2-Steller-Ebene nach WZ03. Die Verknüpfungsmöglichkeiten auf diesen aggregierten Ebenen sind aber aufgrund des Stichprobendesigns und der Fallzahlen eingeschränkt.

4.3.3.5 Identifikation des Evaluationspotenzials

Für die Mindestlohnforschung auf Branchenebene ist PASS für sich allein genommen nur sehr eingeschränkt nutzbar. Hauptgrund hierfür ist, dass die Branchengliederung für die genaue Abgrenzung von Wirtschaftszweigen nicht fein genug ist. Die Stichprobenziehung ist auf Branchenebene auch nicht repräsentativ. Aufgrund des Stichprobendesigns enthalten die Daten außerdem viele Befragte, die überhaupt nicht beschäftigt sind. Die Zahl der direkt Betroffenen ist in vielen Branchen daher sehr klein. Ein weiterer Nachteil von PASS besteht darin, dass die Definition des Bruttoarbeitseinkommens sich über die Wellen mehrmals änderte.

Allerdings ist positiv zu erwähnen, dass mit dem besonderen Fokus auf Bedarfsgemeinschaften in Deutschland eine Bevölkerungsgruppe untersucht werden kann, die überdurchschnittlich durch Mindestlöhne betroffen sein sollte, nämlich die sogenannten „Aufstocker“,

die neben ihrem Arbeitseinkommen zusätzliche Grundsicherungsleistungen erhalten. Durch eine Verknüpfung mit weiteren Datenquellen und eine Ausweitung des Samples könnte das Evaluationspotential erhöht werden.

4.3.4 BIBB/BAuA - Erwerbstätigenbefragung

4.3.4.1 Allgemeine Datensatzinformationen

Das Bundesinstitut für Berufsbildung (BIBB) hat seit 1979 im Abstand von sechs bis sieben Jahren Erwerbstätigenbefragungen durchgeführt. In den Jahren 1979, 1985/86, 1991/92 und 1998/99 wurden die Befragungen in Kooperation mit dem IAB entwickelt und durchgeführt. Seit 2006 erfolgt die Durchführung in Zusammenarbeit mit der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA). Trotz des Wechsels des Kooperationspartners ist nach BIBB und BAuA (2012) eine Vergleichbarkeit der Datensätze zwischen den Erhebungsjahren gegeben. Es ist allerdings zu prüfen, inwieweit dies tatsächlich zutrifft, da sich sowohl die Befragungsmethode als auch die Frageformulierungen, wie im folgenden Absatz näher erläutert, geändert haben (Rohrbach-Schmidt, Hall 2013). In jeder der sechs Erhebungen stehen zwischen 20.000 und 30.000 Beobachtungen zur Verfügung (Rohrbach-Schmidt 2009).

Der Datensatz wird auf Individualebene jeweils als Querschnittsdatsatz erhoben. Als Grundgesamtheit dienen die Erwerbstätigen, die mindestens 15 Jahre alt sind und mindestens zehn Stunden die Woche einer bezahlten Tätigkeit nachgehen. Mit dem Wechsel des Kooperationspartners wurde sowohl das Stichprobendesign als auch die Befragungsmethode geändert. Zwischen 1979 und 1998/99 wurden persönliche Interviews geführt. Ab 2006 sind die Interviews telefonisch (Computer-Assisted Telephone Interviews – CATI). Für die Befragungen von 1979 bis 1998/99 wurde eine mehrstufige Zufallsstichprobe gezogen, während ab 2006 eine Weiterentwicklung des Gabler-Häder-Verfahrens zur Ziehung einer Festnetzstichprobe verwendet wurde. (Rohrbach-Schmidt, Hall 2013, S.15).

4.3.4.2 Verfügbarkeit

Die Datensätze der letzten beiden Wellen können über das FDZ des BIBB beantragt und genutzt werden. Für beide Wellen stehen bereits Scientific-Use-Files (SUF) zur Verfügung, die allerdings aufgrund von Datenschutzbestimmungen keine sensiblen Variablen enthalten. Zudem enthält die SUF-Version des Datensatzes aus 2012 nicht alle generierten Variablen (Rohrbach-Schmidt und Hall 2013). Die Welle 2006 ist ebenfalls On-Site über Datenfernverarbeitung (DFV) oder einen Gastwissenschaftleraufenthalt (GWA) verfügbar. Ab Frühjahr 2014 ist geplant, diese Zugangswege auch für die aktuelle Welle aus 2012 zur Verfügung zu stellen.

4.3.4.3 Merkmale zur Mindestlohnevaluation

Bei den Befragungen handelt es sich zwar um Querschnitte, die aber über eine detaillierte Erfassung des bisherigen Erwerbslebens umfangreiche Informationen retrospektiv enthalten.

Dies beinhaltet Informationen zum höchsten Schulabschluss, zu allen abgeschlossenen Ausbildungen, zum gesamten Erwerbsleben (Erwerbsunterbrechungen, Anzahl Arbeitgeber, Arbeitslosigkeit etc.) sowie für die letzten zwei Jahre Informationen zu Fort- und Weiterbildungen.

Für die Mindestlohnforschung von Interesse sind insbesondere die detaillierten Angaben zur Qualifikation und besuchten Fort- und Weiterbildungen sowie die detaillierten Ausführungen zu den Tätigkeiten und Anforderungen, die sich aus dem ausgeübten Beruf ergeben. Zusätzlich werden auch psychische und physische Belastungen, die sich aus der Erwerbstätigkeit ergeben, erfasst. Die umfangreichen Angaben zum Erwerbsleben werden sowohl durch soziodemografische Merkmale als auch durch Charakteristika des Arbeitgebers ergänzt.

4.3.4.4 Merkmale zur Verknüpfung

Da die Adressdaten der Befragten spätestens Ende 2013 gelöscht werden mussten (BIBB, 2011) und auch keine anderen Verknüpfungsmerkmale auf Individualebene zur Verfügung stehen, ist eine Verknüpfung auf Individualebene nicht möglich. Anhand der Wirtschaftszweige sowie der Gemeindegrenznummer ist eine Verknüpfung technisch möglich. Allerdings ist zu beachten, dass die Wirtschaftszweige ausschließlich als Dreisteller verfügbar sind (Rohrbach-Schmidt und Hall 2013). Trotz der technischen Verknüpfungsmöglichkeit ist fraglich, inwieweit eine Verknüpfung methodisch sinnvoll ist, da vermutlich die Repräsentativität auf Gemeinde- oder Wirtschaftszweigebene aufgrund von zu geringen Fallzahlen nicht gegeben ist.

4.3.4.5 Identifikation des Evaluationspotenzials

Die BIBB/BAuA-Erwerbstätigenbefragung bietet inhaltlich ein sehr großes Evaluationspotenzial, da kaum ein anderer Datensatz in Deutschland so umfangreiche Informationen zu Qualifikationen, Fort- und Weiterbildungen, zu psychischen und physischen Arbeitsbelastungen sowie zur aktuellen und bisherigen Erwerbssituation enthält. Für allgemeine Analysen zu diesen Themen ist eine Stichprobengröße von 20.000 bis 30.000 Beobachtungen durchaus ausreichend. Da aber lediglich etwa 13 % der sozialversicherungspflichtig Beschäftigten einem Mindestlohn nach dem AEntG unterliegen,²⁹ ist die Fallzahl mit etwa 2.700 bis 4.000 Beobachtungen, die einer sozialversicherungspflichtigen Beschäftigung in einer Mindestlohnbranche nachgehen, im Datensatz zu klein, um verlässliche Analysen treffen zu können. Dies gilt insbesondere, da für Analysen die Mindestlohnbranchen nicht als eine Gruppe insgesamt verwendet werden können.

Eine weitere Problematik ergibt sich aus der sehr groben Einteilung der Wirtschaftszweige. Vorausgesetzt, dass bei einer On-Site-Nutzung der Daten sämtliche Freitextfelder zu Aus-

²⁹ Im August 2013 gab es nach Angaben der BA 29,5 Mio. Erwerbstätige, von denen laut BMAS (2013b) rund 4 Mio. in einer Mindestlohnbranche beschäftigt waren.

wertungszwecken zur Verfügung stehen, könnte anhand der Angaben der Befragten eine detailliertere Aufteilung der Wirtschaftszweige vorgenommen werden. Allerdings ist davon auszugehen, dass dies mit einem sehr hohen Aufwand verbunden ist. Zusätzlich ist der Abstand von sechs Jahren zwischen den einzelnen Wellen relativ groß, was das Evaluationspotenzial weiter einschränkt.

4.3.5 Datenbank des Copenhagen Psychological Questionnaire (COPSOQ)

4.3.5.1 Allgemeine Datensatzinformationen

Das Copenhagen Psychological Questionnaire (COPSOQ) als Instrument zur Messung der Arbeitsbelastungen und der Arbeitsqualität wurde erstmals im Auftrag der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA) zwischen 2003 und 2005 durch ein Forschungskonsortium bestehend aus der Freiburger Forschungsstelle Arbeits- und Sozialmedizin (ffas) und den Universitäten Freiburg und Wuppertal entwickelt und validiert. Auf Basis des validierten Fragebogens hat die ffas seit 2005 eine Datenbank auf- und seitdem kontinuierlich ausgebaut. Die Grundgesamtheit besteht aus Individuen, die seit 2005 mindestens einen Tag beschäftigt waren.

Die Auswahl der Befragten erfolgt allerdings nicht auf Basis einer Stichprobenziehung sondern anhand der Teilnahme von Betrieben am Kooperationsmodell Praxis-Wissenschaft, in dessen Rahmen eine Durchführung und Auswertung der COPSOQ-Befragung durch die ffas angeboten werden. Dementsprechend ist die Datenbank eine Datensammlung von Beschäftigten aus Kooperationsbetrieben. Auf diese Weise sind mittlerweile etwa 140.000 Befragungen von Beschäftigten in die Datenbank eingepflegt und ausgewertet worden (Lincke et al. 2013).

4.3.5.2 Verfügbarkeit

Im Auftrag der BAuA wurde 2008 neben der COPSOQ-Datenbank, die mittlerweile etwa 140.000 Beschäftigte umfasst, eine Online-Datenbank aufgebaut, in der derzeit 10.022 Datensätze erfasst sind.³⁰ Die Auswahl der für die Online-Datenbank verwendeten Datensätze wurde anhand der Verteilung der Berufsabschnitte in der Statistik der BA vorgenommen (Nübling et al. 2011).

Der Zugang zur Online-Datenbank ist kostenlos und ohne einen Antrag auf der Homepage www.copsoq-datenbank.de abrufbar. Mit Hilfe einer Onlinemaske können Einzelergebnisse differenziert nach bestimmten Merkmalen erfasst werden. Allerdings werden Werte nur angezeigt, wenn mehr als zehn Fälle vorliegen. Ein Zugriff auf den kompletten Datenbestand ist ausschließlich in Absprache und Kooperation mit der ffas möglich.

³⁰ Abfrage der Fallzahl auf <https://www.copsoq-datenbank.de/db-abfrage.php> [aufgerufen am: 29.10.2013].

4.3.5.3 Merkmale zur Mindestlohnevaluation

Der COPSOQ-Fragebogen ist eine erprobte Möglichkeit die Arbeitsqualität zu messen, die unter anderem in einem aktuellen Projekt des BMAS zur gewünschten und erlebten Arbeitsqualität (siehe Unterabschnitt 4.3.7) verwendet wird. Die Merkmale zur Arbeitsqualität wurden größtenteils mit einer Likert-Skala erfasst und sind damit identisch skaliert. Die gegenüber dem eigentlichen COPSOQ-Fragebogen leicht gekürzte Fassung enthält 87 Items die abgefragt werden (Lincke et al. 2013). Die Items beziehen sich insbesondere auf die Themenbereiche Anforderungen,³¹ Einfluss und Entwicklungsmöglichkeiten sowie soziale Beziehungen und Führung. Weitere zusätzliche erhobene Merkmale betreffen die Arbeitsplatzunsicherheit, berufsspezifische Faktoren, Interesse an Schulungen, Belastungen durch Schichtarbeit, das eigene Gesundheitsverhalten sowie die Existenz von Kundenkontakten bzw. -konflikten. Als Ergebnisindikatoren stehen insbesondere Belastungsfolgen zur Verfügung. Dies sind Arbeitszufriedenheit, Gedanke an Berufsaufgabe, der allgemeine Gesundheitszustand, die Gefahr von Burnout, kognitiver Stress sowie die Lebenszufriedenheit (Lincke et al. 2013, Abb. 1).

Neben den Indikatoren zur Arbeitsbelastung enthalten die COPSOQ-Fragebogen Angaben zur Person und dem Arbeitsplatz, den Namen des Unternehmens, Geschlecht, Geburtsjahr, Beruf und Beschäftigungsverhältnis, nicht aber zum Einkommen (Lincke et al. 2013).³²

4.3.5.4 Merkmale zur Verknüpfung

Über die Wirtschaftszweige – auf Ebene der Viersteller – ist eine Verknüpfung mit anderen Datensätzen technisch möglich. Sofern die Adressdaten der Kooperationsbetriebe noch vorliegen sowie mit der COPSOQ-Datenbank verknüpfbar sind, ist eine Verknüpfung auf Individualebene über Record Linkage technisch möglich.

4.3.5.5 Identifikation des Evaluationspotenzials

Das Evaluationspotenzial der COPSOQ-Datenbank ist sehr eingeschränkt. Es handelt sich mehr um eine Datensammlung einzelner Befragungen in Kooperationsbetrieben als um eine Repräsentativbefragung. Dies hat zum einen den Nachteil, dass die zeitliche Struktur zwar anhand von Jahresangaben nachvollziehbar ist, die Befragungen aber nicht zu vordefinierten Zeitpunkten innerhalb der Jahre stattfanden. Zum anderen ist die COPSOQ-Datenbank nicht repräsentativ, da die Befragung in Kooperationsbetrieben stattfand, das heißt es besteht eine nicht vollständig kontrollierbare Selektivität der erfassten Betriebe, da nichts darüber bekannt ist, welche Betriebe ein grundsätzliches Interesse an einer Messung der Arbeitsqualität haben, welche das Kooperationsmodell Praxis-Wissenschaft der ffas kennen und welche Be-

³¹ Bspw. Quantitative und emotionale Anforderungen, Gefühle verbergen, Work-Privacy-Conflict.

³² Je nach Wunsch des Kooperationsbetriebes können weitere Merkmale vorhanden sein. Die aufgelisteten Merkmale entsprechen der Standardversion.

triebe sich schließlich zu einer Teilnahme entscheiden. Insbesondere bei kleinen Betrieben, wie sie vor allem in den Mindestlohnbranchen im Handwerk existieren, besteht die Gefahr, dass diese nicht ausreichend repräsentiert sind. Die Löhne der Beschäftigten werden nicht erfasst.

Aufgrund der genannten Einschränkungen ist eine direkte Nutzung der COPSOQ-Datenbank für die Mindestlohnforschung nicht sinnvoll. Dennoch könnte insbesondere die Anwendung des COPSOQ-Konzepts zur Messung der Arbeitsqualität sinnvoll sein. Die Online verfügbare COPSOQ-Datenbank der ffas gibt einen guten Überblick über die Analysemöglichkeiten einer Messung der Arbeitsqualität mit diesem Konzept.

4.3.6 Sozio-Oekonomisches Panel (SOEP)

4.3.6.1 Allgemeine Datensatzinformationen

Das Sozio-Oekonomische Panel (SOEP) ist die größte Panelbefragung deutscher Haushalte, die seit 1984 jährlich durchgeführt wird. Seit Mitte November liegen die Daten für 29 Jahre bis einschließlich 2012 vor. Die 30. Welle (2013) wird gegen Ende 2014 vorliegen.

Das SOEP ist eine kombinierte Haushalts- und Personenbefragung auf freiwilliger Basis. Bis 2000 wurden die Interviews persönlich in Papierform als sogenanntes Paper and Pencil Interview (PAPI) durchgeführt. Seit 2000 werden die Interviews auch als Computer-Assistierte Persönliche Interviews (CAPI) durchgeführt (tns infratest 2012, S. 33). Derzeit nehmen etwa 12.000 Haushalte mit mehr als 20.000 Personen ab 16 Jahren³³ am SOEP teil. Allerdings haben nicht alle Personen seit 1984 am SOEP teilgenommen, sondern sind durch spätere Auffrischungsstichproben hinzugekommen. Tabelle 4.1 zeigt die Zusammensetzung der Stichprobe für SOEP Welle 2011.

Tabelle 4.1: Stichprobenzusammensetzung des SOEP Welle 2011

Jahr der ersten Teilnahme	Haushaltsinterviews	Teilnehmende Personen
1984	2.539	4.541
1990	1.355	2.392
1995	266	461
1998	545	961
2000	2.886	4.984
2002	706	1.358
2006	857	1.478
2011	3.136	5.161
Gesamt	12.290	21.336

Quelle: Eigene Darstellung auf Basis von tns infratest (2012, S. 6)

³³ In den befragten Haushalten werden die Personen erst nach Abschluss ihres 16. Lebensjahres in die Befragung aufgenommen.

Aufgrund eines fehlenden zentralen Bevölkerungsregisters in Deutschland werden die Stichproben des SOEP in der Regel auf Grundlage des Stichprobenverfahrens der Arbeitsgemeinschaft Deutscher Marktforschungsinstitute (ADM) gezogen (Haisken-DeNew und Frick 2005). Ausnahmen sind unter anderem die Erweiterung des SOEP 1990 auf die neuen Bundesländer sowie die beiden Migrations- oder Ausländerstichproben. In beiden Fällen wurden Registerdaten für die Ziehung herangezogen (Haisken-DeNew und Frick 2005).

4.3.6.2 Verfügbarkeit

Das SOEP steht kostenlos für wissenschaftliche Zwecke zur Verfügung. Seit diesem Jahr werden die Daten über einen passwortgeschützten Austauschserver den Datennutzerinnen und -nutzern zur Verfügung gestellt. Durch die Umstellung des Verfahrens entfallen die Bereitstellungskosten für den DVD-Versand von 30 Euro zuzüglich Versandkosten von 8 Euro.

Voraussetzung für den Erhalt der SOEP-Daten ist eine ausschließlich wissenschaftliche Nutzung der Daten. Daher werden diese nur an Forschungseinrichtungen weitergegeben, die den Nachweis unabhängiger wissenschaftlicher Forschung erbringen. Die Datenbereitstellung erfolgt auf Grundlage eines Datenweitergabevertrags. Da die Datensätze aufbereitet vorliegen und lediglich der Datenweitergabevertrag unterschrieben werden muss, kann in der Regel mit einer Bereitstellung der Daten innerhalb von drei Tagen gerechnet werden. Dies gilt nicht, wenn Verknüpfungen mit Regionaldaten vorgesehen sind. Hier muss ein individuelles Datenschutzkonzept im Vorfeld abgestimmt werden.

4.3.6.3 Merkmale zur Mindestlohnevaluation

Das SOEP setzt sich aus insgesamt neun verschiedenen Befragungselementen zusammen, die von unterschiedlichen Personen beantwortet werden. Für die Mindestlohnforschung sind vor allem die Personenbiografie sowie der Personen- und der Haushaltsfragebogen von Bedeutung.

Die Personenbiografie ist insbesondere für die Identifikation von Kontrollvariablen relevant. Der Fragebogen umfasst neben den üblichen soziodemografischen Merkmalen ausführliche Informationen über das Elternhaus der Befragten sowie über die bisherige schulische und berufliche Bildungsbiografie und über die Erwerbsbiografie im Allgemeinen.

Anhand des Personenfragebogens, in dem die aktuelle Tätigkeit, inklusive einer genauen Berufsbeschreibung erhoben wird, ist eine Identifikation derjenigen Befragten möglich, die in einer Mindestlohnbranche arbeiten. Zusätzlich enthält der Personenfragebogen Angaben zur Erwerbstätigkeit wie Arbeitszeitregelungen und tatsächlich geleistete Arbeitszeit, Größe des Betriebs, Zufriedenheit mit dem Verdienst, die Entfernung der Arbeitsstelle vom Wohnort sowie weitere Arbeitsbedingungen. Zusätzlich werden sowohl das Brutto- als auch das Nettoeinkommen ohne Zusatzleistungen³⁴ erfasst. Das Fragenspektrum wird ergänzt durch um-

³⁴ Die Zusatzleistungen werden separat erfasst

fangreiche Fragen zum Gesundheitszustand, dem sozialen Netzwerk sowie zu persönlichen Einstellungen.

Gegenstand des Haushaltsfragebogens sind weitere Merkmale zur Wohnsituation, der Ausstattung des Haushalts sowie der Haushaltsgröße und -zusammensetzung. Zusätzlich wird das Haushaltseinkommen erhoben.

4.3.6.4 Merkmale zur Verknüpfung

Das FDZ des SOEP bietet eine Reihe an Regionaldaten an, die mit dem SOEP verknüpft werden können. Die Verfügbarkeit ist dabei sehr stark vom gewählten Nutzungsweg abhängig. Standardmäßig ist für alle Wellen das Bundesland des Wohnorts im SUF integriert. Tiefergehende Analysen sind mit zusätzlichen Datenschutzbestimmungen versehen. So können auf Ebene der Gemeindegrößenklassen oder der Raumordnungsregionen Regionaldaten zum SUF hinzugespielt werden. Allerdings werden hierfür ein erweiterter Datenweitergabevertrag und Datenschutzkonzept verlangt.

Die Daten lassen sich bis auf Postleitzahlen- und Gemeindegemeinschaftsebene aufteilen und mit Regionalmerkmalen verknüpfen. Aufgrund der Sensibilität der Daten können diese Daten ausschließlich im Rahmen eines Gastaufenthalts genutzt werden.³⁵ Auch diese beiden Möglichkeiten gehen über das Standardangebot des FDZ des SOEP hinaus und können lediglich in Form von persönlichen Absprachen eingerichtet werden.

Grundsätzlich ist eine Verknüpfung der SOEP-Daten mit anderen Individualdaten wie den IEB (Unterabschnitt 4.1.1) technisch möglich, da für die Generierung des Panels die Klarnamen und Adressen der Befragten mit dem SOEP verknüpfbar sein müssen.

4.3.6.5 Identifikation des Evaluationspotenzials

Das SOEP bietet ein sehr umfassendes Analysepotenzial, da die Informationen teilweise bis einschließlich 1984 zurückreichen und in den einzelnen Jahren ein sehr umfangreiches Befragungsprogramm durchgeführt wird.

Insbesondere für kleinere Mindestlohnbranchen ist es jedoch wahrscheinlich, dass trotz der großen Beobachtungszahl von 21.336 Personen (2011), von denen maximal die Hälfte einer sozialversicherungspflichtigen Beschäftigung nachgehen, zu wenige in der betroffenen Branche beschäftigt sind, um valide statistische Analysen zu ermöglichen. Die Fallzahlen lassen vermuten, dass insgesamt etwa 1.400 der im SOEP befragten Personen überhaupt von einem Mindestlohn nach dem AEntG betroffen sind.

³⁵ Auf Kreisebene ist eine Verknüpfung bei gleichzeitiger Nutzung der Datenfernverarbeitung möglich.

4.3.7 Befragung zur gewünschten und erlebten Arbeitsqualität

4.3.7.1 Allgemeine Datensatzinformationen

Das BMAS hat ein Forschungskonsortium bestehend aus der Freiburger Forschungsstelle Arbeits- und Sozialmedizin (ffas), dem Forschungszentrum Familienbewusste Personalpolitik (FFP) und dem Institut für angewandte Sozialwissenschaft (infas) beauftragt, eine repräsentative Studie zur Arbeitsqualität durchzuführen. Die Messung der Arbeitsqualität wurde dabei anhand des im Unterabschnitt 4.3.5 vorgestellten Konzepts Copenhagen Psychological Questionnaire (COPSOQ) durchgeführt.

Da das COPSOQ-Konzept eine Messung der Arbeitsqualität auf Individualebene vorsieht, ist auch in diesem Datensatz die Beobachtungseinheit das Individuum. Als Grundgesamtheit wurden sämtliche Beschäftigte mit einer wöchentlichen Arbeitszeit von zehn Stunden oder mehr gewählt. Die ungeschichtete Zufallsstichprobe wurde anhand der Beschäftigtenstatistik der Bundesagentur für Arbeit (BA) gezogen. Insgesamt wurden 5.004 Beschäftigte mit Computer-assistierte Telefoninterviews (CATI) befragt.

Die Befragung fand 2013 statt und wurde als Panelbefragung angelegt, d.h. es wurde die Teilnahmebereitschaft an einer erneuten Befragung ermittelt. Allerdings sieht das aktuell laufende Projekt keine finanziellen Mittel für eine wiederholte Befragung vor. Dementsprechend wird eine Ausweitung des derzeitigen Querschnittsdatensatzes zu einem Panel angestrebt, ist aber abhängig von einer Finanzierung einer erneuten Befragung. Folglich steht weder fest, ob es eine wiederholte Befragung geben wird noch zu welchem Zeitpunkt diese stattfinden wird.

4.3.7.2 Verfügbarkeit

Es ist geplant die Befragungsdaten in Form eines Public-Use-Files (PUF) für Forschungszwecke zur Verfügung zu stellen. Da das Projekt derzeit aber noch in Bearbeitung ist, steht weder der Zeitpunkt der Veröffentlichung fest, noch ist sicher, über welchen Weg die Daten zur Verfügung gestellt werden. Dementsprechend liegen derzeit keine Informationen über die Voraussetzungen, Dauer und Kosten der Beantragung vor.

4.3.7.3 Merkmale zur Mindestlohnevaluation

Der Fragebogen entspricht im Wesentlichen den Merkmalen, die bei den COPSOQ-Befragungen erhoben werden (siehe Unterabschnitt 4.3.5). Dementsprechend sind die gleichen Themenbereiche wie bei COPSOQ enthalten. Dies sind Anforderungen,³⁶ Einfluss und Entwicklungsmöglichkeiten sowie soziale Beziehungen und Führung. Weitere zusätzliche erhobene Merkmale betreffen die Arbeitsplatzunsicherheit, berufsspezifische Faktoren, Interesse an Schulungen, Belastungen durch Schichtarbeit, das eigene Gesundheitsverhalten

³⁶ Bspw. Quantitative und emotionale Anforderungen, Gefühle verbergen, Work-Privacy-Conflict.

sowie die Existenz von Kundenkontakten bzw. -konflikten. Als Ergebnisindikatoren stehen insbesondere Belastungsfolgen zur Verfügung. Dies sind Arbeitszufriedenheit, Gedanke an Berufsaufgabe, der allgemeine Gesundheitszustand, die Gefahr von Burnout, kognitiver Stress sowie die Lebenszufriedenheit (Lincke et al. 2013, Abb. 1). Zudem werden soziodemografische Merkmale wie Alter, Geschlecht, Beruf, sowie wesentliche Merkmale zum Betrieb, wie Betriebsgröße oder Sitz des Betriebs (Lincke et al. 2013) erfasst.

4.3.7.4 Merkmale zur Verknüpfung

Die Ziehunggrundlage der Befragung ist die Beschäftigtenstatistik der BA. In der Befragung wurde zudem die Zuspülerlaubnis der Befragungsergebnisse zu den Prozessdaten der BA erbeten. Dementsprechend ist eine Verknüpfung der Befragung mit der Beschäftigtenstatistik und mit den IEB für die Befragten, die einer Verknüpfung zugestimmt haben möglich. Wie viele Befragte einer Verknüpfung zugestimmt haben, ist nicht bekannt. Erfahrungen aus ähnlichen Befragungen bei Individuen lassen aber eine hohe Zustimmung vermuten.

4.3.7.5 Identifikation des Evaluationspotenzials

Mit 5.004 befragten Beschäftigten über alle Branchen hinweg ist das Evaluationspotenzial des Datensatzes für die Mindestlohnforschung sehr stark eingeschränkt, da die Fallzahlen für die einzelnen Mindestlohnbranchen bzw. auch für alle Mindestlohnbranchen zusammen sehr gering sein werden. Eine Identifikation der Mindestlohn Betroffenheit auf Individualebene ist derzeit nicht möglich. Über eine Ausweitung des Datensatzes um Zusatzstichproben in den Mindestlohnbranchen könnte dieser allerdings als Basis für einen verwendbaren Datensatz dienen. Ein weiterer Nachteil der Befragung ist der Zeitpunkt, da sie für die meisten der derzeit existierenden Mindestlohnbranchen keine Nullmessung vor Mindestlohneinführung darstellt. Ausnahmen bilden die Mindestlohneinführungen im Gerüstbauerhandwerk (seit 1. August 2013) sowie im Steinmetz- und Steinbildhauerhandwerk (seit 1. Oktober 2013). Allerdings ist zu bezweifeln, dass die Fallzahlen im Datensatz in einem der beiden Branchen für Analysen ausreichend sind.³⁷ Zu überprüfen ist, inwieweit die Befragung als Nullmessung für eine Veränderung der Arbeitsqualität aufgrund einer starken Anhebung des Mindestlohns bzw. einer Einführung des Mindestlohns in einer weiteren Branche genutzt werden kann.

5 Daten auf Betriebs-/Unternehmensebene

Die in diesem Abschnitt dargestellten Datensätze nutzen als zugrundeliegende Beobachtungseinheit Betriebe bzw. Unternehmen. Für Analysen auf dieser Ebene können jedoch auch zum Teil Individualdatensätze auf Personenebene genutzt werden, sofern diese über

³⁷ Im Gerüstbauerhandwerk sind 20.000 sozialversicherungspflichtig Beschäftigte vom Mindestlohn betroffen, im Steinmetz- und Steinbildhauerhandwerk sind es 13.100 (BMAS 2013b). Bei 5.004 befragten Beschäftigten ist in jedem der beiden Handwerke mit einer niedrigen einstelligen Fallzahl im Datensatz zu rechnen.

einen Betriebs- oder Unternehmensidentifikator aggregiert werden können. Aus dem vorherigen Abschnitt sind hierfür prinzipiell die IEB (Abschnitt 4.1.1), die SIAB (Abschnitt 4.1.2), die Lohn- und Einkommenssteuerstatistik (Abschnitt 4.1.3), die Daten der Sozialkassen (Abschnitte 4.1.5, 4.1.6 und 4.1.7), die Verdienststrukturerhebung (Abschnitt 4.2.1) und die Befragung im Bauhauptgewerbe (Abschnitt 4.3.1) sowie das Linked Personnel Panel Beschäftigte (Unterabschnitt 4.3.2) geeignet. Allerdings muss zum Teil geprüft werden, ob die aggregierten Daten auf Betriebsebene repräsentativ sind.

5.1 Administrative Daten

5.1.1 Betriebs-Historik-Panel (BHP)

5.1.1.1 Allgemeine Datensatzinformationen

Das Betriebs-Historik-Panel (BHP) ist eine 50%-Stichprobe aller Betriebe in Deutschland, die zum Stichtag 30. Juni eines Jahres mindestens einen sozialversicherungspflichtig oder seit 1999 auch geringfügig Beschäftigten beschäftigen. Das BHP wird aus der Beschäftigten-Historik (BeH) des IAB gezogen, die auch Bestandteil der IEB ist. Sie enthält Daten aus dem „Gemeinsamen Meldeverfahren zur gesetzlichen Kranken-, Pflege-, Renten- und Arbeitslosenversicherung“ und liegt in Westdeutschland seit 1975 und in Ostdeutschland seit 1991 vor. Die Personendaten der BeH werden mit Hilfe eines Betriebsidentifikators auf die Betriebsebene aggregiert.

Das BHP liegt in zwei Versionen vor, die alternativ beantragt werden können. Einerseits eine reine 50%-Zufallsstichprobe, andererseits eine nach Größenklassen geschichtete 50%-Zufallsstichprobe. Die Größenklassen werden über die Zahl der Gesamtbeschäftigten in einem Betrieb festgelegt (<4, 5-9, 10-19, 20-49, 50-99, 100-199, 200-499, >=500).

Das BHP besteht aus drei Modulen. Das Hauptmodul liegt in Jahresscheiben vor und beinhaltet Angaben zu Standort, Wirtschaftszweig, Beschäftigungs- und Gehaltsstruktur. Neben dem Kerndatensatz gibt es zwei Erweiterungsmodule, eines zu Beschäftigungsströmen (d.h. Zu- und Abgängen von Beschäftigten im Betrieb) und eines zu Betriebsdynamiken (d.h. Gründungen und Schließungen). Über eine Zusatzdatei können außerdem die sich im Zeitablauf ändernden Wirtschaftszweigklassifikationen im BHP zeitkonsistent gemacht werden. Diese Datei beruht auf Analysen des FDZ. Die einzelnen Dateien können über eine Betriebs-ID und Jahresangabe zusammengeführt werden.

Das BHP in seiner aktuellen Version (BHP 7510) umfasst insgesamt ca. 4,2 Mio. Betriebe, zwischen 0,6 und 2 Millionen pro Jahr.

5.1.1.2 Verfügbarkeit

Das BHP wird durch das FDZ der BA im IAB angeboten. Das FDZ ist auch für die Erstellung, Aktualisierung und Dokumentation zuständig und übernimmt die Nutzerbetreuung. Der Datensatz ist sowohl als schwach anonymisierte Version über einen Gastaufenthalt als auch

über Datenfernverarbeitung nutzbar. Die Beantragung folgt gemäß den im Unterabschnitt 3.1 beschriebenen Voraussetzungen.

Die Dauer der Beantragung ist abhängig vom Datennutzungsweg und dauert bei Gastaufenthalt aufgrund des zusätzlichen Genehmigungsverfahrens etwas länger. Sicherheitshalber sollten 2 bis 4 Wochen veranschlagt werden. Die Freischaltung für die Datenfernverarbeitung erfolgt bei korrektem Antrag meist innerhalb weniger Tage. Die Datennutzung selbst ist kostenlos.

Während das Hauptmodul standardmäßig angeboten wird, bedarf es bei Beantragung der Erweiterungsmodule einer gesonderten Begründung. Auch bestimmte Merkmale, die eine Deanonymisierung von Betrieben erleichtern können, werden nur dann im Original weitergegeben, wenn es für das Analyseziel notwendig ist und dies im Antrag auf den Datenzugang explizit begründet wird. Diese aus datenschutzrechtlicher Sicht besonders sensiblen Merkmale im Kerndatensatz umfassen Kreisinformationen zum Arbeitsort und fünfstelligen Wirtschaftsunterklassen.

Das BHP wird alle zwei Jahre aktualisiert und ist derzeit für den Zeitraum 1975 bis 2010 verfügbar. Ostdeutsche Betriebe sind ab 1991 enthalten. Aussagekräftige Analysen ostdeutscher Betriebe sind jedoch erst ab dem Jahr 1993 möglich. Das nächste Update ist für Anfang 2014 geplant. Der abgedeckte Zeitraum soll dann bis 2012 erweitert werden. Umfangreiche Dokumentationen und Auszählungen sind auf der Internetpräsenz des FDZ verfügbar.

5.1.1.3 Merkmale zur Mindestlohnevaluation

Das BHP enthält Angaben zur Gehaltsstruktur der Vollzeitbeschäftigten innerhalb eines Betriebs zum Stichtag 30. Juni eines Jahres. Diese umfassen Quartile des Bruttoentgelts insgesamt und getrennt nach verschiedenen Untergruppen. Darüber hinaus kann die Beschäftigungsstruktur im Betrieb sehr genau abgebildet werden, wobei unter anderem nach Alter, Nationalität, Geschlecht, Bildungsabschlüssen und Berufen unterschieden werden kann.

Ein- und Austritte von Beschäftigten im jeweiligen Betrieb können mit dem entsprechenden Zusatzmodul ebenfalls abgebildet werden, wobei wiederum nach verschiedenen Dimensionen der Beschäftigungsstruktur unterschieden werden kann. Neugründungen und Betriebs-schließungen können über das Betriebsdynamik-Modul identifiziert werden.

Das BHP enthält einen Regionalschlüssel bis maximal auf Kreisebene (zu einem einheitlichen Gebietsstand) und Angaben zum Wirtschaftszweig bis hin zur 5-Steller-Ebene. Diese Definition änderte sich im Zeitablauf mehrmals. Vor 1998 sind nur Informationen zur dreistelligen Wirtschaftsklasse 73 vorhanden. Die Zusatzdatei ermöglicht jedoch die Generierung zeitkonsistenter Wirtschaftszweige.

Bei den Angaben zum Arbeitsort ist insbesondere auch die Logik der Betriebsnummernvergabe zu beachten. Hierbei werden mehrere Arbeitsstätten des Arbeitgebers zu einem Betrieb zusammengefasst, wenn sie sich in derselben Gemeinde befinden und dieselbe wirt-

schaftsfachliche Zuordnung aufweisen. Dadurch werden Beschäftigte demselben Betrieb zugeordnet, auch wenn sie an unterschiedlichen Standorten innerhalb einer Gemeinde tätig sind.

5.1.1.4 Merkmale zur Verknüpfung

Das BHP enthält in der schwach anonymisierten Version Betriebsidentifikatoren, die über Schlüsselstabellen rein technisch Verknüpfungen mit der IEB und anderen Datenbeständen der BA zulassen. Dazu gehört beispielsweise auch Adressmaterial der Betriebe, wodurch ein Record Linkage mit anderen Datenquellen möglich wäre. Die Vertragsbestimmungen des FDZ lassen allerdings eine Verknüpfung des BHP mit anderen Mikrodaten nicht zu.

Auf aggregierter Ebene ermöglichen der Regionalschlüssel sowie der Wirtschaftszweig eine Verknüpfung mit anderen Daten.

5.1.1.5 Identifikation des Evaluationspotenzials

Zunächst erscheint das BHP aufgrund seiner Größe, seinen detaillierten Angaben zur Beschäftigungsstruktur und der genauen Abgrenzbarkeit bestimmter regionaler Einheiten oder Wirtschaftszweige sehr gut für die Mindestlohnevaluation auf Betriebsebene geeignet. Die Herkunft aus administrativen Quellen sichert zumindest bei einigen zentralen Angaben eine sehr hohe Datenqualität. Allerdings gibt es klare Einschränkungen, die bei der Projektgestaltung beachtet werden müssen.

Ein betriebspezifisches Maß für die Mindestlohnbetreffenheit zu berechnen ist aus zwei Gründen problematisch. Der erste Grund besteht darin, dass Stundenlöhne in den administrativen Daten der BA nicht präzise zu berechnen sind und nur unter relativ restriktiven Annahmen abgeschätzt werden können. Der zweite Grund ist, dass die Entgeltverteilung im Betrieb lediglich durch vier Quartile dargestellt ist. Daher ist das BHP am ehesten geeignet, eine Mindestlohnbranche mit einer sorgsam gewählten Kontrollgruppe aus anderen Branchen zu vergleichen.

Aufgrund des quellenbedingt eingeschränkten Informationsgehalts im BHP sind außerdem einige aus Betriebssicht interessante Dimensionen wie beispielsweise Wettbewerb, Output und Investitionen nicht analysierbar.

5.1.2 Umsatzsteuerstatistik

5.1.2.1 Allgemeine Datensatzinformationen

Die Umsatzsteuerstatistik erfasst seit 1996 jährlich die Angaben der Unternehmen bei den Umsatzsteuervoranmeldungen. Seit 1998 existiert mit der Umsatzsteuer-Identifikationsnummer ein eindeutiger Identifikator, mit dem die einzelnen Jahre zu einem Panel verknüpft werden können (FDZ der Statistischen Ämter, 2013c).

In der Umsatzsteuerstatistik sind sämtliche Unternehmen, die zu einer Voranmeldung ihrer Umsatzsteuer bei den Finanzämtern verpflichtet sind und deren Jahreswert der Lieferungen und Leistungen die Freigrenze der Umsatzsteuer von 17.500 Euro übersteigt, enthalten. Es handelt sich dabei um eine Sekundärerhebung aus Angaben, die in den Rechenzentren der Landesfinanzbehörden zusammengestellt werden (Statistisches Bundesamt, 2013b).

Derzeit sind Daten bis 2011 verfügbar. Da für die Aufbereitung der Daten zwölf bis maximal 16 Monate benötigt werden ist im Frühjahr 2014 mit einer Aktualisierung der Umsatzsteuerstatistik um ein Jahr zu rechnen.

5.1.2.2 Verfügbarkeit

Die Umsatzsteuerstatistik ist ausschließlich über einen On-Site-Zugang (Datenfernverarbeitung oder Gastwissenschaftlerarbeitsplatz) nutzbar. Der Beantragungsweg entspricht dem anderer Produkte der FDZ der Statistischen Ämter und wird im Unterabschnitt 3.2 näher beschrieben. Die Dauer der Beantragung ist sehr stark von der Sensibilität der beantragten Merkmale abhängig.

5.1.2.3 Merkmale zur Mindestlohnevaluation

Die Umsatzsteuerstatistik enthält sämtliche Merkmale, die als Berechnungsgrundlage für die Umsatzsteuer dienen. Dies bedeutet eine sehr umfangreiche Auflistung von Erträgen, absetzbaren Vorleistungen sowie zu bereits geleisteter Umsatzsteuer, bspw. über die Einfuhrumsatzsteuer. Eine komplette Variablenliste befindet sich in der Beschreibung der Metadaten der FDZ der Statistischen Ämter (2013c).

5.1.2.4 Merkmale zur Verknüpfung

Anhand der Steuernummer sowie der Umsatzsteuer-Identifikationsnummer ist eine Verknüpfung der Umsatzsteuerstatistik mit dem URS möglich.³⁸ Weitere Verknüpfungsmerkmale sind der Wirtschaftszweig, der eine branchenspezifische Verknüpfung ermöglicht, sowie der amtliche Gemeindeschlüssel für eine regionale Verknüpfung auf der Gemeindeebene.

5.1.2.5 Identifikation des Evaluationspotenzials

Die Umsatzsteuer ist ein geeigneter Indikator für die Wirtschaftsaktivität eines Unternehmens. Dies betrifft neben der Umsatzentwicklung, die auch im URS enthalten ist (siehe Unterabschnitt 5.1.3), ebenfalls Angaben zu Vorleistungen und Einfuhren. Dies ermöglicht eine detailliertere branchenspezifische Analyse der Mindestlohneffekte auf die Umsatzentwicklung der betroffenen Unternehmen.

³⁸ Zu den weiteren Verknüpfungsmöglichkeiten über das URS siehe Unterabschnitt 5.1.3.

Problematisch ist, dass die Umsatzstatistik auf Unternehmensebene erhoben wird. Dadurch erhält die Abgrenzung des Geltungsbereichs eines Mindestlohnes neben der unscharfen Abgrenzungsmöglichkeit von Tarifgebieten anhand der Wirtschaftszweige, eine weitere Unschärfe. Mindestlohnbranchen mit einem hohen Anteil an Betrieben, die zu einem Unternehmen eines anderen Wirtschaftszweiges gehören, werden über die Abgrenzung anhand der Wirtschaftszweige nicht vollständig erfasst.

5.1.3 Unternehmensregister – System 95 (URS)

5.1.3.1 Allgemeine Datensatzinformationen

Das Unternehmensregister – System 95 (URS) ist eine Zusammenführung verschiedener administrativer Datenquellen. Als Vollerhebung aller in Deutschland wirtschaftlich aktiven Unternehmen und Betriebe³⁹ ist das URS die Datengrundlage für fast alle amtlichen Erhebungen auf Betriebs- und Unternehmensebene.

Das URS setzt sich aus einer meist jährlich durchgeführten Verknüpfung folgender administrativer Datenquellen zusammen:

- Steuerverwaltung:
 - Umsatzsteuerdateien der Oberfinanzdirektionen
 - Organschaftsdatei des Bundeszentralamtes für Steuern
- Bundesagentur für Arbeit: Betriebsdatei
- Kammern:
 - Dateien der Handwerkskammern
 - Dateien der Industrie- und Handelskammern

Die Angaben aus diesen Datenquellen werden durch Rückläufe aus statistischen Bereichserhebungen, Registerumfragen, Gewerbeanzeigen und Internet- sowie Datenbankrecherchen ergänzt (Statistisches Bundesamt 2011).

Die Angaben im URS werden auf Jahresebene berichtet und sind aufgrund eindeutig vergebenen Identifikatoren als Paneldatensatz verknüpfbar. Dies wird auch von den FDZ der Statistischen Ämter im Rahmen des Projekts AFiD als AFiD-Panel Unternehmensregister durchgeführt und angeboten (FDZ der Statistischen Ämter 2010c).⁴⁰ Derzeit deckt das URS die Jahre 2002 bis 2010 ab. Die Zeitscheibe 2011 wird für das vierte Quartal 2013 erwartet. Da die Aufbereitung des URS aus den unterschiedlichen Statistiken sehr aufwändig ist, liegt der aktuelle Rand mit etwa zwei Jahren vergleichsweise weit in der Vergangenheit.

³⁹ Ausgenommen sind Landwirtschaftliche Betriebe.

⁴⁰ Da die Voraussetzungen und Gegebenheiten für das AFiD-Panel identisch sind, wird hier nicht näher auf das AFiD-Panel eingegangen.

5.1.3.2 Verfügbarkeit

Das URS ist ausschließlich über einen On-Site-Zugang (Datenfernverarbeitung oder Gastwissenschaftlerarbeitsplatz) nutzbar. Der Beantragungsweg entspricht dem anderer Produkte der FDZ der Statistischen Ämter und wird im Unterabschnitt 3.2 näher beschrieben. Die Dauer der Beantragung ist sehr stark von der Sensibilität der beantragten Merkmale abhängig.

5.1.3.3 Merkmale zur Mindestlohnevaluation

Das URS hat als Registerdatensatz wenige inhaltliche Merkmale, was die Analysemöglichkeiten stark einschränkt. Das URS erlaubt anhand der gleichzeitigen Verfügbarkeit von Beschäftigten- und Umsatzdaten Analysen zur Mindestlohnwirkung unter Berücksichtigung beider Merkmale (IAW 2011a, IAW 2011b).

5.1.3.4 Merkmale zur Verknüpfung

Auch wenn das URS aufgrund fehlender inhaltlicher Merkmale für direkte Analysen ungeeignet scheint, ist es in der Datenlandschaft ein zentrales Element, da es die Verknüpfung verschiedener Datenquellen anhand unterschiedlicher Identifikatoren ermöglicht.

Wie bereits erwähnt, dient das URS als Ziehungsgrundlage für die meisten Erhebungen auf Betriebs- oder Unternehmensebene der Statistischen Ämter. Dementsprechend ist die Identifikationsnummer des URS in den meisten Betriebs- und Unternehmensdatensätzen vertreten, was sowohl eine direkte Verknüpfung verschiedener Datensätze als auch eine Verknüpfung mit dem URS ermöglicht.

Gleichzeitig enthält das URS die Identifikationsnummern aus sämtlichen Datenquellen, die für die Erstellung des URS verwendet wurden. Dies macht eine Verknüpfung mit dem Datenangebot des FDZ der BA am IAB möglich, da beispielsweise die Betriebsnummer der BA im URS erfasst ist. Weiterhin sind die Nummern im Handelsregister, die Mitgliedsnummern der Handwerkskammer oder der Industrie- und Handelskammer sowie die Identifikationsnummer von Creditreform bekannt. Dies ermöglicht grundsätzlich die Verknüpfung mit externen Datenbeständen wie dem MUP oder den Daten von Bureau van Dijk. Da auch diverse Steuernummern dokumentiert sind, ist auch eine Berücksichtigung detaillierter Informationen aus dem Steuerverfahren technisch möglich.

5.1.3.5 Identifikation des Evaluationspotenzials

Das isolierte Evaluationspotenzial des URS ist wie bereits erwähnt sehr begrenzt, da es neben der Beschäftigtenanzahl und dem Umsatz kaum inhaltliche Variablen enthält. Für die Forschung allgemein besitzt das URS eine Schlüsselfunktion, da es die Verknüpfung des Datenangebots der FDZ der Statistischen Ämter mit den Datenangeboten des FDZ der BA am IAB, den Daten der Steuerbehörden sowie mit weiteren nichtamtlichen Datenquellen er-

möglich. Bisher konnten die technischen Verknüpfungsmöglichkeiten des URS aus Datenschutzgründen allerdings noch nicht in voller Form genutzt werden.

5.2 Amtliche Befragungsdaten

5.2.1 Jahreserhebung einschließlich der Investitionserhebung im Bauhaupt- sowie im Ausbaugewerbe

5.2.1.1 Allgemeine Datensatzinformationen

Die statistischen Ämter führen als Input für die Volkswirtschaftliche Gesamtrechnung die Jahreserhebung einschließlich der Investitionserhebung im Bauhauptgewerbe (bzw. im Ausbaugewerbe) durch. Es handelt sich hierbei um Befragungen mit Auskunftspflicht.

Grundgesamtheit bilden die Unternehmen im deutschen Bauhauptgewerbe (bzw. im Ausbaugewerbe) mit mindestens 20 beschäftigten Personen. Zugrunde gelegt wird dabei die Beschäftigtenzahl Ende September des Berichtsjahres. Als Unternehmen gilt die kleinste rechtliche Einheit, die aus handels- und/oder steuerrechtlichen Gründen Bücher führt und bilanziert.

Das Bauhauptgewerbe umfasst die Gruppen 45.1 „Vorbereitende Baustellenarbeiten“ und 45.2 „Hoch- und Tiefbau“ der Klassifikation der Wirtschaftszweige 2003 (WZ03). Das Ausbaugewerbe umfasst die Gruppen 45.3 „Bauinstallation“, 45.4 „Sonstiges Ausbaugewerbe“ und 45.5 „Vermietung von Baumaschinen und -geräten mit Bedienungspersonal“ der WZ03.

2003 erfolgte eine Neustrukturierung der Erhebung, u.a. mit einer Neudefinition der Grundgesamtheit anhand der WZ03 anstatt der WZ93. Die Änderungen sind sowohl für das Bauhauptgewerbe als auch das Ausbaugewerbe gering, so dass von einer Vergleichbarkeit über die Jahre ausgegangen werden kann (FDZ der Statistischen Ämter 2009a, 2009b).

5.2.1.2 Verfügbarkeit

Die Jahreserhebung einschließlich der Investitionserhebung im Bauhauptgewerbe (bzw. im Ausbaugewerbe) sind Produkte der FDZ der Statistischen Ämter, die ausschließlich über die On-Site-Nutzung zur Verfügung stehen. Die Beantragung erfolgt auf Basis des in Unterabschnitt 3.2 beschriebenen Vorgehens. Bisher stehen sechs Wellen zwischen 2003 und 2008 zur Verfügung. Metadaten sind über das Metadatenystem der FDZ zu erhalten.

5.2.1.3 Merkmale zur Mindestlohnevaluation

Die Jahreserhebung einschließlich der Investitionserhebung im Bauhauptgewerbe (bzw. im Ausbaugewerbe) enthält Informationen zur Jahresbauleistung in In- und Ausland, zu Investitionen in und den Verkauf von Sachanlagen, zur Gesamtbeschäftigung sowie zu Löhnen und Gehältern.

5.2.1.4 Merkmale zur Verknüpfung

Über die Unternehmensnummer ist sowohl eine Verknüpfung mit der VSE (Datensatzbeschreibung siehe Unterabschnitt 4.2.1) als auch mit dem URS (Datensatzbeschreibung siehe Unterabschnitt 5.1.3) technisch möglich. Insbesondere die Verknüpfung mit dem URS ermöglicht eine Vielzahl weiterer technischer Verknüpfungsmöglichkeiten.

Der Regionalschlüssel bis auf Gemeindeebene und der Wirtschaftszweig ermöglichen eine Verknüpfung der Daten auf diesen aggregierten Ebenen. Dadurch können eine Vielzahl anderer Statistiken zugespielt und auch Makroanalysen durchgeführt werden.

5.2.1.5 Identifikation des Evaluationspotenzials

Die Jahresherhebung einschließlich der Investitionserhebung im Bauhauptgewerbe (bzw. im Ausbaugewerbe) deckt nicht den Zeitpunkt der Einführung des Mindestlohns im Bauhauptgewerbe ab. Damit sind lediglich Erhöhungen des Mindestlohns analysierbar. Darüber hinaus sind sie nur für Kontrollgruppenansätze einsetzbar, falls man bereit ist, das Ausbaugewerbe als akzeptable Kontrollgruppe zu werten. Durch technische Verknüpfungsmöglichkeiten ist die Datengrundlage hier aber vermutlich erweiterbar.

Schließlich erlaubt die Nichtberücksichtigung kleiner Unternehmen mit weniger als 20 Beschäftigten keine Rückschlüsse auf die Auswirkungen des Mindestlohns in weiten Teilen des Bauhauptgewerbes. Dadurch können keine repräsentativen Aussagen für die Branche insgesamt ermittelt werden.

5.2.2 Pflegestatistik

5.2.2.1 Allgemeine Datensatzinformationen

Seit 1999 wird jeweils im Zweijahresrhythmus die Pflegestatistik erhoben. Die Pflegestatistik setzt sich aus drei unterschiedlichen Statistiken zusammen:

- Statistik über ambulante Pflegeeinrichtungen – Pflegedienste
- Statistik über stationäre Pflegeeinrichtungen – Pflegeheime
- Statistik über die Empfänger/innen von Pflegeleistungen

Für die Mindestlohnforschung haben ausschließlich die beiden ersten Statistiken Relevanz. Daher beziehen sich die folgenden Angaben ausschließlich auf diese beiden Statistiken. Zur besseren Lesbarkeit wird allgemein von Pflegestatistik gesprochen. Damit sind beide Statistiken gemeint.

Zurzeit liegen sieben Wellen der Pflegestatistik vor. Die aktuellste Welle wurde im Frühjahr 2012 mit dem Stichtag 15. Dezember 2011 durchgeführt. Die 2013 durchgeführte Erhebung wird voraussichtlich im März 2015 verfügbar sein. Die Pflegestatistik ist eine Vollerhebung

der zugelassenen ambulanten und stationären Pflegeeinrichtungen⁴¹ mit Auskunftspflicht, die je nach Statistischem Landesamt entweder ausschließlich postalisch oder zusätzlich mit der Option eines Online-Uploads durchgeführt wird (Statistisches Bundesamt 2007).

Im Unterschied zu anderen Erhebungen der Statistischen Ämter wird die Grundgesamtheit nicht anhand von Merkmalen im Unternehmensregister (URS) definiert, sondern ergibt sich aus der Weitergabe der Kontaktdaten der zugelassenen Pflegeeinrichtungen durch die Pflegekassen an die Statistischen Landesämter (IAW 2011c, S. 16).

5.2.2.2 Verfügbarkeit

Die Pflegestatistik ist ausschließlich über einen On-Site-Zugang (Datenfernverarbeitung oder Gastwissenschaftlerarbeitsplatz) nutzbar. Der Beantragungsweg entspricht dem anderer Produkte der FDZ der Statistischen Ämter und wird im Unterabschnitt 3.2 näher beschrieben. Da es für die Pflegestatistik ein Standardprodukt für Forschungszwecke gibt, ist die Dauer der Beantragung relativ kurz. Das entsprechende Anonymisierungskonzept ist bereits erstellt und abgestimmt.

5.2.2.3 Merkmale zur Mindestlohnevaluation

Die Pflegestatistik enthält eine Reihe von Merkmalen auf Einrichtungsebene, die beispielsweise zur Differenzierung der Einrichtungen nach Trägerform genutzt werden können. Zusätzlich sind für stationäre Einrichtungen detaillierte Informationen zu den Pflegesätzen, d.h. zu den Zahlungen der Pflegekassen an die Einrichtungen, vorhanden.

Auf Ebene der Beschäftigten sind sowohl soziodemografische Merkmale als auch Variablen zum Beschäftigungsverhältnis sowie zur ausgeübten Tätigkeit für sämtliche Beschäftigte verfügbar. Keine Angaben liegen hingegen zur Entlohnung vor.

5.2.2.4 Merkmale zur Verknüpfung

Da die Pflegestatistik nicht das URS als Basis hat, ist eine Verknüpfung der Daten mit dem URS schwierig. Da aber die Kontaktdaten vorliegen, ist eine Verknüpfung über Record Linkage technisch möglich. Insgesamt enthält die Pflegestatistik wenige Angaben, die für eine Verknüpfung mit anderen Datenquellen genutzt werden können. Neben der beschriebenen Verknüpfungsmöglichkeit über den Adressbestand ist eine Verknüpfung der Daten anhand des amtlichen Gemeindeschlüssels technisch möglich.

⁴¹ Eine Pflegeeinrichtung hat eine Zulassung, wenn sie entweder mit den Pflegekassen einen Versorgungsvertrag nach § 75 SGB XI abgeschlossen hat oder nach § 73 Abs. 3 & 4 SGB XI Bestandsschutz hat (Statistisches Bundesamt 2007).

5.2.2.5 Identifikation des Evaluationspotenzials

Das Evaluationspotenzial der Pflegestatistik für die Mindestlohnforschung ist sehr stark eingeschränkt. Als Fachstatistik im Pflegebereich ist der Geltungsbereich des Mindestlohns fast deckungsgleich mit der Grundgesamtheit in der Pflegestatistik (IAW 2011c). Da weiterhin weder Angaben zur Entlohnung noch ausführliche Informationen zur Arbeitszeit vorliegen, fehlen zwei entscheidende Merkmale für die Mindestlohnforschung.

Die vorhandenen Angaben können allerdings genutzt werden, um eine Veränderung in der Vergütung der Einrichtungen, Veränderungen in der Mitarbeiterstruktur oder hinsichtlich der Gründungen und Schließungen zu untersuchen. Ohne entsprechende Informationen von Einrichtungen oder Betrieben, die als Vergleichsgruppe dienen können, beschränken sich die Auswertungsmöglichkeiten ausschließlich auf deskriptive Analysen.

Eine Verknüpfung der Pflegestatistik mit dem URS würde das Evaluationspotenzial deutlich erhöhen, da dadurch die Pflegebranche in anderen Datenquellen trennscharf abgegrenzt werden kann, was derzeit anhand der Wirtschaftszweige nur sehr schwer möglich ist.⁴²

5.3 Amtliche Firmendaten für Deutschland (AFiD)

Mit dem Projekt Amtliche Firmendaten für Deutschland (AFiD) wollen die FDZ der Statistischen Ämter eine Datenbasis schaffen, bei der alle Mikrodaten der Wirtschafts- und Umweltstatistiken verknüpft werden. Dies erweitert das bisherige Analysepotenzial der unverknüpften Einzelstatistiken, da Merkmale aus unterschiedlichen Statistiken genutzt werden können (FDZ der Statistischen Ämter, 2012a).

Bisher sind AFiD-Paneldaten für folgende Bereiche verfügbar:

- Agrarstruktur
- Industriebetriebe
- Industrieunternehmen
- Energiebetriebe
- Energieunternehmen
- Unternehmensregister
- Handel
- Gastgewerbe
- Dienstleistungen
- Körperschaftssteuer

⁴² Für eine genauere Analyse dieser Problematik siehe (IAW 2011c).

Für die Mindestlohnforschung sind die AFiD-Paneldaten Industriebetriebe (Unterabschnitt 5.3.1) sowie Dienstleistungen (Unterabschnitt 5.3.2) relevant. Das AFiD-Panel Unternehmensregister wurde bereits im Unterabschnitt 5.1.3 zum Unternehmensregister – System 95 vorgestellt, da es sich dabei im Unterschied zu den drei anderen Paneldatensätzen um keine Kombination verschiedener Datenquellen handelt, sondern ausschließlich um eine Zusammenführung der Jahrgänge des URS seit 2002.

5.3.1 AFiD-Panel Industriebetriebe

5.3.1.1 Allgemeine Datensatzinformationen

Das AFiD-Panel Industriebetriebe setzt sich aus insgesamt fünf verschiedenen Erhebungen mit Auskunftspflicht zusammen, die auf Jahresebene miteinander verknüpft werden. Bei den Erhebungen handelt es sich um folgende Statistiken:

- Monatsbericht für Betriebe im Bereich Verarbeitendes Gewerbe, Bergbau und Gewinnung von Steinen und Erden
- Investitionserhebung im Bereich Verarbeitendes Gewerbe, Bergbau und Gewinnung von Steinen und Erden
- Vierteljährliche Produktionserhebung im Bereich Verarbeitendes Gewerbe, Bergbau und Gewinnung von Steinen und Erden
- Erhebung für industrielle Kleinbetriebe im Bereich Verarbeitendes Gewerbe, Bergbau und Gewinnung von Steinen und Erden
- Jahresbericht für Betriebe im Bereich Verarbeitendes Gewerbe, Bergbau und Gewinnung von Steinen und Erden

Die unterjährigen Erhebungen (Monatsbericht und Vierteljährliche Produktionserhebung) werden für die Verknüpfung auf Jahresangaben aggregiert. Die Verknüpfung wird über die eindeutig vergebene Betriebsnummer durchgeführt.

Zur Grundgesamtheit gehören sämtliche Betriebe, d.h. örtlich abgegrenzte Produktionseinheiten, mit Betriebsort in Deutschland, die ausschließlich oder überwiegend Tätigkeiten im Bereich Verarbeitendes Gewerbe, Bergbau und Gewinnung von Steinen und Erden ausüben. Dies sind die Wirtschaftszweige der Abschnitte C und D der Klassifikation der Wirtschaftszweige 2003 (WZ03). Damit sind auch Betriebe aus den Abschnitten C und D der WZ03, die zu Unternehmen gehören, die einem Wirtschaftszweig außerhalb der beiden genannten Abschnitte zugerechnet werden, explizit mit berücksichtigt (FDZ der Statistischen Ämter 2010b).

Bei allen genannten Erhebungen handelt es sich um Vollerhebungen mit Abschneidegrenzen. Diese Abschneidegrenze ist unterschiedlich, je nachdem ob es sich um einen Betrieb von Unternehmen aus dem Bereich Verarbeitendes Gewerbe, Bergbau und Gewinnung von Steinen und Erden oder um einen Betrieb mit dem Tätigkeitsschwerpunkt aus diesem Be-

reich, der aber zu einem Unternehmen mit einem Tätigkeitsschwerpunkt aus einem anderen Bereich gehört,⁴³ handelt. Gleichzeitig gibt es in einigen ausgewählten kleinbetrieblich strukturierten Wirtschaftszweigen niedrigere Abschneidegrenzen (FDZ der Statistischen Ämter 2010b, S. 4).⁴⁴

Im Monatsbericht, der Investitionserhebung und der Produktionserhebung werden Betriebe von Unternehmen aus den Abschnitten C und D der WZ03 berücksichtigt, die 20 oder mehr Beschäftigte im Unternehmen haben. Betriebe von Unternehmen aus anderen Wirtschaftszweigen werden nur berücksichtigt, wenn der Betrieb 20 oder mehr Personen beschäftigt. Ab 2007 werden im Monatsbericht nur noch Betriebe mit 50 oder mehr Beschäftigten erfasst. Kleinere Betriebe werden seitdem im Jahresbericht erfasst. Die Erhebung industrieller Kleinbetriebe umfasste Betriebe mit weniger als 20 Beschäftigten, allerdings keine Betriebe des verarbeitenden Handwerks.

5.3.1.2 Verfügbarkeit

Das AFiD-Panel Industriebetriebe ist ein Produkt der FDZ der Statistischen Ämter, das ausschließlich über die On-Site-Nutzung zur Verfügung steht. Die Beantragung erfolgt auf Basis des in Unterabschnitt 3.2 beschriebenen Vorgehens. Bisher stehen 16 Wellen zwischen 1995 und 2010 zur Verfügung, wobei nicht alle Statistiken über den kompletten Zeitraum enthalten sind. Die Erhebung für industrielle Kleinbetriebe wurde 2002 eingestellt. Der Jahresbericht wird erst seit 2007 erhoben. Tabelle 5.1 zeigt die Verfügbarkeit der einzelnen Statistik im AFiD-Panel Industriebetriebe.

Im Unterschied zu anderen Datenquellen der FDZ der Statistischen Ämter gibt es beim AFiD-Panel Industriebetriebe keine vorgeschriebenen Zeitpunkte der Aktualisierungen. Nach Auskunft der FDZ der Statistischen Ämter laufen zurzeit die Vorbereitungsarbeiten für eine Aktualisierung des AFiD-Panels Industriebetriebe. Es ist geplant, das Panel im Frühjahr 2014 um das Jahr 2011 und mit Ausnahme der Kostenstrukturerhebung⁴⁵ um das Jahr 2012 zu erweitern.

Derzeit wird das AFiD-Panel Industriebetriebe nicht als eigenständige Statistik gewertet, sondern als Produkt mehrerer Erhebungen. Dies hat zur Konsequenz, dass die Kosten der Beantragung für jede Einzelstatistik anfallen. Somit liegen die Kosten pro Erhebungsjahr zwischen 750 und 1.000 Euro, wenn alle Einzelstatistiken genutzt werden.

⁴³ Ausgenommen sind Baubetriebe und Betriebe der Energie- und Wasserversorgung.

⁴⁴ Die abweichenden Abschneidegrenzen sind abhängig von den Statistiken und können in den jeweiligen Metadaten und Qualitätsberichten der Einzelstatistiken eingesehen werden.

⁴⁵ Die Kostenstrukturerhebung liegt erst 18 Monate nach dem Berichtsjahr mit verlässlichen Ergebnissen für Aufbereitungsarbeiten an den FDZ der Statistischen Ämter vor.

Tabelle 5.1: Zeitliche Verfügbarkeit der Einzelerhebungen

Statistik	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09	10
Monatsbericht																
Investitionserhebung																
Vierteljährliche Produktionserhebung																
Erhebung für industrielle Kleinbetriebe																
Jahresbericht																

Quelle: Eigene Darstellung auf Basis von FDZ der Statistischen Ämter (2010b).

5.3.1.3 Merkmale zur Mindestlohnevaluation

Durch die Zusammenführung verschiedener Statistiken auf Betriebsebene ist das AFiD-Panel Industriebetriebe ein sehr umfangreicher und detaillierter Datensatz mit Informationen zu Beschäftigung, Umsatz und Investitionsaktivitäten. Diese können bis auf Gemeindeebene differenziert werden (FDZ der Statistischen Ämter 2010b).

5.3.1.4 Merkmale zur Verknüpfung

Da das AFiD-Panel Industriebetriebe bereits eine Verknüpfung verschiedener Datensätze ist, wurden einige Verknüpfungsmöglichkeiten bereits ausgeschöpft. Über die Betriebsnummer ist sowohl eine Verknüpfung mit der VSE (Datensatzbeschreibung siehe Unterabschnitt 4.2.1) als auch mit dem URS (Datensatzbeschreibung siehe Unterabschnitt 5.1.3) technisch möglich. Insbesondere die Verknüpfung mit dem URS ermöglicht eine Vielzahl weiterer technischer Verknüpfungsmöglichkeiten sowohl auf Betriebs- als auch auf Unternehmensebene.

5.3.1.5 Identifikation des Evaluationspotenzials

Das AFiD-Panel Industriebetriebe bietet hinsichtlich der Merkmale eine ideale Grundlage für eine Analyse der Auswirkungen von Mindestlohneinführungen. Gleichzeitig deckt der verfügbare Zeitraum alle Mindestlohneinführungen nach dem AEntG ab. Die Aufteilung sowohl nach Wirtschaftszweigen als auch nach Regionen ist sehr feingliedrig, was detaillierte Auswertungen ermöglicht. Durch die weitergehenden technischen Verknüpfungsmöglichkeiten ist die Datengrundlage zudem sehr gut erweiterbar.

Trotz dieser sehr guten Voraussetzungen ist das Evaluationspotenzial stark eingeschränkt. Die Nichtberücksichtigung kleiner Betriebe mit weniger als 20 Beschäftigten erlaubt keine Rückschlüsse auf die Auswirkungen des Mindestlohns in diesen Betrieben. Insbesondere in den Handwerksbranchen, die sich größtenteils aus Betrieben mit weniger als 20 Beschäftig-

ten zusammensetzen, können dadurch keine repräsentativen Aussagen für die Branche getätigt werden.

Zusätzlich schließt die fachliche Abgrenzung des AFiD-Panels Industriebetriebe auf die Abschnitte C und D der WZ03 die meisten Mindestlohnbranchen aus, da sich diese hauptsächlich im Abschnitt F (Baugewerbe)⁴⁶ befinden. Lediglich ein kleiner Teil der Abfallwirtschaft (37.00.0: Recycling, WZ03) sowie die Steinmetze (26.70.1, WZ03) sind Bestandteil der beiden integrierten Abschnitte.

5.3.2 AFiD-Panel Dienstleistungen

5.3.2.1 Allgemeine Datensatzinformationen

Das AFiD-Panel Dienstleistungen ist die Längsschnittverknüpfung der Strukturhebung im Dienstleistungsbereich. Zur Grundgesamtheit gehören sämtliche Unternehmen und Einrichtungen zur Ausübung einer freiberuflichen Tätigkeit mit einem Umsatz von mehr als 17.500 Euro (FDZ der Statistischen Ämter 2010a). Dies sind die Wirtschaftszweige der Abschnitte I (Verkehr und Nachrichtenübermittlung) und K (Grundstücks- und Wohnungswesen, Vermietung beweglicher Sachen, Erbringung von wirtschaftlichen Dienstleistungen, anderweitig nicht genannt) der Klassifikation der Wirtschaftszweige 2003 (WZ03).

Die Strukturhebung im Dienstleistungsbereich ist eine repräsentative Befragung mit Auskunftspflicht. Die Stichprobe wurde aus dem URS (Unterabschnitt 5.1.3) gezogen und ist nach Bundesländern, Branchengruppen und Umsatzgrößenklassen geschichtete Zufallsstichprobe mit einem Auswahlsatz von 15 %. Die Befragung wurde 2003 als Panel angelegt. Bis einschließlich 2007 wurde die Stichprobe jährlich durch eine Neuzugangsstichprobe aus den seit der letzten Ziehung neu im Register auftauchenden Unternehmen/Einrichtungen ergänzt. Dadurch ist eine Repräsentativität für alle Jahre gegeben.

2008 erfolgte eine Neustrukturierung der Erhebung, u.a. mit einer Neudefinition der Grundgesamtheit anhand der Wirtschaftszweigklassifikation 2008 (WZ08). Durch die Neudefinition verteilt sich die im Rahmen der Strukturhebung im Dienstleistungsgewerbe definierte Grundgesamtheit auf die fünf Abschnitte H (Verkehr und Lagerei), J (Information und Kommunikation), L (Grundstücks- und Wohnungswesen), M (Erbringung von freiberuflichen, wissenschaftlichen und technischen Dienstleistungen) und N (Erbringung von sonstigen wirtschaftlichen Dienstleistungen) sowie auf die Abteilung S/95 (Reparatur von Datenverarbeitungsgeräten und Gebrauchsgütern). Dies ist auch eine inhaltliche Erweiterung der Grundgesamtheit, da die meisten Einrichtungen der Abteilung S/95 (WZ08) nach der alten Definition (WZ03) in den Abschnitten D (Verarbeitendes Gewerbe) und G (Handel; Instandhaltung und Reparatur von Kraftfahrzeugen und Gebrauchsgütern) vertreten waren und somit nicht Teil der Strukturhebung im Dienstleistungsgewerbe. Zudem beinhaltet der Abschnitt J (In-

⁴⁶ Bauhauptgewerbe, Dachdeckerhandwerk, Elektrohandwerk, Maler- und Lackiererhandwerk und Gerüstbauergewerbe

formation und Kommunikation) der WZ08 einige Einrichtungen, die zuvor dem Abschnitt O (Erbringung von sonstigen öffentlichen und persönlichen Dienstleistungen) der WZ03 zugeordnet waren.

Aufgrund dieser Neustrukturierung wurde eine vollständig neue Stichprobe gezogen, wobei die Schichtungsmerkmale beibehalten wurden. Diese Stichprobe wurde bis einschließlich 2010 genutzt. Zum Berichtsjahr 2011 wurde wieder eine neue Stichprobe gezogen (Statistisches Bundesamt 2013e). Nach Auskunft der FDZ der Statistischen Ämter wird mit der neuen Stichprobe vermutlich ein neuer AFiD-Paneldatensatz begonnen.

5.3.2.2 Verfügbarkeit

Das AFiD-Panel Dienstleistungen ist ein Produkt der FDZ der Statistischen Ämter, das ausschließlich über die On-Site-Nutzung zur Verfügung steht. Die Beantragung erfolgt auf Basis des in Unterabschnitt 3.2 beschriebenen Vorgehens. Bisher stehen für 2003 bis 2007 fünf Wellen als ein Datenprodukt sowie die Wellen von 2008 bis 2010 als zweites Datenprodukt zur Verfügung. Da das AFiD-Panel Dienstleistungen aus einer Statistik besteht, entstehen im Unterschied zum AFiD-Panel Industriebetriebe keine höheren Kosten pro Erhebungsjahr.

Wie beim AFiD-Panel Industriebetriebe (Unterabschnitt 5.3.1) ist eine Aktualisierung des AFiD-Panels Dienstleistungen für Frühjahr 2014 geplant. Allerdings beträgt der Aktualisierungszeitraum hier lediglich ein Jahr, da die Strukturhebung im Dienstleistungsgewerbe erst 18 Monate nach dem Berichtsjahr mit einer verlässlichen Datengrundlage für Aufbereitungszwecke vorliegt.

5.3.2.3 Merkmale zur Mindestlohnevaluation

Das AFiD-Panel Dienstleistungen enthält umfangreiche Informationen zu Umsatz, Einnahmen und Investitionen der befragten Einrichtungen. Zudem existieren Angaben zum Personalbestand sowie den Personalkosten (FDZ der Statistischen Ämter 2010a).

5.3.2.4 Merkmale zur Verknüpfung

Grundlage für die Stichprobenziehung der Strukturhebung im Dienstleistungsbereich ist das URS. Dementsprechend ist eine Verknüpfung mit dem URS technisch möglich. Dies ermöglicht eine weitere Verknüpfung mit einer Vielzahl weiterer Datenquellen (siehe dazu Unterabschnitt 5.1.3).

5.3.2.5 Identifikation des Evaluationspotenzials

Das AFiD-Panel Dienstleistungen bietet hinsichtlich der Merkmale eine gute Grundlage für eine Analyse der Auswirkungen von Mindestlohneinführungen auf Betriebsebene. Die Aufteilung sowohl nach Wirtschaftszweigen als auch nach Regionen ist sehr feingliedrig, was detaillierte Auswertungen ermöglicht. Durch die weitergehenden technischen Verknüpfungsmöglichkeiten ist die Datengrundlage zudem sehr gut erweiterbar.

Trotz dieser guten Voraussetzungen ist das Evaluationspotenzial stark eingeschränkt. Die fachliche Abgrenzung erlaubt lediglich Untersuchungen für die Mindestlohnbranchen Gebäudereinigung und Sicherheitsdienstleistungen. Die Einführung des allgemeinverbindlichen Mindestlohns nach dem AEntG in der Gebäudereinigung erfolgte im Jahr 2008, das den Zeitpunkt des strukturellen Bruchs in den Daten darstellt. Dadurch müssen die Gebäudereinigungen anhand zweier unterschiedlicher Wirtschaftszweigklassifikationen abgegrenzt werden. Zusätzlich sind keine Panelanalysen möglich, da 2008 eine neue Stichprobe gezogen wurde. Vor der Verwendung der Daten muss daher überprüft werden, ob trotz des strukturellen Bruchs in den Daten eine Datengrundlage für eine Mindestlohnevaluation geschaffen werden kann.

Ähnlich gelagerte Probleme treten bei den Sicherheitsdienstleistungen auf. Die erstmalige Einführung eines Mindestlohns nach dem AEntG (2011) liegt derzeit außerhalb des verfügbaren Zeitraums. Gleichzeitig ist bereits bekannt, dass 2011 eine neue Stichprobe gezogen wurde (Statistisches Bundesamt 2013e). Auch hier müssen die Daten geprüft werden, ob trotz neuer Stichprobenziehung Analysen durchführbar sind.

5.4 Nichtamtliche Daten

5.4.1 IAB-Betriebspanel (IAB-BP)

5.4.1.1 Allgemeine Datensatzinformationen

Das IAB-Betriebspanel (IAB-BP) ist eine jährliche Betriebsbefragung, die in Zusammenarbeit mit TNS Infratest Sozialforschung durchgeführt wird. Sie wurde zunächst 1993 für Westdeutschland eingeführt, umfasst aber seit 1996 das gesamte Bundesgebiet. Die Befragung ist bundesweit für alle Branchen und Betriebsgrößen repräsentativ, auf Bundeslandebene seit 2000. Das Panel ist als eine Wiederholungsbefragung ausgelegt, so dass jedes Jahr ein Großteil derselben Betriebe wieder befragt wird. Daher können Entwicklungen im Zeitverlauf durch den Vergleich von Querschnittsdaten über die Jahre oder aber auch als betriebsindividuelle Längsschnittuntersuchungen analysiert werden.

Die Grundgesamtheit des IAB-BP umfasst alle Betriebe mit mindestens einem sozialversicherungspflichtig Beschäftigten zum 30.06. des Vorjahres. Datengrundlage bildet hierfür die Betriebsdatei der BA, die aus den meldepflichtigen Angaben der Arbeitgeber zur Beschäftigtenstatistik hervorgeht und etwa 2 Millionen Betriebe enthält.

Der Datensatz basiert auf einer nach Betriebsgröße, Branche und Bundesland disproportional geschichteten Stichprobe. Dadurch sind große Betriebe, kleine Bundesländer und kleine Branchen sowie das Verarbeitende Gewerbe in Ostdeutschland überrepräsentiert. Um repräsentative Aussagen über die Grundgesamtheit treffen zu können, müssen daher bei deskriptiven Auswertungen entsprechende Gewichtungsfaktoren verwendet werden. Durch diese Hochrechnung sind regionale, branchen- und größendifferenzierte Auswertungen möglich. Unterhalb dieser Dimensionen der Schichtungsmatrix sind valide Aussagen nur einge-

schränkt möglich. Aufgrund der Änderungen in der Wirtschaftszweigsystematik sind vergleichende Branchenanalysen dabei gewissen zeitlichen Einschränkungen unterworfen. Insbesondere im Jahr 2000 ergibt sich hier ein Bruch. Außerdem kann seit 2007 nicht mehr zwischen West- und Ostberlin unterschieden werden, so dass Berlin seither vollständig zu Ostdeutschland gezählt wird.

Der Fragebogen setzt sich aus mehreren Blöcken zusammen. Ein Teil der Fragen wird jedes Jahr gestellt, andere Themenbereiche werden in regelmäßigen Abständen wiederholt. Hinzu kommen einmalige Schwerpunktbefragungen, die insbesondere aktuell relevante Themenkomplexe umfassen. Die Befragungen finden im dritten Quartal eines Jahres statt. Die Interviews sind dabei zumeist persönlich-mündlich, in einigen Bundesländern aber auch schriftlich-postalisch. Dabei zeigen sich für wiederholt befragte Betriebe und persönliche Befragungen höhere Rückläufe als für neue Betriebe und schriftliche Befragungen (Janik/Kohaut 2012).

Durch die finanzielle Beteiligung der Bundesländer konnte die Stichprobe 2000 deutlich aufgestockt werden. Dies ermöglicht nicht nur Auswertungen auf Bundeslandebene, sondern auch ein erhöhtes Analysepotenzial auf Bundesebene. Mittlerweile werden im IAB-BP knapp 16.000 Betriebe zu einer Vielzahl beschäftigungsrelevanter Themen befragt.

5.4.1.2 Verfügbarkeit

Das IAB-BP wird durch das FDZ der BA im IAB angeboten. Das FDZ übernimmt auch die Betreuung externer Nutzer/innen. Der Datensatz ist sowohl über einen Gastaufenthalt oder über Datenfernverarbeitung nutzbar. Die Beantragung folgt gemäß den im Unterabschnitt 3.1 beschriebenen Voraussetzungen.

Die Dauer der Beantragung ist abhängig vom Datennutzungsweg und dauert bei Gastaufenthalt aufgrund des zusätzlichen Genehmigungsverfahrens etwas länger. Sicherheitshalber sollten 2 bis 4 Wochen veranschlagt werden. Die Freischaltung für die Datenfernverarbeitung erfolgt bei vollständigem Antrag meist innerhalb weniger Tage. Die Datennutzung selbst ist kostenlos.

Bestimmte Merkmale, die eine Deanonymisierung von Betrieben erleichtern, werden nur dann im Original weitergegeben, wenn es für das Analyseziel notwendig ist und dies im Antrag auf den Datenzugang explizit begründet wird. Diese aus datenschutzrechtlicher Sicht besonders sensiblen Merkmale umfassen Kreisinformationen und detaillierte Wirtschaftsunterklassen.⁴⁷

Das IAB-BP wird jedes Jahr aktualisiert und ist derzeit für den Zeitraum 1993 bis 2011 verfügbar. Das nächste Update ist für Ende 2013 geplant. Der abgedeckte Zeitraum soll dann

⁴⁷ Angaben zum Bundesland und vergrößerte Wirtschaftszweige werden immer weitergegeben.

bis 2012 erweitert werden. Umfangreiche Dokumentationen und Auszählungen sind auf der Internetpräsenz des FDZ verfügbar.

5.4.1.3 Merkmale zur Mindestlohnevaluation

Das IAB-BP enthält Angaben der Betriebe zu Beschäftigungsstruktur und Beschäftigungsentwicklung sowie Löhnen und Gehältern. Es wird aber auch eine Reihe von Themen abgefragt, die in den administrativen Datensätzen, die das IAB anbietet, nicht enthalten sind. Für die Mindestlohnforschung insbesondere von Bedeutung ist hierbei die Frage nach der Geltung von Branchen- und Firmentarifverträgen, betriebsspezifischer Mitarbeitervertretung, der betrieblichen Arbeitszeit und Personalsuche durch den Betrieb. Regelmäßig werden auch Fragen zur beruflichen Ausbildung gestellt. Um den Erfolg der Betriebe und ihre Planung für die Zukunft abzubilden, werden Fragen zu Geschäftsvolumen, Investitionsverhalten und Wettbewerbsdruck gestellt.

Das IAB-BP enthält einen Regionalschlüssel bis maximal auf Kreisebene und Angaben zum Wirtschaftszweig bis hin zur 5-Steller-Ebene. Diese Definition änderte sich im Zeitablauf mehrmals.

Bei den Angaben zum Arbeitsort ist insbesondere auch die Logik der Betriebsnummernvergabe zu beachten. Hierbei werden mehrere Arbeitsstätten des Arbeitgebers zu einem Betrieb zusammengefasst, wenn sie sich in derselben Gemeinde befinden und dieselbe wirtschaftsfachliche Zuordnung aufweisen. Dadurch werden Beschäftigte demselben Betrieb zugeordnet, auch wenn sie an unterschiedlichen Standorten innerhalb einer Gemeinde tätig sind. Für das IAB-BP bedeutet das, dass die durch das Befragungsinstitut kontaktierte Einheit immer diejenige ist, deren Kontaktdaten in der Betriebsdatei hinterlegt sind.

5.4.1.4 Merkmale zur Verknüpfung

Das IAB-BP enthält Betriebsidentifikatoren, die über Schlüssel Tabellen rein technisch Verknüpfungen mit der IEB und anderen Datenbeständen der BA zulassen. Dazu gehört beispielsweise auch Adressmaterial der Betriebe, wodurch ein Record Linkage mit anderen Datenquellen möglich wäre. Die Vertragsbestimmungen des FDZ lassen allerdings eine Verknüpfung des IAB-BP mit anderen Mikrodaten nicht zu. Eine Ausnahme bilden sowohl das BHP, aus dem in begründeten Fällen Zusatzmerkmale an das IAB-BP gespielt werden können, als auch eine mit der SIAB vergleichbare Ziehung von Individualdaten für die Betriebspanelbetriebe. Eine Kombination dieser drei Datensätze wird am FDZ unter dem Namen LIAB (Linked-Employer-Employee-Daten des IAB) angeboten.

Auf aggregierter Ebene ermöglichen der Regionalschlüssel sowie der Wirtschaftszweig eine Verknüpfung mit anderen Daten. Es ist aber zu beachten, dass eine Aggregation auf Ebenen, die nicht durch die Schichtungslogik abgedeckt sind, in der Regel nicht zu repräsentativen Ergebnissen führt.

5.4.1.5 Identifikation des Evaluationspotenzials

Das IAB-BP deckt eine Reihe von Themenbereichen ab, die von den anderen Datenprodukten des IAB nicht abgedeckt werden, unter anderem zu Personalsuche, Wettbewerb und Investitionen. Insofern kann es bei der Mindestlohnforschung als sinnvolle Ergänzung zu diesen Produkten betrachtet werden. Für branchenspezifische Untersuchungen ist einschränkend anzumerken, dass die Schichtungsmatrix auf Branchenebene für viele relevante Branchen zu grob ist, und daher nicht von einer repräsentativen Stichprobe gesprochen werden kann. Darüber hinaus kann der Grad der Betroffenheit von einem Mindestlohn anhand der im IAB-BP vorhandenen Variablen kaum approximiert werden. Hier kann aber die Kombination mit Individualdaten im Rahmen des LIAB helfen, wobei die Beschränkungen gelten, die in Abschnitt 4.1.2 zum SIAB besprochen wurden.

Des Weiteren ist zu beachten, dass einige Merkmale im IAB-BP nicht jährlich abgefragt werden, was die Durchführung von Paneluntersuchungen erschwert. Sollte diese Datenstruktur für eine projektspezifische Untersuchung notwendig sein, muss also schon im Vorfeld geklärt werden, ob sie sich für den relevanten Zeitraum überhaupt erzeugen lässt.

5.4.2 Linked Personnel Panel Betriebsbefragung (LPP Betriebe)

5.4.2.1 Allgemeine Datensatzinformationen

Im Rahmen des Forschungsprojekts „Arbeitsqualität und wirtschaftlicher Erfolg“ wird derzeit das Linked Personnel Panel (LPP) aufgebaut. Der Datensatz entsteht in Kooperation zwischen IAB, ZEW, BMAS und der Universität Köln. Ziel des Projekts ist der Aufbau eines Linked Employer-Employee Datensatzes, mit dem sich die Auswirkung des Personalmanagements auf den Unternehmenserfolg sowie Arbeitszufriedenheit, Gesundheit und Unternehmensbindung der Mitarbeiter untersuchen lassen. Der Datensatz ist als Panel geplant und besteht aus einem Betriebsbefragungsteil (LPP Betriebe) und einem Beschäftigtenbefragungsteil (LPP Beschäftigte). In diesem Abschnitt wird die Betriebsbefragung besprochen.

Grundgesamtheit für die Betriebsbefragung ist wie beim IAB-BP die Betriebsdatei der BA (siehe Abschnitt 5.4.1). In die Auswahl gelangen aber nur Betriebe, die im selben Jahr für das IAB-Betriebspanel befragt werden, wobei einige zusätzliche Einschränkungen gelten. So wird die Ziehung auf Betriebe mit mehr als 50 sozialversicherungspflichtig Beschäftigten eingegrenzt und Betriebe aus der Land-, Forstwirtschaft und Fischerei, dem öffentlichen Dienst und mildtätige und kirchliche Organisationen werden zunächst ausgeschlossen.

Der Datensatz basiert auf einer nach vier Betriebsgrößenklassen und 5 Branchengruppen geschichteten Stichprobe. Um repräsentative Aussagen über die Grundgesamtheit treffen zu können, müssen daher bei deskriptiven Auswertungen entsprechende Gewichtungsfaktoren verwendet werden. Diese sollen sowohl die Disproportionalität der Stichprobe für das LPP als auch für das IAB-BP selbst ausgleichen. Dadurch sind branchen- und größendifferenzier-

te Auswertungen für Deutschland möglich. Unterhalb dieser Dimensionen der Schichtungsmatrix sind valide Aussagen nur eingeschränkt möglich.

Die Befragung findet im direkten Anschluss an die Befragung für das IAB-BP statt und wird vom selben Befragungsinstitut durchgeführt. Die Interviews sind dabei zumeist persönlich-mündlich, in einigen Bundesländern aber auch schriftlich-postalisch.

In der ersten Welle der Befragung wurde bei einer Bruttostichprobe von 2.222 Betrieben ein Rücklauf von 1.219 auswertbaren Interviews realisiert.

5.4.2.2 Verfügbarkeit

Das LPP wird derzeit aufgebaut und ist bislang noch nicht für externe Forscher zugänglich. Später soll der Datensatz durch das FDZ der BA im IAB angeboten werden. Ein Veröffentlichungstermin steht noch nicht fest.

5.4.2.3 Merkmale zur Mindestlohnevaluation

Das LPP Betriebe enthält Angaben der Betriebe zu Personalplanung und -beschaffung, Personalentwicklung und Unternehmenskultur. Für die Mindestlohnforschung könnten dabei insbesondere die Fragen zur Dauer von Stellenbesetzungen, internen Leistungsbeurteilungen und Maßnahmen der Arbeitnehmerbindung interessant sein.

5.4.2.4 Merkmale zur Verknüpfung

Das LPP Betriebe enthält Identifikatoren, die sowohl eine Verknüpfung mit dem LPP Beschäftigte als auch mit den anderen IAB-Datenprodukten, also dem IAB-BP, der IEB und dem BHP zulässt. Dadurch soll am Ende ein Linked Employer-Employee Datensatz entstehen.

Auf aggregierter Ebene ermöglichen der Regionalschlüssel sowie der Wirtschaftszweig eine Verknüpfung mit anderen Daten. Es ist aber zu beachten, dass eine Aggregation auf Ebenen, die nicht durch die Schichtungslogik abgedeckt sind, in der Regel nicht zu repräsentativen Ergebnissen führt. Da die Schichtungsgruppen beim LPP sehr grob sind, ist eine solche Aggregation in der Regel wohl nicht sinnvoll.

5.4.2.5 Identifikation des Evaluationspotenzials

Für sich allein genommen ist das Evaluationspotential des LPP Betriebe sehr eingeschränkt. Die Kombination mit anderen IAB-Produkten zum LPP ergibt jedoch deutlich verbesserte Analysemöglichkeiten, da über das LPP Beschäftigte beispielsweise der Stundenlohn der Beschäftigten ermittelt werden kann. Somit ist die Untersuchung der Zusammenhänge zwischen Mindestlöhnen und Arbeitsqualität bzw. Unternehmenskultur möglich.

Einschränkend ist anzumerken, dass sich das Panel erst im Aufbau befindet und noch nicht genutzt werden kann. Außerdem fand die erste Welle im Jahr 2012 statt, so dass bisherige

Mindestlohneinführungen bzw. -erhöhungen nicht untersucht werden können. Für Branchenanalysen ist zu beachten, dass die sehr grobe Schichtungsmatrix hier kaum repräsentative Untersuchungen zulässt. Daher wäre der Datensatz eher für die Untersuchung eines allgemeinen Mindestlohns nutzbar.

5.4.3 BIBB-Betriebspanel zu Qualifizierung und Kompetenzentwicklung

5.4.3.1 Allgemeine Datensatzinformationen

Das BIBB hat 2011 eine Panelstudie aufgesetzt, die sich mit der betrieblichen Qualifizierung und Kompetenzentwicklung auseinandersetzt. Ziel ist es einen Datensatz aufzusetzen, der eine kontinuierliche Untersuchung von Aus- und Weiterbildungsaktivitäten von Betrieben ermöglicht (Gerhards et al. 2013).

Zu diesem Zweck wurde aus der Grundgesamtheit aller Betriebe mit mindestens einer sozialversicherungspflichtig beschäftigten Person⁴⁸ eine anhand der Beschäftigtenzahl disproportionale, geschichtete Zufallsstichprobe gezogen. Zusätzlich wurden Betriebe in Ostdeutschland und ausbildende Betriebe überproportional gezogen, um Auswertungen für diese Gruppen zu ermöglichen (Gerhards et al. 2013, S. 10f).

Im Jahr 2011 wurden 2.004 Betriebe mit der CAPI-Methode erstmalig befragt. Die erste Welle ist bereits im FDZ des BIBB verfügbar. Die zweite Welle (2012) soll Ende 2013 zur Verfügung stehen.

5.4.3.2 Verfügbarkeit

Ein Datenzugang kann beim FDZ des BIBB kostenlos beantragt werden. Derzeit liegen die Daten ausschließlich für die On-Site-Nutzung vor. Für die Beantragung ist ein formloser Antrag mit Namen der Projektbearbeiter/innen, der beantragenden Institution sowie einer Projektbeschreibung notwendig. Der Datenzugang wird in der Regel zeitnah innerhalb von etwa drei Werktagen ermöglicht (Gerhards et al. 2013, S. 26f).

5.4.3.3 Merkmale zur Mindestlohnevaluation

Der Datensatz enthält ein umfangreiches Merkmalsset zum Thema Aus- und Weiterbildung. Dies beinhaltet sowohl Einstellungsfragen als auch detaillierte Angaben zu den derzeitigen Aktivitäten als Ausbildungsbetrieb bzw. hinsichtlich Weiterbildungsmaßnahmen. Ergänzt werden die Merkmale durch ausführliche Betriebsinformationen, der Arbeitsorganisation sowie der Personalstruktur und -bewegungen (Gerhards et al. 2013, S. 33-35).

⁴⁸ Grundlage ist die Betriebsdatei der Beschäftigtenstatistik der BA (siehe Unterabschnitt 5.1.1 zum BHP).

Für die Mindestlohnforschung von Interesse sind insbesondere die detaillierten Angaben zur Qualifikation und angebotenen Fort- und Weiterbildungen, die auch anhand der vorliegenden betrieblichen Strukturmerkmale verglichen werden können.

5.4.3.4 Merkmale zur Verknüpfung

Die Erhebung ist als Panel angelegt. Folglich ist eine Zuordnung der Betriebe zu ihren Angaben technisch möglich. Da die Ziehungsgrundlage die Betriebsdatei der Beschäftigtenstatistik der BA ist, ist eine Verknüpfung des BIBB-Qualifizierungspanels mit dem Datenangebot der BA und des IAB technisch möglich. Über die BA-Betriebsnummer ist ferner eine Verknüpfung mit dem URS und somit mit dem Datenbestand der FDZ der Statistischen Ämter technisch möglich.

Eine Verknüpfung auf Branchen- oder Berufsklassifikationsebene ist mit Einschränkungen möglich. Die Wirtschaftszweige liegen ausschließlich auf der Ebene der Zweisteller vor. Die Berufe sind sowohl nach der Berufsklassifikation nach ISCO-88,⁴⁹ als auch nach Blossfeld (1985) und der nationalen Klassifikation der Berufe 1992 (Statistisches Bundesamt 1992) erfasst.

5.4.3.5 Identifikation des Evaluationspotenzials

Das BIBB-Qualifizierungspanel bietet eine umfangreiche Datenbasis für die Analyse der betrieblichen Aktivitäten in der Ausbildung sowie der Weiterbildung ihrer Beschäftigten. Für gesamtwirtschaftliche Analysen ist der Datensatz mit 2.004 Beobachtungen auch ausreichend groß. Es bleibt abzuwarten, wie hoch die tatsächliche Panelausschöpfung ist.⁵⁰

Trotz dieser Voraussetzungen ist der Datensatz für die Evaluation branchenspezifischer Mindestlöhne nur sehr eingeschränkt verwendbar, da die erste Befragung nach der Einführung der Mindestlöhne in den meisten Branchen liegt. Zudem ist die Stichprobe zu klein und die Repräsentativität für Einzelbranchen wahrscheinlich nicht gegeben.

Allerdings bietet die Studie gute Anknüpfungsmöglichkeiten. Durch eine Zusatzziehung aus der Betriebsdatei der Beschäftigtenstatistik der BA kann die Stichprobe sinnvoll erweitert werden. Dies muss allerdings in Abstimmung mit dem BIBB geschehen, was eine Vorlaufzeit von mindestens neun Monaten voraussetzt.

⁴⁹ Internationale Standardklassifikation der Berufe (International Standard Classification of Occupations) in der Fassung von 1988.

⁵⁰ 79,6 % der befragten Betriebe gaben eine Wiederbefragungsbereitschaft an (Gerhards et al. 2013).

5.4.4 Befragung des Instituts für Angewandte Wirtschaftsforschung e.V. (IAW) zum Mindestlohn im Elektrohandwerk

5.4.4.1 Allgemeine Datensatzinformationen

Im Rahmen der 2011 durch das BMAS in Auftrag gegebenen Evaluation der gesetzlichen Mindestlohnregelungen im Elektrohandwerk hat das IAW eine standardisierte schriftliche Befragung bei Betrieben durchgeführt, die potenziell unter die Mindestlohnregelung im Elektrohandwerk fallen. Dies waren Betriebe, die dem Wirtschaftszweig „Elektroinstallation“ (WZ03: 45.31.0) angehören, bzw. den Wirtschaftszweigen „Rohrleitungs- und Kabelleitungstiefbau“ (WZ03: 45.21.7) oder „Sonstige Bauinstallationen“ (WZ03: 45.34.0) angehören und gleichzeitig mehr als 30 % ihrer Belegschaft den Berufsgruppen Elektroinstallateur, Fernmeldemonteur, -handwerker, Elektrogeräte-, Elektroteilemontierer, Elektroingenieur oder Techniker des Elektrofaches zuzurechnen sind.

Auf Basis dieser Abgrenzung wurden aus dem BHP 4.000 Betriebe gezogen. Dies entspricht einem Auswahlsatz von 13,1 % der Betriebe in der beschriebenen Grundgesamtheit (IAW 2011a). Die Stichprobe wurde nach Ost- und Westdeutschland sowie nach Betriebsgrößenklasse geschichtet. Als Ziehungsstichtag wurde der 30. Juni 2008 gewählt.

Die Befragung fand im Frühjahr 2011 statt und erzielte einen Rücklauf von 25,3 % (1.010 Betriebe). Es handelte sich um eine einmalige Querschnittsbefragung. Eine detaillierte Beschreibung der Stichprobenziehung sowie des Feldphasenverlaufs kann im Endbericht zur Mindestlohnevaluation des Elektrohandwerks (IAW 2011a, S. 16-19, 44f) eingesehen werden.

5.4.4.2 Verfügbarkeit

Die Befragungsdaten wurden für die Mindestlohnevaluation erhoben und ausgewertet (IAW 2011a). Da keine konkrete weitere Nutzung der Daten geplant war, ist eine erneute Nutzung der Daten mit dem BMAS als Auftraggeber der Evaluation sowie dem IAW als Bereitsteller der Daten zu klären.

5.4.4.3 Merkmale zur Mindestlohnevaluation

Da die Befragung explizit zur Mindestlohnevaluation durchgeführt wurde, enthält der Datensatz umfangreiche Merkmale, die für die Mindestlohnevaluation von Interesse sind. Dies beinhaltet eine konkrete Zuordnung der Betriebe in den Geltungsbereich des Mindestlohns als auch Informationen zur Tarifbindung der Betriebe. Weiterhin wurden die Wettbewerbssituation sowohl auf dem Arbeitsmarkt als auch auf dem Dienstleistungsmarkt für das Elektrohandwerk abgefragt, eine Bewertung des Mindestlohns erbeten sowie betriebliche Charakteristika erfasst. Ein detaillierter Überblick zu den erhobenen Merkmalen kann unter IAW (2011a, Unterabschnitt 3.1.2, S. 39ff) nachgelesen werden.

5.4.4.4 Merkmale zur Verknüpfung

Da das BHP als Ziehungsgrundlage dient, ist eine Verknüpfung der Befragungsdaten mit dem BHP und somit auch den IEB über eine durch IAB-ITM für dieses Projekt neu vergebene Betriebsnummer technisch möglich. 54 % der befragten Betriebe haben auch einer Verknüpfung ihrer Angaben zugestimmt (IAW 2011a, S. 46). Da die BA-Betriebsnummer auch im Unternehmensregister vermerkt ist, ist eine Verknüpfung mit diesem ebenfalls technisch möglich.

5.4.4.5 Identifikation des Evaluationspotenzials

Das zusätzliche Evaluationspotenzial des Befragungsdatensatzes ist relativ eingeschränkt, da die wesentlichen Auswertungen bereits in der Evaluationsstudie in IAW (2011a) vorgenommen wurden. Allerdings ermöglicht die Verknüpfung sowohl mit dem BHP als auch mit den IEB eine Identifikation der Mindestlohnbranche in beiden Datensätzen. Da die Befragung nicht als Panel konzipiert wurde und daher auch die Panelbereitschaft nicht ermittelt wurde, ist eine Ausweitung der Befragung zum Panel nicht möglich.

5.4.5 Befragung des Instituts für Angewandte Wirtschaftsforschung e.V. (IAW) zum Mindestlohn im Maler- und Lackiererhandwerk

5.4.5.1 Allgemeine Datensatzinformationen

Im Rahmen der 2011 durch das BMAS in Auftrag gegebenen Evaluation der gesetzlichen Mindestlohnregelungen im Maler- und Lackiererhandwerk hat das IAW eine standardisierte schriftliche Befragung bei Betrieben durchgeführt, die potenziell unter die Mindestlohnregelung im Maler- und Lackiererhandwerk fallen. Dies waren Betriebe, die den Wirtschaftszweigen „Stuckateurgewerbe“ (WZ03: 45.41.0), „Sonstige Fußbodenlegerei und -kleberei“ (WZ03: 45.43.4), „Tapetenkleberei“ (WZ03: 45.43.5) oder „Maler- und Lackierergewerbe“ (WZ03: 45.44.1) angehören, bzw. den Wirtschaftszweigen „Oberflächenveredelung und Wärmebehandlung“ (WZ03: 45.32.0) oder „Dämmung gegen Kälte, Wärme, Schall und Erschütterung“ (WZ03: 45.32.0) angehören und gleichzeitig mehr als 30 % ihrer Belegschaft den Berufsgruppen Stuckateure, Gipser, Verputzer, Maler und Lackierer im Ausbau oder Dekorationen- und Schildermaler zuzurechnen sind.

Auf Basis dieser Abgrenzung wurden aus dem BHP 4.000 Betriebe gezogen. Dies entspricht einem Auswahlsatz von 11,9 % der Betriebe in der beschriebenen Grundgesamtheit (IAW 2011b). Die Stichprobe wurde nach Ost- und Westdeutschland sowie nach Betriebsgrößenklasse geschichtet. Als Ziehungsstichtag wurde der 30. Juni 2008 gewählt.

Die Befragung fand im Frühjahr 2011 statt und erzielte einen Rücklauf von 22,3 % (890 Betriebe). Es handelte sich um eine einmalige Querschnittsbefragung. Eine detaillierte Beschreibung der Stichprobenziehung sowie des Feldphasenverlaufs kann im Endbericht zur Mindestlohnevaluation des Maler- und Lackiererhandwerks (IAW 2011b, S. 16-18, 43-45) eingesehen werden.

5.4.5.2 Verfügbarkeit

Die Befragungsdaten wurden für die Mindestlohnevaluation erhoben und ausgewertet (IAW 2011b). Da keine konkrete weitere Nutzung der Daten geplant war, ist eine erneute Nutzung der Daten mit dem BMAS als Auftraggeber der Evaluation sowie dem IAW als Bereitsteller der Daten zu klären.

5.4.5.3 Merkmale zur Mindestlohnevaluation

Da die Befragung explizit zur Mindestlohnevaluation durchgeführt wurde, enthält der Datensatz umfangreiche Merkmale, die für die Mindestlohnevaluation von Interesse sind. Dies beinhaltet eine konkrete Zuordnung der Betriebe in den Geltungsbereich des Mindestlohns als auch Informationen zur Tarifbindung der Betriebe. Weiterhin wurden die Wettbewerbssituation sowohl auf dem Arbeitsmarkt als auch auf dem Dienstleistungsmarkt für das Maler- und Lackiererhandwerk abgefragt, eine Bewertung des Mindestlohns erbeten sowie betriebliche Charakteristika erfasst. Ein detaillierter Überblick zu den erhobenen Merkmalen kann unter IAW (2011b, Unterabschnitt 3.1.2, S. 39ff) nachgelesen werden.

5.4.5.4 Merkmale zur Verknüpfung

Da das BHP als Ziehungsgrundlage dient, ist eine Verknüpfung der Befragungsdaten mit dem BHP und somit auch den IEB über eine durch IAB-ITM für dieses Projekt neu vergebene Betriebsnummer technisch möglich. 56 % der befragten Betriebe haben auch einer Verknüpfung ihrer Angaben zugestimmt (IAW 2011b, S. 45). Da die BA-Betriebsnummer auch im Unternehmensregister vermerkt ist, ist eine Verknüpfung mit diesem ebenfalls technisch möglich.

5.4.5.5 Identifikation des Evaluationspotenzials

Das zusätzliche Evaluationspotenzial des Befragungsdatensatzes ist relativ eingeschränkt, da die wesentlichen Auswertungen bereits in der Evaluationsstudie in IAW (2011b) vorgenommen wurden. Allerdings ermöglicht die Verknüpfung sowohl mit dem BHP als auch mit den IEB eine Identifikation der Mindestlohnbranche in beiden Datensätzen. Da die Befragung nicht als Panel konzipiert wurde und daher auch die Panelbereitschaft nicht ermittelt wurde, ist eine Ausweitung der Befragung zum Panel nicht möglich.

5.4.6 Befragung zum Mindestlohn in der Pflegebranche durch T.I.P. Biehl & Partner

5.4.6.1 Allgemeine Datensatzinformationen

Im Rahmen der 2011 durch das BMAS in Auftrag gegebenen Evaluation der gesetzlichen Mindestlohnregelungen in der Pflegebranche wurde durch das Befragungsinstitut T.I.P. Biehl & Partner eine standardisierte assistierte Onlinebefragung bei Einrichtungen durchgeführt, die potenziell unter die Mindestlohnregelung in der Pflegebranche fallen. Dies sind sämtliche Einrichtungen, die im Rahmen der Pflegestatistik durch die Statistischen Ämter des Bundes

und der Länder regelmäßig befragt werden. Die Grundgesamtheit der Einrichtungen in der Pflegestatistik wurde um eine Adressrecherche in den Branchenbüchern der Deutschen Telekom durch das Befragungsinstitut erweitert. Insgesamt bestand die Grundgesamtheit aus 33.360 Einrichtungen (IAW 2011c, S. 43).

Aus der Grundgesamtheit wurde eine geschichtete Zufallsstichprobe anhand der Merkmale Pflegeart (ambulant oder stationär), Trägerschaft (privat, freigemeinnützig, öffentlich-rechtlich) sowie Region (ost- und westdeutsche Bundesländer) von insgesamt 17.400 Einrichtungen gezogen. Dies entspricht einem Auswahlsatz von 52,2 % (IAW 2011c, S. 43).

Die Befragung wurde als Panel mit zwei Wellen konzipiert. Die erste Welle wurde im September und Oktober 2010 durchgeführt, die zweite Welle fand im März und April 2011 statt. Insgesamt haben an mindestens einer der beiden Wellen 1.513 Einrichtungen teilgenommen, was einem Rücklauf von 8,7 % entspricht. Eine detaillierte Beschreibung der Stichprobenziehung sowie des Feldphasenverlaufs kann im Endbericht zur Mindestlohnevaluation Pflegebranche (IAW 2011c, Abschnitt 4.4, S. 42ff) eingesehen werden.

5.4.6.2 Verfügbarkeit

Die Befragungsdaten wurden für die Mindestlohnevaluation durch T.I.P. Biehl & Partner erhoben und durch das IAW ausgewertet (IAW 2011c). Da keine konkrete weitere Nutzung der Daten geplant war, ist eine erneute Nutzung der Daten mit dem BMAS als Auftraggeber der Evaluation sowie T.I.P. Biehl & Partner als Bereitsteller der Daten zu klären.

5.4.6.3 Merkmale zur Mindestlohnevaluation

Da die Befragung explizit zur Mindestlohnevaluation durchgeführt wurde, enthält der Datensatz umfangreiche Merkmale, die für die Mindestlohnevaluation von Interesse sind. Dies gilt vor allem für eine detaillierte und nach Geschlecht, Pflegefach- und -hilfskräften und nach Ost- und Westdeutschland differenzierte Abfrage der Lohnstruktur sowie eine Zuordnung der Einrichtung in den Geltungsbereich des Mindestlohns. Weiterhin wurden Fragen zu den Eigenschaften der Einrichtung, der Beschäftigtenstruktur, der Kostenstruktur, dem Ertrag und offenen Stellen sowie zur Wirkung des Mindestlohns (IAW 2011c, S. 44) gestellt. Eine detaillierte Beschreibung des Datensatzes befindet sich in IAW (2011c, S. 44f).

5.4.6.4 Merkmale zur Verknüpfung

Das Adressmaterial stammt größtenteils aus dem Adressbestand der Statistischen Ämter des Bundes und der Länder zur Pflegestatistik. Sofern dieses Adressmaterial mit einem in der Befragung vorhandenen Identifikator beim Befragungsinstitut vorliegt, ist eine Verknüpfung der Daten sowohl mit der Pflegestatistik als auch mit anderen Datenbeständen wie dem URS oder dem BHP über Record Linkage technisch möglich.

5.4.6.5 Identifikation des Evaluationspotenzials

Das zusätzliche Evaluationspotenzial des Befragungsdatensatzes ist relativ begrenzt, da die wesentlichen Auswertungen bereits in der Evaluationsstudie in IAW (2011c) vorgenommen wurden. Zusätzlich ist bei dem relativ niedrigen Rücklauf sowie nur eingeschränkt durchführbaren Selektivitätsanalysen (IAW 2011c, S. 48-51) die Repräsentativität des Datensatzes schwierig zu bewerten.

5.4.7 Befragung des Instituts Arbeit und Qualifikation (IAQ) zum Mindestlohn in der Branche Wäschereidienstleistungen

5.4.7.1 Allgemeine Datensatzinformationen

Im Rahmen der 2011 durch das BMAS in Auftrag gegebenen Evaluation der gesetzlichen Mindestlohnregelungen in der Branche Wäschereidienstleistungen im Objektkundengeschäft hat das Sozialwissenschaftliche Umfragezentrum (SUZ GmbH) eine standardisierte telefonische Befragung (CATI) bei Betrieben durchgeführt, die potenziell unter die Mindestlohnregelung in der Branche Wäschereidienstleistungen im Objektkundengeschäft fallen. Dies waren Betriebe, die dem Wirtschaftszweig „Wäscherei und chemische Reinigung“ (WZ08: 96.01.0) angehörten (IAQ 2011b, S. 36).

Auf Basis dieser Abgrenzung wurde aus der Betriebsstatistik der BA eine Stichprobe gezogen. Diese wurde durch die Adressdatenbestände der Arbeitgeberverbände ergänzt. Insgesamt wurde eine nach Betriebsgrößenklassen und nach Ost- und Westdeutschland geschichtete Stichprobe von 1.504 Betrieben gezogen (IAQ 2011b). Als Ziehungsstichtag wurde der 30. September 2010 gewählt.

Die Befragung fand im Juni und Juli 2011 statt und erzielte einen Rücklauf von 21,7 % (327 Betriebe). Es handelte sich um eine einmalige Querschnittsbefragung. Eine detaillierte Beschreibung der Stichprobenziehung sowie des Feldphasenverlaufs kann im Endbericht zur Mindestlohnevaluation für die Branche der Wäschereidienstleistungen (IAQ 2011b, S. 35-45) eingesehen werden.

5.4.7.2 Verfügbarkeit

Die Befragungsdaten wurden für die Mindestlohnevaluation erhoben und ausgewertet (IAQ 2011b). Da keine konkrete weitere Nutzung der Daten geplant war, ist eine erneute Nutzung der Daten mit dem BMAS als Auftraggeber der Evaluation sowie dem IAQ als Bereitsteller der Daten zu klären.

5.4.7.3 Merkmale zur Mindestlohnevaluation

Da die Befragung explizit zur Mindestlohnevaluation durchgeführt wurde, enthält der Datensatz umfangreiche Merkmale, die für die Mindestlohnevaluation von Interesse sind. Dies beinhaltet eine konkrete Zuordnung der Betriebe in den Geltungsbereich des Mindestlohns, den

Umsatzanteil im Objektkundengeschäft sowie die Lohnstruktur vor und nach Mindestlohneinführung. Ebenfalls vor und nach Mindestlohneinführung wurden die Beschäftigtenstruktur, die Wettbewerbssituation und weitere Beschäftigungsmerkmale erfasst. Weiterhin wurden eine Bewertung des Mindestlohns erbeten sowie betriebliche Charakteristika erfasst. Ein detaillierter Überblick zu den erhobenen Merkmalen kann unter IAQ (2011b, S. 38-40) nachgelesen werden.

5.4.7.4 Merkmale zur Verknüpfung

Sofern das Adressmaterial der Befragung noch vorhanden und mit der Befragung verknüpfbar ist, ist eine Verknüpfung der Daten durch Record Linkage technisch möglich. Über die Betriebsstatistik der BA, die als Ziehungsgrundlage diente, müsste eine direkte Verknüpfung anhand der Betriebsnummer mit dem BHP möglich sein.

5.4.7.5 Identifikation des Evaluationspotenzials

Das zusätzliche Evaluationspotenzial des Befragungsdatensatzes ist stark eingeschränkt, da die wesentlichen deskriptiven Auswertungen bereits in der Evaluationsstudie in IAQ (2011b) vorgenommen wurden. Allerdings ermöglicht eine Verknüpfung mit dem BHP eine Identifikation der Mindestlohnbranche in den BHP.

5.4.8 Dafne

5.4.8.1 Allgemeine Datensatzinformationen

Das Bureau van Dijk ist ein weltweiter Anbieter von Firmen- und Finanzdaten. Ein Produkt des BvD ist die Datenbank Dafne, die Geschäftsdaten von über einer Million Unternehmen in Deutschland enthält⁵¹. Grundgesamtheit bilden dabei die bilanzierungspflichtigen Unternehmen. In Dafne sind Finanz- und Bilanzdaten sowohl für solvente als auch insolvente Unternehmen verfügbar.

Die Daten sind frühestens seit 1999 verfügbar, aber insbesondere bis 2007 lückenhaft. Aufgrund einer Gesetzesänderung zur Bilanzierungspflicht sind die Variablen ab 2007 für publizierende Unternehmen vollständiger gefüllt. BvD sammelt für seine Datenbanken alle öffentlich zugänglichen Datenquellen⁵², bereitet diese auf und bietet sie umfassend an.

5.4.8.2 Verfügbarkeit

Dafne kann über BvD kostenpflichtig bezogen werden. Dabei hängen die Kosten von der Zugangsart (Online-Softwaremaske vs. Datenextrakt), sowie der Menge der gewünschten

⁵¹ Weitere Informationen unter <http://www.bvdinfo.com/de-de/products/company-information/national/dafne>.

⁵² Hauptlieferanten des BvD sind die Creditreform Rating AG und der Verband der Vereine Creditreform e.V.

Variablen und der Anzahl der Nutzer/innen ab. Um die individuelle Nutzbarkeit der Daten zu evaluieren, kann eine Testversion gratis beantragt werden⁵³.

Der Datenextrakt des BvD, den das FDZ der BA im IAB gekauft hat, unterliegt den Nutzungsbestimmungen des BvD. Diese sehen vor, dass ausschließlich Mitarbeiter/innen des FDZ der BA im IAB die Daten nutzen. Kooperationsprojekte sind möglich, sofern die Kooperationspartner außerhalb des FDZ der BA im IAB nicht die Daten selbst, sondern lediglich die Ergebnisse einsehen. Für die Zukunft ist eine Ausweitung der Verfügbarkeit geplant.

5.4.8.3 Merkmale zur Mindestlohnevaluation

Die Dafne-Datenbank enthält Kennzahlen, die bei der Evaluation von Mindestlohnwirkungen auf den Unternehmenserfolg relevant sein könnten. Hierzu zählen insbesondere Informationen über den Gewinn und die Profitabilität (z.B. return on assets, return on equity, etc.), den Umsatz, aber auch über Kapitalbestände, Mitarbeiterzahlen und Exportquoten.

Dafne enthält Angaben zum Wirtschaftszweig (WZ 2008 5-Steller Code, NACE Rev. 2 Code, NAICS 2007 Code, US SIC Code). Des Weiteren sind die genauen Adressen der Firmen bekannt, sowie das Bundesland und der Stadtkreis, wodurch die Möglichkeit besteht, Analysen auf Regionen- oder Kreisebene durchzuführen.

5.4.8.4 Merkmale zur Verknüpfung

Jede Firma in Dafne ist eindeutig durch eine Identifikationsnummer (BvD-ID) identifiziert. Des Weiteren enthält Dafne detaillierte Adressangaben (Straße, Hausnummer, Ort, Postleitzahl) und Rechtsformangaben für jede im Datensatz enthaltene Firma. Durch Record Linkage können somit Verknüpfungen mit anderen Datensätzen erstellt werden. Beispielsweise wird derzeit in einem Projekt geklärt, inwieweit die Betriebe in den administrativen Daten des IAB mit den Firmen in Dafne verknüpft werden können. Sobald eine BvD-ID einer Betriebsnummer zugeordnet ist, ist die Verknüpfung theoretisch mit jedem Datensatz des IAB möglich, der eine Betriebsnummer enthält.

Auf aggregierter Ebene ermöglichen der Wirtschaftszweig, sowie regionale Angaben, eine Verknüpfung mit anderen Daten.

5.4.8.5 Identifikation des Evaluationspotenzials

Dafne fokussiert auf Bilanz- und Firmendaten. Daher fehlen genauere Angaben zur Beschäftigtenstruktur und zu Löhnen. Jedoch sind regionale Einheiten und Wirtschaftszweige genau abgrenzbar, wodurch eine Mindestlohnevaluierung auf Branchen- oder Regionalebene möglich wäre. So wäre Dafne geeignet eine Mindestlohnbranche mit einer Kontrollgruppe aus einer anderen Branche zu vergleichen. In den meisten Fällen wird jedoch eine Verknüpfung

⁵³ Datenzugang über <http://www.bvdinfo.com/de-de/home>.

mit anderen Daten notwendig sein, um Informationen über Beschäftigtenstrukturen und Löhnen zu erhalten. Daraufhin sind jedoch Analysen interessanter Aspekte wie beispielsweise die Outcomes Gewinn, Umsatz und Profitabilität möglich. Hierbei ist zu berücksichtigen, dass es sich bei Dafne nicht um Betriebs- sondern Unternehmensinformationen handelt.

5.4.9 Mannheimer Unternehmenspanel (MUP)

5.4.9.1 Allgemeine Datensatzinformationen

Das MUP ist ein Paneldatensatz zu allen rechtlich selbstständigen Unternehmen mit Sitz in Deutschland, der vom Zentrum für Europäische Wirtschaftsforschung (ZEW) in Zusammenarbeit mit Creditreform seit 1999 als Vollerhebung existiert. Es handelt sich dabei um eine Zusammenstellung von Strukturdaten, die von Creditreform dem ZEW für wissenschaftliche Zwecke zur Verfügung gestellt werden und vom ZEW hierfür aufbereitet werden.⁵⁴

Die Informationen werden von Creditreform durch eine systematische Recherche in öffentlichen Registern, insbesondere im Handelsregister, in Tageszeitungen, in Geschäftsberichten und in veröffentlichten Bilanzen gewonnen. Diese Informationen werden durch Auskunftsanfragen von Creditreform ergänzt (Almus et al. 2000, S. 9).

Zum Zeitpunkt der ersten Evaluationsrunde der Mindestlöhne enthielt das MUP Einträge für insgesamt fast 6,8 Mio. Unternehmen, wovon etwa 3,1 Mio. aktiv waren (ZEW, infas 2011a & 2011b). Ab der Beobachtungswelle Juli 2012 liegen Einträge für etwa 10 Mio. Unternehmen⁵⁵ mit Sitz in Deutschland vor.⁵⁶

5.4.9.2 Verfügbarkeit

Auf Anfrage können Standardauswertungen des MUP zur Verfügung gestellt werden. Die Kosten liegen abhängig von der gewünschten Auswertung und der Ebene der Auswertung zwischen 100 Euro und 750 Euro pro Auswertung. Die Daten werden mit dazugehörigen Variablenbeschreibungen im Excel-Format elektronisch zur Verfügung gestellt. Eine detaillierte Beschreibung der Auswertungen sowie der Kosten kann unter ZEW (2012) nachgelesen werden. Da die Standardtabellen bereits aufbereitet vorliegen, ist mit einer sehr geringen Bearbeitungszeit zu rechnen.

Aufgrund rechtlicher Restriktionen sind tieferegehende Analysen ausschließlich in Kooperation mit dem ZEW möglich, da die Daten ausschließlich durch Mitarbeiterinnen und Mitarbeiter des ZEW genutzt werden dürfen. Sowohl die Kosten als auch die Dauer der Bearbeitung sind dabei sehr stark vom Aufwand für die Auswertungen abhängig.

⁵⁴ Für das genaue Vorgehen siehe Almus et al. (2000).

⁵⁵ Von diesen sind nicht alle aktiv.

⁵⁶ Information von der Projektseite des ZEW (<http://www.zew.de/de/forschung/projekte.php?action=detail&nr=102&abt=ioek> [abgerufen: 25.10.2013]).

5.4.9.3 Merkmale zur Mindestlohnevaluation

Für die Evaluationen der Mindestlöhne in der Abfallwirtschaft (ZEW 2011a) und im Dachdeckerhandwerk (ZEW 2011b) wurde das MUP insbesondere für eine Schätzung der Auswirkungen des Mindestlohns auf Gründungen und Schließungen verwendet. Für die Auswertungen lagen diverse Strukturmerkmale der Unternehmen wie Umsatz, Mitarbeiterzahl, oder Bonität vor (ZEW 2011a, S. 74). Die Daten sind teilweise unterjährig verfügbar, teilweise werden die Merkmale allerdings zu einem Stichtag im Jahr erhoben.

Auf Basis des MUP werden regelmäßig Befragungen bei neugegründeten Unternehmen sowie bei Bestandsunternehmen durchgeführt, die weitere Informationen insbesondere zu Inputfaktoren beinhalten. Dies sind sowohl einmalige als auch wiederholte und längerfristige Befragungen. Welche Befragungen zu welchen Zeitpunkten vorliegen, kann im Datenangebot des FDZ des ZEW eingesehen werden.⁵⁷

5.4.9.4 Merkmale zur Verknüpfung

Über die Nummer im Handelsregister ist eine Verknüpfung des MUP mit dem URS und damit mit einer Reihe weiterer Datenquellen (siehe dazu den Unterabschnitt 5.1.3 zum URS) technisch möglich. Zudem existieren im MUP die Klarnamen der Unternehmen und deren Adresse sowie die fünfstelligen Wirtschaftszweige und Gemeindeschlüssel. Dies ermöglicht nicht nur eine Zuspiegelung von regionalen Indikatoren.

5.4.9.5 Identifikation des Evaluationspotenzials

Wie das ZEW (2011a & 2011b) bereits gezeigt hat, kann das MUP für Analysen hinsichtlich betrieblicher Ergebnisindikatoren wie Umsatz, Marktein- und -austritte sowie zur Unternehmensgröße anhand der Mitarbeiterzahl genutzt werden. Allerdings ist zu beachten, dass es keine Angaben zur Entlohnung gibt und es sich um eine Datenbank auf Unternehmensebene handelt, während die Mindestlöhne nach dem AEntG auf Betriebsebene gelten. In den beiden vom ZEW (2011a & 2011b) evaluierten Branchen wird das Problem als eher gering eingestuft, da im Dachdeckerhandwerk Unternehmen und Betrieb in der Regel deckungsgleich sind und in der Abfallwirtschaft die regionalen Niederlassungen häufig als Tochtergesellschaften organisiert sind und folglich im MUP vertreten sind. In anderen Branchen muss untersucht werden, inwieweit eine Untersuchung auf Unternehmensebene zu starken Verzerrungen führt.

⁵⁷ Das Datenangebot des FDZ des ZEW befindet sich unter <http://kooperationen.zew.de/de/zew-fdz/datenangebot.html> [abgerufen am: 05.11.2013]. Da sich das FDZ des ZEW noch in Aufbauphase befindet, kann das Datenangebot und die Dokumentationen für wissenschaftliche Zwecke sowohl als SUF als auch über einen GWA kostenlos genutzt werden. Detailliertere Informationen und Kontaktdaten können unter <http://kooperationen.zew.de/de/zew-fdz/nutzungsbedingungen.html> [abgerufen am: 05.11.2013] eingesehen werden.

6 Daten auf aggregierter Ebene

6.1 Datenangebot der Statistik der Bundesagentur für Arbeit

6.1.1.1 Allgemeine Datensatzinformationen

Die Bundesagentur für Arbeit erstellt und veröffentlicht als Teil der amtlichen Statistik Statistiken zum Arbeitsmarkt. Diese basieren weitestgehend auf der Auswertung von Prozessdaten, die bei der Durchführung der Aufgaben der Bundesagentur nach den Sozialgesetzbüchern anfallen. Eine weitere Grundlage bilden Arbeitgebermeldungen zu sozialversicherungspflichtig und geringfügig Beschäftigten (DEÜV-Meldungen) sowie Meldungen von Verleihbetrieben (Zeitarbeitsfirmen) zur Überlassung von Arbeitnehmern (AÜG).

Die Hauptarbeitsgebiete der BA-Statistik umfassen die Statistiken der Arbeitslosen, Arbeitssuchenden und der gemeldeten Arbeitsstellen, die Statistiken zum Ausbildungsstellenmarkt, die Förderstatistik, die Leistungsstatistiken nach dem SGB III, die Leistungsstatistik der Grundsicherung für Arbeitsuchende (SGB II), die Beschäftigungsstatistik, die Arbeitnehmerüberlassungsstatistik sowie die Eingliederungsbilanzen der Agenturen für Arbeit.

Die Statistiken basieren auf Vollerhebungen, wodurch eine sehr tiefe Ausdifferenzierung möglich wird. Sie liegen gegliedert nach regionalen Einheiten, soziodemografischen sowie erwerbsbiografischen Merkmalen vor.

6.1.1.2 Verfügbarkeit

Das Standardangebot der Statistik der BA ist über deren Internetauftritt online verfügbar und kostenlos. Die Beantragung tiefer gegliederter Datenauszüge erfolgt entsprechend Abschnitt 3.1. Dauer und Kosten der Beantragung sind abhängig vom genauen Datenbedarf und der Komplexität der Datenziehung.

Die Aktualisierung der einzelnen Statistiken erfolgt in der Regel monatlich. Zeitreihen reichen zum Teil bis 1950 zurück (Arbeitslosigkeit), zumeist aber deutlich weniger weit. Ein Grund hierfür ist, dass gesetzliche Änderungen immer wieder zu Brüchen führen, so dass die Daten nicht konsistent über lange Zeiträume abgebildet werden können. Umfangreiche Dokumentationen sind auf der Internetpräsenz der Statistik der BA verfügbar.

6.1.1.3 Merkmale zur Mindestlohnevaluation

Für die Analyse von Mindestlohnwirkungen sind insbesondere die Statistiken zur Entwicklung der sozialversicherungspflichtigen Beschäftigung, der sozialversicherungspflichtigen Bruttoarbeitsentgelte, der Betriebe, der Arbeitslosen und der gemeldeten Stellen, eventuell aber auch der Ausbildungsstellenmarkt relevant.

6.1.1.4 Merkmale zur Verknüpfung

Bei den BA-Statistiken handelt es sich um aggregierte Datenbestände. Verknüpfungen auf individueller Ebene sind nicht möglich. Für personen- oder betriebsbezogene Daten sind die IAB-Produkte zu wählen.

Allerdings lassen die Regional- und Wirtschaftszweigschlüssel, wenn vorhanden, eine Verknüpfung mit anderen Datensätzen auf dieser aggregierten Ebene zu.

6.1.1.5 Identifikation des Evaluationspotenzials

Das Statistikangebot der BA ist vor allem dann interessant, wenn Analysen durch zusätzliche Aggregatdaten auf Wirtschaftszweig-, Bundesland- oder Kreisebene ergänzt werden sollen, etwa in Makromodellen, Panelansätzen auf Kreisebene oder Kontrollgruppenansätzen mit verschiedenen Wirtschaftszweigen.

Für sich allein genommen haben die Daten ein eingeschränktes Analysepotenzial, da die Betroffenheit vom Mindestlohn nicht abgebildet wird. Eine Ausnahme könnten hier Kontrollgruppenansätze auf Basis der Wirtschaftszweige sein, wenn die Mindestlohnbranche und die Kontrollbranche(n) über die Klassifikation der Wirtschaftszweige korrekt abgegrenzt werden können.

6.2 Datenangebot der Statistischen Ämter des Bundes und der Länder

6.2.1.1 Allgemeine Datensatzinformationen

Neben den über die FDZ der Statistischen Ämter erhältlichen Mikrodatensätzen sind über das Onlineangebot der Statistischen Ämter eine Vielzahl von Ergebnissen aus der amtlichen Statistik in den entsprechenden Fachserien und Datenbanken zu beziehen. Für ein breites Themenspektrum auf nationaler bzw. regional stark aggregierter Ebene ist hier zunächst das Datenportal GENESIS-Online zu nennen⁵⁸ und für regional tiefer gegliederte Tabellen die Regionaldatenbank Deutschland.⁵⁹ Da es sich in der Regel um amtliche Befragungen handelt, beruhen die Statistiken entweder auf Vollerhebungen oder Stichproben mit Auskunftspflicht.

6.2.1.2 Verfügbarkeit

Es handelt sich um Online-Datenbanken, die kostenlos genutzt werden können. Zusätzliche Nutzungsmöglichkeiten stehen registrierten Kunden gegen eine Jahresgebühr von 50 Euro zur Verfügung. Periodizität und abgedeckte Zeiträume sind abhängig von der jeweiligen Fachserie. Dokumentationen sind in den entsprechenden Datenportalen verfügbar.

⁵⁸ <https://www-genesis.destatis.de/>

⁵⁹ <https://www.regionalstatistik.de/>

6.2.1.3 Merkmale zur Mindestlohnevaluation

Die Nutzbarkeit für die Mindestlohnevaluation hängt stark von der genauen Fragestellung und dem empirischen Vorgehen ab. Dennoch sind ein Großteil der aggregierten Daten der GENESIS-Datenbank für branchenspezifische Analysen entweder aufgrund der Thematik uninteressant oder aber nicht spezifisch genug. Ausnahmen könnten beispielsweise die Konjunktur- und Strukturhebungen in den jeweiligen Wirtschaftsbereichen, die Preisindizes für die Bauwirtschaft, die Vierteljährliche Verdiensterhebung oder die Insolvenzstatistik sein.

Die Regionaldatenbank ist vor allem für Paneldatenmodelle auf regionaler Ebene (z.B. den Kreisen) relevant, da sie wichtige Kovariate im Bereich Bevölkerung, Bautätigkeit, Konjunktur- und Strukturhebung, Gewerbean- und Abmeldungen, Insolvenzen oder Baulandpreise enthält.

6.2.1.4 Merkmale zur Verknüpfung

Bei den angesprochenen Statistiken handelt es sich um aggregierte Datenbestände. Verknüpfungen auf individueller Ebene sind nicht möglich. Allerdings lassen die Regional- und Wirtschaftszweigschlüssel, wenn vorhanden, eine Verknüpfung mit anderen Datensätzen auf dieser aggregierten Ebene zu.

6.2.1.5 Identifikation des Evaluationspotenzials

Das Statistikangebot der Statistischen Ämter ist vor allem dann interessant, wenn Analysen durch zusätzliche Aggregatdaten auf Wirtschaftszweig-, Bundesland- oder Kreisebene ergänzt werden sollen, etwa in Makromodellen, Panelansätzen auf Kreisebene oder Kontrollgruppenansätzen mit verschiedenen Wirtschaftszweigen.

Für sich allein genommen haben die Daten ein eingeschränktes Analysepotenzial, da die Betroffenheit vom Mindestlohn nicht abgebildet wird. Eine Ausnahme könnten hier Kontrollgruppenansätze auf Basis der Wirtschaftszweige sein, wenn die Mindestlohnbranche und die Kontrollbranche(n) über die Klassifikation der Wirtschaftszweige korrekt abgegrenzt werden können.

6.3 Elektronisches Verbandsinformations-, Recherche- und Analysesystem (ELVIRA) des Hauptverbands der Deutschen Bauindustrie e.V.

6.3.1.1 Allgemeine Datensatzinformationen

Das Elektronische Verbandsinformations-, Recherche- und Analysesystem (ELVIRA) des Hauptverbandes der Deutschen Bauindustrie e.V. ist ein Datenportal mit den für die Bauwirtschaft relevanten Statistiken und Daten unterschiedlicher Anbieter. Diese können durch die Nutzer/innen über eine Eingabemaske ausgewertet und grafisch dargestellt werden.

Ein wesentlicher Bestandteil der Datenbank wird aus administrativen Datenquellen der BA, der Statistischen Ämter oder der SOKA-Bau gespeist. Allerdings beinhaltet ELVIRA ebenfalls

Statistiken aus dem eigenen Datenbestand sowie Informationen aus Befragungen oder Auswertungen anderer Institutionen wie dem DIW oder dem ifo Institut.

Die Datenbank wird kontinuierlich beim Erscheinen neuer Statistiken aktualisiert und enthält somit bei allen Statistiken die am aktuellen Rand verfügbaren Zahlen. Auswertungen sind ausschließlich auf Branchen- und/oder Kreisebene möglich. Dabei ist die genaue Aufteilung abhängig von der jeweiligen Statistik. Ebenso ist von der jeweiligen Statistik abhängig, für welchen Zeitraum Auswertungen vorgenommen werden können und ob diese auf Monats- oder Jahresebene erfolgen. Die längsten Zeitreihen reichen dabei bis 1991 zurück.

6.3.1.2 Verfügbarkeit

ELVIRA ist eine Online-Datenbank, die für Mitglieder des Hauptverbandes kostenlos genutzt werden kann. Nichtmitglieder können einen kostenpflichtigen personalisierten Zugang beantragen, der in der Regel innerhalb eines Tages gewährt wird. Grundsätzlich kann zwischen den zwei Bezahlmodellen Abonnement und Einzelabrechnung der Zeitreihen gewählt werden. Zusätzlich kann einmalig ein Testzugang für einen Monat für 40 Euro beantragt werden. Die Bezahlmodelle sind in Tabelle 6.1 dargestellt.

Tabelle 6.1: Zugangskosten zu ELVIRA

Zugangsweg	Kosten
Abonnement	980 Euro/Jahr zzgl. 19 % MwSt.
Zeitreihen/Einzelabrechnung	Grundbeitrag: 30 Euro/Jahr zzgl. MwSt. Variable Kosten: 1,80 Euro/Zeitreihe zzgl. 19 % MwSt.
Testzugang	40 Euro zzgl. 19 % MwSt. für 1 Monat Einmalig möglich

Quelle: Eigene Darstellung auf Basis von Informationen von der Homepage des Hauptverbandes der Deutschen Bauindustrie e.V.⁶⁰

6.3.1.3 Merkmale zur Mindestlohnevaluation

Wie bereits beschrieben, setzt sich ELVIRA aus sehr vielen unterschiedlichen Datenquellen zusammen. Dementsprechend viele Themenbereiche werden auch abgedeckt. Bezüglich des Arbeitsmarktes können beispielsweise Statistiken zu Arbeitslosigkeit, offenen Stellen oder der Situation auf dem Ausbildungsmarkt ausgewertet werden. Zusätzlich sind Auswertungen zur Kurzarbeit, den Bruttolohnsummen sowie zur geleisteten Arbeitszeit möglich.

Weitere Merkmale hinsichtlich der wirtschaftlichen Entwicklung der Bauwirtschaft sind der Auftragsbestand, Statistiken zur Auslandstätigkeit der Branche, Baufertigstellungen, Baugenehmigungen, Bauinvestitionen, Betriebsanzahl, Insolvenzen, Konjunkturentwicklung, Kostenstruktur, öffentliche Ausgaben für Baumaßnahmen, Produktion und Umsatz im Bauhauptgewerbe sowie die Unternehmensstruktur und Jahresbauleistung.

⁶⁰ Seite: <http://www.bauindustrie.de/zahlen-fakten/datenbank-elvira/preise-und-leistungen/> [abgerufen am: 11.11.2013]

6.3.1.4 Merkmale zur Verknüpfung

Die verfügbaren Zeitreihen können aus Branchenebene – teilweise auch untergliedert – und auf Kreisebene mit anderen aggregierten Datenbeständen verknüpft werden.

6.3.1.5 Identifikation des Evaluationspotenzials

Das Potenzial der Datenbank ELVIRA für Mindestlohnevaluationen ist eingeschränkt. Die Datenbank beschränkt sich mit wenigen Ausnahmen auf das Bauhauptgewerbe. Zudem ist eine detaillierte branchenspezifische Aufteilung nur teilweise möglich. Allerdings gibt es Bestrebungen die Datenbank um das Ausbaugewerbe zu erweitern. Auch um eine Erweiterung der Statistiken, die nach Wirtschaftszweigen unterteilt werden können, wird sich bemüht.

Insgesamt bietet die Datenbank eine gute Ergänzung einer Mindestlohnevaluation für die Erstellung eines Branchenbildes in der Bauwirtschaft, da die wesentlichen Informationen zusammengefasst verfügbar sind. Die meisten Daten lassen sich aber auch auf anderem Wege, etwa direkt über die statistischen Ämter beziehen. Ausnahmen bilden hier zum Beispiel die Daten über den Auslandsbau, die vom Hauptverband selbst erhoben werden oder Informationen zur Ausbildung im Bau, die durch die SOKA-Bau aufbereitet werden.

6.4 Tarifarchiv des Wirtschafts- und Sozialwissenschaftlichen Instituts der Hans-Böckler-Stiftung (WSI-Tarifarchiv)

6.4.1.1 Allgemeine Datensatzinformationen

Das Tarifarchiv des Wirtschafts- und Sozialwissenschaftlichen Instituts der Hans-Böckler-Stiftung (WSI-Tarifarchiv) ist eine Sammlung sämtlicher Tarifverträge, die seit 1950 mit Gewerkschaften des Deutschen Gewerkschaftsbundes (DGB) als Tarifpartner geschlossen wurden. Teilweise sind Tarifverträge seit den 1920er Jahren erfasst. Neue Tarifabschlüsse werden zeitnah in das Tarifarchiv übernommen. Das Tarifarchiv ermöglicht dementsprechend eine detaillierte Rückverfolgung zu sämtlichen tarifrechtlichen Regelungen.

6.4.1.2 Verfügbarkeit

Ein Großteil des Tarifarchivs ist Online frei verfügbar und kann kostenlos genutzt werden. Für detailliertere Auswertungen besteht die Möglichkeit eines Gastwissenschaftleraufenthalts am WSI. Auch dieser Zugangsweg ist kostenlos und erfolgt ohne formalisierte Beantragung sondern in Absprache mit den zuständigen Mitarbeitern und Mitarbeiterinnen. Da kein formaler Antrag existiert, wird keine Zeit für die Bearbeitung der Anträge benötigt. Insbesondere bei der Nutzung des Zugangs über einen Gastwissenschaftleraufenthalt ist es jedoch zu empfehlen ausreichend Zeit einzuplanen, um genügend Flexibilität bei der Planung des Aufenthalts zu haben.

6.4.1.3 Merkmale zur Mindestlohnevaluation

Da die Mindestlohnregelungen nach dem Arbeitnehmer-Entsendegesetz (AEntG) auf Tarifabschlüssen aufbauen, beinhaltet das Tarifarchiv sowohl die für allgemein verbindlich erklärten Mindestlohntarifverträge als auch die zu dem Zeitpunkt gültigen Rahmentarifverträge, die vorangegangenen Tarifabschlüsse sowie die weitere Entwicklung der Abschlüsse.

6.4.1.4 Merkmale zur Verknüpfung

Im Tarifarchiv wurde eine grobe Zuteilung der Tarifverträge zu den Wirtschaftszweigen der Wirtschaftszweigsystematik vorgenommen. Über diese ist eine Verknüpfung des Tarifarchivs mit sämtlichen Datensätzen, die den Wirtschaftszweig enthalten möglich. Hierzu gehören unter anderem das BHP (Abschnitt 5.1.1), das URS (Abschnitt 5.1.3) sowie das SOEP (Abschnitt 4.3.6) und der Mikrozensus (Abschnitt **Fehler! Verweisquelle konnte nicht gefunden werden.**). Eine weitere Verknüpfungsmöglichkeit besteht anhand der Bezeichnung der Tarifverträge über Record Linkage mit der VSE (Abschnitt 4.2.1) enthalten sind.

6.4.1.5 Identifikation des Evaluationspotenzials

Für eigenständige Auswertungen und Analysen zum Mindestlohn ist das Tarifarchiv nicht geeignet. Allerdings bietet das Tarifarchiv eine Fülle an Informationen und Sachverhalten, die die Datenbasis anderer Datensätze sinnvoll erweitern kann und damit umfassendere und detailliertere Auswertungen mit diesen Datensätzen ermöglicht. Abhängig davon, wie zuverlässig die Zuordnung zu Wirtschaftszweigen ist, können Mindestlohnbranchen besser in anderen Datensätzen identifiziert werden.

Komplementär zum WSI-Tarifarchiv kann die Tarifdatenbank des Statistischen Bundesamtes⁶¹ genutzt werden. Diese enthält – im Unterschied zum WSI Tarifarchiv – ausschließlich Entgeltinformationen sowie Informationen zu Lohnzusatzleistungen, allerdings keine Angaben zu Arbeitszeiten oder anderen nicht monetären tariflichen Vereinbarungen. In Verbindung mit der VSE (4.2.1) ist die Verwendung der Tarifdatenbank des Statistischen Bundesamtes sinnvoll, da diese die Tarifvertragsnummer enthält, die auch in der VSE enthalten ist.

7 Nicht verwendbare Datensätze

7.1 Arbeiten und Lernen im Wandel (ALWA)

Der Datensatz Arbeiten und Lernen im Wandel (ALWA) wurde am IAB im Rahmen des Projekts „Qualifikationen, Kompetenzen und Erwerbsverläufe“ erhoben. Die Befragung wurde telefonisch zwischen August 2007 und April 2008 durch infas durchgeführt. Insgesamt wurden 10.404 Personen befragt. Die Grundgesamtheit besteht aus allen Personen, die zwischen 1956 und 1988 geboren sind und in Deutschland leben. In einem zweistufigen Stich-

⁶¹ <https://www.destatis.de/tdblive/> [abgerufen am: 27.11.2013].

probenverfahren wurden zunächst 250 Gemeinden ausgewählt und anschließend aus deren Einwohnermeldeamtsregistern eine Zufallsauswahl getroffen.

ALWA enthält retrospektive Angaben zu Bildungs- und Erwerbsverläufen sowie Informationen zu Wohnort, Partner und Kindern. Für die Mindestlohnforschung auf Branchenebene ist der Datensatz aber aus einer Reihe von Gründen ungeeignet. Die Konzeption als Bevölkerungsstichprobe und die Fallzahlen führen dazu, dass nur sehr wenige Befragte in einer Mindestlohnbranche beschäftigt sind. Der Anteil der direkt Betroffenen ist entsprechend noch geringer, auch weil ein gewisser Bildungsbias vorhanden ist (Antoni et al. 2010). Geeignete Lohninformationen sind in ALWA selbst nicht enthalten. Zwar ist eine Verlinkung mit administrativen Daten im Datenprodukt ALWA-ADIAB vorhanden. Da die Verlinkung aber auf die Person und nicht auf den Spell bezogen ist, ist mit zusätzlichen Unschärfen zu rechnen. ALWA wurde im Rahmen des NEPS (Unterabschnitt 7.10) zum Panel ausgebaut.

7.2 IAB-Stellenerhebung

Die IAB-Stellenerhebung (zuvor IAB-Erhebung des Gesamtwirtschaftlichen Stellenangebots) ist eine jährlich stattfindende Querschnittsbefragung von Betrieben in Deutschland, um repräsentative Daten zur Entwicklung des gesamtwirtschaftlichen Arbeitskräftebedarfs und von Stellenbesetzungsprozessen zu erhalten. Sie enthält außerdem Informationen zu abgebrochenen Suchvorgängen und der betrieblichen Einschätzungen aktueller arbeitsmarktpolitischer Entwicklungen (Müller et al., 2013). Die Befragung begann 1989 und wurde 1992 auf Ostdeutschland ausgeweitet. Die Grundgesamtheit bilden alle Betriebe mit mindestens einem sozialversicherungspflichtig Beschäftigten in Deutschland. Die Befragung ist disproportional nach 28 Wirtschaftszweigen (WZ03) bzw. 23 Wirtschaftszweigen (WZ08) sowie sieben Betriebsgrößenklassen (Anzahl der SV-Beschäftigten) für Ost- und Westdeutschland geschichtet.

Die IAB-Stellenerhebung ist momentan für die Mindestlohnforschung wenig nützlich, da die Brancheneinteilung für die meisten Untersuchungen zu grob ist. Da die befragten Betriebe jedes Jahr wechseln und die IAB-Stellenerhebung bislang nicht mit den anderen IAB-Datenprodukten (z.B. SIAB, BHP) verknüpft ist, können auch keine Entwicklungen für einzelne Betriebe untersucht oder aber die Mindestlohn Betroffenheit festgestellt werden. Eine solche Verknüpfung ist jedoch geplant und könnte in Zukunft Analysepotenzial ergeben.

7.3 Berufliche Weiterbildung als Bestandteil Lebenslangen Lernens (WeLL)

Beim Panel WeLL handelt es sich um eine telefonische Beschäftigtenbefragung, die im Rahmen des Projekts Berufliche Weiterbildung als Bestandteil Lebenslangen Lernens entwickelt wurde. Es handelt sich um eine Längsschnittbefragung mit vier Befragungszeitpunkten in 2007, 2008, 2009 und 2010, die von infas durchgeführt wurde. In der ersten Welle zwischen Herbst 2007 und Januar 2008 wurden ca. 6.400 Beschäftigte aus 149 Betrieben des IAB-Betriebspanels befragt. Das Erhebungsdesign ist eine zweistufige Stichprobe, bei der zunächst Betriebe geschichtet nach den Kriterien Betriebsgröße, Branche, Standort, Weiter-

bildungs- und Investitionsaktivitäten ausgewählt und im zweiten Schritt die Befragten in diesen Betrieben gezogen wurden.

WeLL umfasst Fragen zu Weiterbildungsaktivitäten und Erwerbsbiografie, aber auch zu Soziodemografie, Einkommen, Haushalt, Arbeitszufriedenheit und Zukunftserwartungen. Dennoch ist der Datensatz für die Mindestlohnevaluation auf Branchenebene nicht geeignet. Dies liegt vor allem daran, dass das Stichprobendesign für die Untersuchung innerbetrieblicher Weiterbildungsaktivitäten ausgelegt ist. Eine Repräsentativität auf der Ebene von Betrieben oder Beschäftigten ist nicht gegeben.

7.4 AFiD-Panel Industrieunternehmen

Das AFiD-Panel Industrieunternehmen setzt sich aus insgesamt drei verschiedenen Erhebungen mit Auskunftspflicht zusammen, das für die Jahre 2001 bis 2010 verfügbar ist (FDZ der Statistischen Ämter 2012b). Bei den Erhebungen handelt es sich um folgende Statistiken:

- Jahresbericht für (Mehrbetriebs-)Unternehmen im Bereich Verarbeitendes Gewerbe, Bergbau und Gewinnung von Steinen und Erden
- Investitionserhebung im Bereich Verarbeitendes Gewerbe, Bergbau und Gewinnung von Steinen und Erden
- Kostenstrukturerhebung im Bereich Verarbeitendes Gewerbe, Bergbau und Gewinnung von Steinen und Erden

Der Aufbau sowie die Definition der Grundgesamtheit gleichen dem Aufbau des AFiD-Panels Industriebetriebe (Unterabschnitt 5.3.1), mit dem Unterschied, dass die Bezugseinheit nicht der Betrieb sondern das Unternehmen ist. Während auch auf Unternehmensebene der Jahresbericht sowie die Investitionserhebung Vollerhebungen sind, ist die Kostenstrukturerhebung eine einstufig geschichtete Zufallsstichprobe von maximal 18.000 Unternehmen mit 20 oder mehr Beschäftigten (Statistisches Bundesamt 2013c).

Das AFiD-Panel Industrieunternehmen weist im Vergleich zum AFiD-Panel Industriebetriebe die gleichen Vor- und Nachteile hinsichtlich der Verwendbarkeit für die Mindestlohnforschung auf. Da der betriebliche Geltungsbereich der Mindestlöhne über den Tätigkeitsschwerpunkt der Betriebe und nicht der Unternehmen definiert wird, ist die Abgrenzung der Mindestlohnbranche im AFiD-Panel Industrieunternehmen ungenauer. Aus diesen Gründen ist das AFiD-Panel Industrieunternehmen für die Mindestlohnforschung nicht geeignet.

7.5 Einkommens- und Verbrauchsstichprobe (EVS)

Die Einkommens und Verbrauchsstichprobe (EVS) ist eine amtliche Befragung, die seit 1962/63 im Fünfjahresrhythmus durchgeführt wird. Im Unterschied zu anderen amtlichen Befragungen gibt es bei der EVS keine Auskunftspflicht. Es ist eine Kombination aus Online- und schriftlicher Befragung, die dezentral durch die Statistischen Landesämter durchgeführt wird (Statistisches Bundesamt 2012a). In den Jahren zwischen den einzelnen Wellen der

EVS wird die Laufende Wirtschaftsrechnung erhoben, die inhaltlich eng mit der EVS abgestimmt ist und somit eine Fortschreibung der EVS für die Zwischenperioden darstellt. Im Unterschied zu anderen Haushaltsbefragungen ist die EVS aber kein Paneldatensatz. Vor jeder Erhebung wird eine neue Stichprobe gezogen.

Die Stichprobe der EVS, 0,2 % aller Privathaushalte (77.648), ist zumindest für die großen Mindestlohnbranchen (Gebäudereinigung, Zeitarbeit, Pflege, Bauhauptgewerbe) ausreichend. Zwar enthält der Datensatz nur unzureichende Informationen über die Branche, in der die befragte Person arbeitet, sowie über die ausgeübte Tätigkeit, was aber mit einer Verknüpfung der EVS mit dem Mikrozensus behoben werden kann, da die Stichprobe der EVS aus den im Mikrozensus befragten Haushalten gezogen wird (Statistisches Bundesamt 2012a).

Trotz dieser sehr detaillierten Informationen über die Einkommenssituation und der technischen Verknüpfungsmöglichkeiten mit dem Mikrozensus ist die EVS für die Mindestlohnforschung nicht verwendbar. Dies hat mit allgemeinen methodischen Einwänden hinsichtlich der Stichprobenziehung der EVS zu tun. Da es sich hier um eine Untergruppe der bereits für den Mikrozensus befragten Personen handelt, die sich *freiwillig* zu einer weiteren umfangreichen Befragung bereit erklärt haben, ist mit sehr starken Selektivitätseffekten zu rechnen, die eine Repräsentativität der Auswertungen sehr stark einschränkt.

7.6 Daten der Zusatzversorgungskasse des Steinmetz- und Steinbildhauerhandwerks VVaG

Im Unterschied zu den Datensätzen der anderen untersuchten Sozialkassen, die in den Unterabschnitten 4.1.5 bis 4.1.7 dargestellt sind, sind die Daten der Zusatzversorgungskasse des Steinmetz- und Steinbildhauerhandwerks VVaG nicht für die Mindestlohnforschung geeignet. Vergleichbar mit den anderen Sozialkassen handelt es sich auch hier um prozessproduzierte Daten zu den rund 11.000 gewerblich Beschäftigten, die bei Betrieben im fachlichen und räumlichen Geltungsbereich des Tarifvertrages des Steinmetz- und Steinbildhauerhandwerks liegen. Die Daten werden jährlich erhoben und liegen seit 2005 als Panel auf Individualebene vor. Als Merkmale liegen die Sozialversicherungsnummern und die Lohnsummen, nicht aber die Arbeitszeit vor. Aufgrund der fehlenden Arbeitszeit ist eine Verwendung der Daten für die Mindestlohnforschung unnötig, da die vorhandenen Informationen auch in anderen Datensätzen wie den IEB (Unterabschnitt 4.1.1) vorliegen.

7.7 Datenbestand der Deutschen Rentenversicherung

Die Mikrodaten der Deutschen Rentenversicherung stammen wie die Prozessdaten der BA, die bereits im Unterabschnitt 4.1.1 beschrieben wurden, aus dem DEÜV-Meldeverfahren. Auch die Prozessdaten der Deutschen Rentenversicherung stehen über ein FDZ der Forschung zur Verfügung. Aus Datenschutzgründen und der Datensparsamkeit erhält die Deutsche Rentenversicherung ausschließlich die für sie relevanten Daten aus dem DEÜV-Meldeverfahren. Dementsprechend liegen ausschließlich beschäftigungsrelevante Merkmale

für das letzte Jahr einer Beschäftigung vor dem Renteneintritt vor. Da für eine Mindestlohnevaluation vor allem Daten aus der Beschäftigungsphase benötigt werden, ist der Datenbestand der Deutschen Rentenversicherung für die Mindestlohnforschung nicht geeignet.

Neben den Daten, die im FDZ vorliegen, ist die Deutsche Rentenversicherung für die Betreuung der Entsendestatistik zuständig. Auch dieser Datensatz ist für die Mindestlohnforschung nicht geeignet, da die Daten weder Angaben zu den Lohnsummen oder den Arbeitszeiten enthalten. Lediglich die Sozialversicherungspflicht sowie Angaben zu Arbeitnehmer/in und Arbeitgeber werden erfasst.

7.8 Mikrodaten bei der Deutschen Bundesbank

Dieser Bericht spart die Mikrodatenbestände bei der Deutschen Bundesbank aus. Die Mikrodatenbank Direktinvestitionen (MiDi) beinhaltet Informationen über Kapitalbeteiligungen deutscher Investoren an Unternehmen im Ausland und ausländischer Investoren an Unternehmen im Inland. Diese Informationen sind thematisch für die Mindestlohnevaluation nicht relevant.

Die Unternehmensbilanzen in der Datenbank Ustan werden nicht näher besprochen, da mit der Dafne-Datenbank bereits ein umfassender Datensatz zu deutschen Bilanzdaten besprochen wurde (siehe Abschnitt 5.4.8), dessen Verknüpfbarkeit mit anderen Datenprodukten nach momentanem Stand einfacher sein sollte. Ob und inwieweit Ustan Informationen bietet, die Dafne ergänzen oder dortige Angaben verifizieren könnten, wurde im Rahmen des Projekts nicht geklärt.

7.9 Mindestlohndatenbank des Wirtschafts- und Sozialwissenschaftlichen Instituts der Hans-Böckler-Stiftung

Die Mindestlohndatenbank des Wirtschafts- und Sozialwissenschaftlichen Instituts der Hans-Böckler-Stiftung ist eine Online-Datenbank,⁶² die einen internationalen Vergleich nationaler Mindestlöhne ermöglicht. In der Datenbank sind sämtliche Mitgliedsstaaten der EU, die einen nationalen Mindestlohn haben, vertreten. Zusätzlich sind die nationalen Mindestlöhne aus neun weiteren Ländern⁶³ vertreten (WSI 2013).

Die Datenbank erfasst die Mindestlöhne in den nationalen Währungen teilweise ab 1984 (USA). Zur besseren Vergleichbarkeit werden ab 2000 die jeweils zum 1. Januar des Jahres geltenden Mindestlöhne auf Stundenbasis mit Hilfe des Jahresdurchschnitts der Euro-Referenzkurse 2012 umgerechnet. Zusätzlich sind Informationen zu den Kaufkraftstandards 2011 für Vergleiche der Mindestlöhne anhand der Kaufkraft verfügbar. Neben den Mindest-

⁶² Die Online-Datenbank befindet sich unter http://www.boeckler.de/wsi-tarifarchiv_43610.htm [abgerufen am: 07.11.2013].

⁶³ Argentinien, Australien, Brasilien, Japan, Kanada, Korea, Neuseeland, Türkei und USA. In Kanada und den USA sind neben den nationalen Mindestlöhnen auch die in den Provinzen (Kanada; ab 2000) und in den Bundesstaaten (USA; ab 2009) gültigen Mindestlohnregelungen aufgeführt.

loohnhöhen enthält die Datenbank ebenfalls Informationen zur durchschnittlichen tarifvertraglichen Arbeitszeit (WSI 2013).

Für eine Evaluation der in Deutschland geltenden branchenspezifischen Mindestlöhne ist die Mindestlohn Datenbank ungeeignet, da ausschließlich nationale branchenübergreifende Mindestlöhne erfasst werden. Diese sind nicht mit branchenspezifischen Mindestlöhnen vergleichbar.

7.10 Nationales Bildungspanel (NEPS)

Das Nationale Bildungspanel (National Educational Panel Study – NEPS) ist eine groß angelegte Studie, mit der der Bildungserwerb und die Folgen für individuelle Lebensverläufe untersucht werden sollen. Im Rahmen unterschiedlicher Befragungen werden insgesamt 60.000 Teilnehmende mehrfach befragt (NEPS 2012). Diese teilen sich auf insgesamt sieben Startkohorten auf: Kindergarten (etwa 3.000), Klasse 5 (etwa 6.300), Klasse 9 (etwa 15.000, davon etwa 1.500 Förderschüler/innen), Studierende (etwa 18.300), Erwachsene (etwa 11.700), Zusatzstudie Thüringen (etwa 1.400) und Zusatzstudie Baden-Württemberg (etwa 1.300).⁶⁴ Da einige der Befragungen bereits während des Besuchs einer allgemeinbildenden Schule einsetzen, wären die Datensätze von der Konzeption gut geeignet um Bildungsentscheidungen zu untersuchen und inwieweit ein Mindestlohn in einer Branche die Bildungsentscheidungen beeinflusst. Für die Mindestlohnforschung wäre daher die Startkohorte Klasse 9 von besonderer Relevanz.

Trotz dieser Voraussetzungen ist das NEPS nicht für Studien zur Mindestlohnevaluation geeignet, da die ersten Befragungen 2007 stattfanden, was nach der Einführung der Mindestlöhne in den meisten Branchen liegt. Zusätzlich ist bei den verfügbaren Fallzahlen in der Startkohorte Klasse 9 nicht davon auszugehen, dass die Anzahl der Beobachtungen in einzelnen Mindestlohnbranchen ausreichend ist, um repräsentative und valide Aussagen treffen zu können.

8 Handlungsempfehlungen

Ziel dieses Gutachten war es neue Datenquellen für die Evaluation branchenspezifischer Mindestlöhne zu erschließen, das Analysepotenzial von bereits für die Evaluation genutzten Datenquellen zu bewerten und mögliche Erweiterungs- bzw. Verbesserungsvorschläge herauszuarbeiten. Hierfür wurden primär leitfadengestützte Experteninterviews mit Datenanbietern in Deutschland geführt und systematisch ausgewertet. Zusätzlich wurde mit Hilfe statistischer Methoden das Analysepotenzial von Veränderungen in bereits genutzten Daten (z.B. die Einführung der Stundenlöhne in den DEÜV-Meldungen) untersucht.

⁶⁴ Die Zusatzstudien in Baden-Württemberg und Thüringen betreffen ausschließlich Schüler/innen, die die (Fach-)Hochschulreife anstreben.

Insgesamt wurden auf dieser Basis mit 13 Datenanbietern Gespräche über das jeweilige Datenangebot geführt und 43 Datensätze sowie die allgemeinen Datenbestände der Statistischen Ämter, der Statistik der Bundesagentur für Arbeit sowie des FDZ der Deutschen Rentenversicherung und der Deutschen Bundesbank untersucht und auf ihre Verwendungsmöglichkeiten für die Mindestlohnforschung durchleuchtet. Für die Bewertung des Analysepotenzials der betrachteten Datensätze hinsichtlich der Evaluation branchenbezogener Mindestlöhne wurde geprüft, ob die Datensätze klar definierte Beobachtungseinheiten, geeignete Ergebnisvariablen und eine adäquate zeitliche Abdeckung aufweisen. Als Beobachtungseinheiten waren Beschäftigte, Betriebe, Regionen oder Wirtschaftszweige von Interesse. Als Ergebnisgrößen wurde eine umfangreiche Reihe an Indikatoren betrachtet, die auch bislang noch nicht betrachtete Ergebnisdimensionen wie beispielsweise Arbeitszufriedenheit umfasst. Im Hinblick auf die zeitliche Abdeckung war relevant, ob die Datensätze für einen Zeitraum vorliegen, der im Idealfall die Einführung und die Geltungsdauer von bestehenden Branchenmindestlöhnen umfasst oder sich zumindest auf eine Periode vor und nach mindestens einer Mindestlohnerhöhung während des Geltungszeitraums erstreckt.

Bei der (qualitativen) Gesamtbeurteilung der Geeignetheit aller untersuchten Datenquellen wurde zunächst unterschieden, ob sich die Datensätze in ihrer jeweils vorliegenden Form für Kausalanalysen oder lediglich für deskriptive Untersuchungen eignen. Dabei wurde jeweils herausgearbeitet, wo etwaige Schwachstellen liegen und inwieweit dies die Geeignetheit der Datensätze einschränkt. Darüber hinaus wurde unterschieden, ob es *technisch* möglich ist, einen bestimmten Datensatz mit anderen Datenquellen über eindeutige Identifikatoren oder Record Linkage zu verknüpfen, um das Analysepotenzial zu erweitern.

In der Gesamtschau kann festgehalten werden, dass es zwar nicht einen allumfassenden Datensatz für die Evaluation von Mindestlöhnen in Deutschland gibt, die existierenden Datensätze insgesamt jedoch eine ausreichende Basis für eine Evaluation darstellen, sofern sie sinnvoll miteinander kombiniert werden (dürfen). Dies impliziert zunächst, dass es keiner umfassenden und mit enormen Kosten verbundenen Primärerhebung speziell für eine Mindestlohnevaluation bedarf, sondern einer „kreativen“ Nutzung bereits vorhandener Datenquellen. Eine solche „kreative“ Nutzung setzt allerdings Modifikationen bei einigen dieser Datenquellen voraus und hängt entscheidend davon ab, dass diese Datenquellen zusammengeführt werden (dürfen). Sind diese Voraussetzungen erfüllt, so lässt sich eine Datenbasis generieren, die für die Evaluation branchenspezifischer Mindestlöhne geeignet ist.

Welche Modifikation im Einzelnen geeignet ist, hängt stark von der zu untersuchenden Branche und Fragestellung ab. Für die Handwerksbranchen erscheint beispielsweise eine Erweiterung der VSE sinnvoll, da in diesen Branchen die Betriebe sehr häufig nur wenige Beschäftigte haben und die VSE aufgrund der Abschneidegrenze diese Branchen nur unzureichend repräsentiert. Schwierigkeiten bei der Abgrenzung der Pflegebranche könnten hingegen vermindert werden, wenn eine Verknüpfung von Sozialversicherungs- und Umfragedaten mit der Pflegestatistik gelänge.

Die beiden Hauptanbieter der für eine Mindestevaluation zu verwendenden Datenquellen sind zum einen die Statistischen Ämter und zum anderen das Institut für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit. Ergänzend könnten noch – überwiegend für Detailanalysen – Unternehmensdaten von Bureau van Dijk, sowie Daten der Sozialkassen relevant werden. Darüber hinaus erscheint es sinnvoll, auf Basis dieses Datenmaterials punktuell ergänzende und mit überschaubaren Kosten verbundene Befragungen durchzuführen, mit deren Hilfe vertiefende Informationen zu bestimmten Aspekte oder Sachverhalten erhoben werden können, wenn die Befragungsinformationen mit diesen Datensätzen verknüpfbar sind.

Von den Statistischen Ämtern sollten das Unternehmensregister, der Mikrozensus und die Verdienststrukturerhebung für eine Evaluation des Mindestlohns herangezogen werden. Zumindest für die Verdienststrukturerhebung muss allerdings das Erhebungsprogramm angepasst werden. Es muss sichergestellt werden, dass auch Unternehmen mit weniger als zehn Beschäftigten, die für die Mindestlohnevaluation von besonderer Bedeutung sind, in das Befragungsprogramm aufgenommen werden. Auch sollte über eine Modifikation oder eine Erweiterung des Befragungsprogramms nachgedacht werden, um umfassender die relevanten Dimensionen für eine Evaluation abzudecken.

Dies ist möglicherweise nur über eine Änderung des Verdienststatistikgesetzes, das die Ausgestaltung und das Frageprogramm der VSE regelt, zu erreichen. Ähnliches gilt, wenn der Mikrozensus um weitere Ergebnisindikatoren wie Angaben zum Konsum- und Sparverhalten sowie detailliertere Angaben zur regionalen und branchenbezogenen Mobilität der Arbeitnehmer/innen erweitert werden soll. Die Nutzbarkeit des Mikrozensus würde sich auch deutlich erhöhen, wenn die Betroffenheit von Mindestlöhnen durch entsprechende Fragen genauer identifiziert werden könnte. Wie Tabelle 8.1 zeigt, decken die fünf vorgeschlagenen Datensätze einen Großteil der relevanten Ergebnisindikatoren ab. Diese können mit themenspezifischen Datensätzen abgedeckt werden, die in der Regel mit dem URS verknüpft werden können.

Seitens des Instituts für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit sind die Integrierten Erwerbsbiografien und das Betriebs-Historik-Panel für eine Evaluation des Mindestlohns von herausragender Bedeutung. Beide Datenquellen speisen sich aus den Meldungen zur Sozialversicherung (DEÜV), deren Qualität weiter zu prüfen sein wird, was möglicherweise dazu führen wird, dass das Meldeverfahren zumindest für die Stundenlöhne präzisiert werden muss.

Der wichtigste Bestandteil unserer Empfehlung besteht jedoch darin, die Grundlage für ein Zusammenspielen der in Tabelle 8.1 aufgeführten Datenquellen zu schaffen. Wie im Bericht ausgeführt, erscheint dies technisch möglich, wobei es für die Identifikation etwaiger Fallstricke und Detailprobleme eines entsprechenden Testlaufs bedarf (vgl. hierzu auch unten). Den zentralen Hemmschuh stellen derzeit allerdings datenschutzrechtliche Hürden dar, die ausgeräumt werden müssen. Denn nur durch die Kombination der benannten Datenquellen

ergibt sich ein Analysedatensatz, der den Anforderungen für eine umfassende und stichhaltige Ergebnisse generierende Mindestlohnevaluation genügt. Zur Schaffung der rechtlichen Rahmenbedingungen für eine entsprechende Verknüpfung erscheint ein zum Zensus vergleichbares Gesetz für die Mindestlohnevaluation notwendig.

Tabelle 8.1: Ergebnisindikatoren in den wichtigsten Mikrodatsätzen

	IAB		Statistische Ämter		
	IEB	BHP	URS	VSE	Mikrozensus
Individualebene					
Einkommen: Stundenlohn/Entgelte	Bruttoentgelt, kein Stundenlohn	Aggregiert auf Betriebsebene		Ja	Nettoeinkommen
Arbeitszeit: tariflich und/oder tatsächlich	seit 2010 geleistete Stunden, aber Definition vom Bedarf abweichend			Tarifliche und tatsächliche Arbeitszeit	Ja, auf Stundenbasis
Qualifikation/Tätigkeitsbereiche	Vorhanden	Vorhanden		Vorhanden	Vorhanden
Wirtschaftszweige/tarifvertragliche Regelungen	Wirtschaftszweige	Wirtschaftszweige	Wirtschaftszweige	Wirtschaftszweige, Tarifvertragsschlüssel und ID für Mindestlohnbranche	Wirtschaftszweige (3-Steller)
Qualität von Beschäftigungsverhältnissen (COPSOQ-/ERI-Ansatz)					Gesundheitszustand, Schichtarbeit, Wochenend-/Abendarbeit, Weiterbildung
Arbeitsmotivation/Produktivität					Zufriedenheit mit Arbeitszeit, Wunsch nach Erhöhung/ Reduktion der Arbeitszeit
Arbeitsbeziehungen					Arbeitsabteilung im Betrieb, Stellung im Beruf
Bildungsentscheidung junger Kohorten	Ausbildungsstand/Änderungen ab erster Beschäftigung bekannt				
Regionale und branchenbezogene Mobilität	ja				Arbeitsweg

Quelle: Eigene Darstellung

noch Tabelle 8.1: Ergebnisindikatoren in den wichtigsten Mikrodatsätzen

	IAB		Statistische Ämter		
	IEB	BHP	URS	VSE	Mikrozensus
Betriebsebene					
Umsatz			Ja		
Beschäftigungsstruktur (Qualifikation, Alter, Anzahl Beschäftigter)	Berechenbar	Ja		Berechenbar	
Marktein- und -austritte		Ja	Ja		
Output und Investitionen					
Preise					
Gesamtwirtschaft/Branche					
Konsumausgaben/Ersparnis					
Schattenwirtschaft					
Steuereinnahmen (nach Steuerarten)					
Sozialversicherungsbeiträge	Aus dem Brutto-lohn abschätzbar			Ja	

Quelle: Eigene Darstellung

Darüber hinaus müssen die technischen Möglichkeiten für ein derartiges Zusammenspielen geschaffen werden. Hierfür dürfte eine Treuhänderstelle einzurichten sein, die diese Daten nach internationalen Standards zusammenführt. Eine Zusammenführung der Daten des BHP, des URS und der VSE auf Betriebsebene ist unproblematisch, da die URS als Schnittstelle sowohl die BA-Betriebsnummer (BHP) als auch eine eigene Identifikationsnummer enthält, die in der VSE ebenfalls integriert ist.

Eine Verknüpfung auf Individualebene erscheint zwar machbar, aber alles andere als trivial, da das Zusammenführen momentan für Personen nur über Namen, Adressen und Geburtsangaben möglich ist und hierzu Record Linkage Techniken zum Einsatz kommen müssen, für die beide Datenanbieter durch den Zensus bzw. das German Record Linkage Center (GRLC) jedoch über ausreichende Vorerfahrungen verfügen (siehe Exkurs). Schließlich erachten wir es für notwendig, die skizzierten Schritte baldmöglichst zu erproben und auch vertiefende Analysen zur Qualität der hierdurch generierten Datensätze durchzuführen.

Exkurs: German Record Linkage Center (GRLC)

Die Möglichkeit einer Verknüpfung unterschiedlicher Mikrodatenquellen wird in den Wirtschafts- und Sozialwissenschaften zunehmend genutzt. Die allgemein anerkannten Vorteile des Record Linkage von verschiedenen Datenquellen sind u.a. die Gewinnung von bereits vorhandenen Informationen, die sonst mit hohem Aufwand erneut erhoben werden müssten und die dadurch reduzierte Belastung von Befragten. Dies wiederum führt auch zu einer höheren Akzeptanz gegenüber der ursprünglichen Befragung.

Eine der deutschlandweit führenden Einrichtungen auf dem Feld des Record Linkage ist das German Record Linkage Center (GRLC, vgl. www.record-linkage.de). Das GRLC wird im Rahmen eines von der Deutschen Forschungsgemeinschaft (DFG) geförderten Drittmittelprojektes an zwei verschiedenen Standorten geführt: Der Standort am Lehrstuhl von Prof. Dr. Rainer Schnell (Universität Duisburg-Essen) führt Forschung zu Methoden des Record Linkage durch. Die dabei entstehende Methodenentwicklung, z.B. auf den Feldern des effizienten Blockings bei der anonymen Verknüpfung sehr großer Datensätze sowie generell zum Thema Privacy Preserving Record Linkage, findet international großen Anklang, sowohl durch vielzitierte Publikationen als auch durch zahlreiche Vorträge auf renommierten Konferenzen.

Diese Forschungsergebnisse werden am zweiten Standort des GRLC unmittelbar in praktischen Record Linkage-Projekten angewendet. Dieser Standort am FDZ der BA im IAB wird von Manfred Antoni und Stefan Bender geleitet und fungiert als Serviceeinrichtung. Zum Angebot zählen u.a. Record Linkage-Schulungen, die Beratung von externen Record Linkage-Projekten und die Durchführung von Datenverknüpfungen im Auftrag externer Forscher/innen. Der Standort Nürnberg allein hat bis heute knapp 20 solcher Auftragsverknüpfungen durchgeführt. Die meisten dieser Aufträge bezogen sich auf die Verknüpfung von externen Datenquellen (Befragungen, Daten von kommerziellen Anbietern etc.) mit den administrativen Daten des IAB. Dabei wurden sowohl Verknüpfungen auf der Personen- als auch auf der Betriebsebene durchgeführt, letztere meist durch eine Aggregation auf die höhere Unternehmensebene. Die Erfolgsquoten bei Verknüpfungen von Betriebsdaten liegen zwischen 50% und 80%, die bei Personendaten bei knapp 80% bis 98%. Beispielsweise lag die Erfolgsquote bei einem Verknüpfungsprojekt für den Datensatz ALWA bei 86%. Für PASS lag sie wellenabhängig bei bis zu 90%.

Eine besondere Form von Auftragsverknüpfungen stellen Projekte dar, in denen das GRLC eine Treuhänderfunktion übernimmt. Dabei werden die Datenbanken zweier oder mehrerer Auftraggeber durch das GRLC verknüpft und die dabei entstehende pseudonymisierte Zuordnung der Beobachtungseinheiten der Datenlieferanten an die Ausgangs- oder weitere Parteien übermittelt. So können die Forschungsdaten der ursprünglichen Datenhalter verknüpft und gemeinsam ausgewertet werden, während deren datenschutzrechtlichen Ansprüchen voll Rechnung getragen wird.

Der Grad der Anonymität der zu verknüpfenden Einheiten kann noch einmal gesteigert werden, indem ein Privacy Preserving Record Linkage durchgeführt wird. Dabei werden die identifizierenden Merkmale bereits bei den Dateneignern mit einem abgestimmten Algorithmus verschlüsselt. Das GRLC als Treuhänder erhält in diesem Fall ausschließlich die verschlüsselten Merkmale und einen Satz von Pseudonymen pro Ausgangsdatensatz, um eine spätere Zuordnung der verknüpften Einheiten zu ermöglichen. Bei entsprechender Verschlüsselung ist es dem GRLC nach menschlichem Ermessen unmöglich, Rückschlüsse auf die Originalidentifikatoren zu ziehen. Das GRLC führt anschließend auf Grundlage der verschlüsselten Merkmale ein Record Linkage durch und verfährt im Weiteren wie bei sonstigen Projekten im Treuhänderverfahren.

Der Standort des GRLC am FDZ der BA im IAB ist aus verschiedenen Gründen prädestiniert für eine solche Funktion des Treuhänders: zum einen findet das Record Linkage dort auf einem gesicherten Server statt, der den Ansprüchen des IAB an Datensicherheit entspricht. Die IT-Sicherheitsmaßnahmen des IAB wurden in Anlehnung an das IT-Sicherheitshandbuch des Bundesamtes für die Sicherheit in der Informationstechnik (BSI) der Schutzbedarfsklasse pD-3/ VS-V für Daten nach § 3 Abs. 9 des Bundesdatenschutzgesetzes (BDSG) eingerichtet. Zum anderen hat das GRLC ein kryptografisches Verfahren auf Grundlage von so genannten Bloom-Filtern entwickelt, das ein fehlertolerantes Record Linkage unter Verwendung von verschlüsselten Identifikatoren erlaubt.

Zusammenfassend schlagen wir folgendes vor:

- Betriebsbefragungen mit unteren Abschneidegrenzen (z.B. VSE) sollten so ausgeweitet werden, dass sie auch für Branchen mit relativ kleiner durchschnittlicher Betriebsgröße repräsentative Ergebnisse liefern können.
- Datenquellen, aus denen sich keine für die Mindestlohnforschung nutzbare Stundenlohninformation ableiten lassen, sollten (wenn möglich) durch entsprechende Merkmale ergänzt werden.
- Bei Befragungen, deren Nutzbarkeit trotz relevanter Merkmale bisher vor allem durch zu große Abstände zwischen den Befragungswellen eingeschränkt war, sollte die Periodizität der Befragung entsprechend erhöht werden.
- Es sollten Rahmenbedingungen geschaffen werden, unter denen eine Verknüpfung verschiedener Datenquellen bei Einhaltung datenschutzrechtlicher Standards möglich ist. Dies ist vor allem dann von Bedeutung, wenn wichtige Gründe den ersten drei Erweiterungsvorschlägen im Wege stehen.

Diese Handlungsempfehlungen bilden eine wichtige Grundlage für die Fortführung der Mindestlohnforschung in Deutschland. Sie sind darüber hinaus auch weitgehend unabhängig davon, welche Themenkomplexe in zukünftigen Evaluierungen im Vordergrund stehen sollen.

Literatur

- Almus, Matthias; Dirk Engel und Susanne Prantl (2000), Die „Mannheimer Gründungspanels“ des Zentrums für Europäische Wirtschaftsforschung, ZEW-Dokumentation Nr. 00-02, Mannheim.
- Antoni, Manfred; Drasch, Katrin; Kleinert, Corinna; Matthes, Britta; Ruland, Michael; Trahms, Annette (2010): [Arbeiten und Lernen im Wandel](#). Teil I: Überblick über die Studie - März 2011 (2. aktualisierte Fassung des Berichtes vom Mai 2010). (FDZ-Methodenreport, 05/2010 (de)), Nürnberg, 49 S.
- Berg, M., Cramer, R., Dickmann C., Gilberg, R., Jesske, B., Kleudgen, M., Bethmann, A., Fuchs, B., Trappmann, M. und Wurdack, A. (2012), Codebuch und Dokumentation des „Panel Arbeitsmarkt und soziale Sicherung (PASS)“ – Datenreport Welle 5, FDZ-Datenreport 06/2012, Nürnberg.
- BIBB, BAuA (2012), BIBB/BAuA-Erwerbstätigenbefragung 2012 – Arbeit und Beruf im Wandel – Erwerb und Verwertung beruflicher Qualifikation, Informationsblatt des BIBB und der BAuA zum Datensatz, Bonn/Dortmund.
- Blossfeld, Hans-Peter (1985), Bildungsexpansion und Berufschancen, CAMPUS, Frankfurt.
- Bundesagentur für Arbeit (2011) Schlüsselverzeichnis für die Angaben zur Tätigkeit in den Meldungen zur Sozialversicherung – Ausgabe 2010, Nürnberg.
- BMAS (2013a), Merkblatt über Beitragsverpflichtungen zu tariflichen Sozialkassen – Stand: April 2013, Berlin.
- BMAS (2013b), Übersicht über die geltenden tariflichen Mindestlöhne – Stand: September 2013, Berlin.
- DGUV (2013), DGUV-Statistiken für die Praxis 2012 – Aktuelle Zahlen und Zeitreihen aus der Deutschen Gesetzlichen Unfallversicherung, Berlin.
- Dundler, A. und Frank, T. (2012), Beschäftigungsstatistik - Umstellung der Erhebungsinhalte bei den Merkmalen „ausgeübte Tätigkeit“ (Beruf), „Arbeitszeit“ und „Ausbildung“, Methodenbericht der Statistik der Bundesagentur für Arbeit, Nürnberg.
- FDZ der Statistischen Ämter des Bundes und der Länder (2009a), Investitionserhebung im Bauhauptgewerbe - Metadaten für die On-Site-Nutzung, FDZ der Statistischen Ämter der Länder Regionaler Standort Baden-Württemberg, April 2009, Stuttgart.
- FDZ der Statistischen Ämter des Bundes und der Länder (2009b), Investitionserhebung im Ausbaugewerbe - Metadaten für die On-Site-Nutzung, FDZ der Statistischen Ämter der Länder Regionaler Standort Baden-Württemberg, April 2009, Stuttgart.
- FDZ der Statistischen Ämter des Bundes und der Länder (2010a), Metadaten AFiD-Panel Dienstleistungen, FDZ der Statistischen Ämter der Länder Regionaler Standort Berlin-Brandenburg, Juni 2010, Berlin.
- FDZ der Statistischen Ämter des Bundes und der Länder (2010b), Metadaten AFiD-Panel Industriebetriebe, FDZ der Statistischen Ämter der Länder Regionaler Standort Sachsen-Anhalt, 20. Juli 2010, Halle (Saale).
- FDZ der Statistischen Ämter des Bundes und der Länder (2010c), AFiD-Panel Unternehmensregister – Metadaten für die On-Site-Nutzung, FDZ der Statistischen Ämter der Länder Regionaler Standort Niedersachsen, 5. November 2010, Hannover.

- FDZ der Statistischen Ämter des Bundes und der Länder (2011), Schlüsselverzeichnis des Mikrozensus 2011, FDZ der Statistischen Ämter der Länder Regionaler Standort Nordrhein-Westfalen, Düsseldorf.
- FDZ der Statistischen Ämter des Bundes und der Länder (2012a), Amtliche Firmendaten für Deutschland (AFiD) – Neue Daten für die Wirtschafts- und Umweltforschung, Juli 2012, Berlin.
- FDZ der Statistischen Ämter des Bundes und der Länder (2012b), Metadaten AFiD-Panel Industrieunternehmen, FDZ der Statistischen Ämter der Länder Regionaler Standort Sachsen-Anhalt, 2. Oktober 2012, Halle (Saale).
- FDZ der Statistischen Ämter des Bundes und der Länder (2013a), Verdienststrukturerhebung 2006 und 2010 – EVAS 62111, Metadaten für die On-Site-Nutzung, FDZ der Statistischen Ämter der Länder Regionaler Standort Hessen, 22. März 2013, Wiesbaden.
- FDZ der Statistischen Ämter des Bundes und der Länder (2013b), Mikrozensus 2012, FDZ der Statistischen Ämter der Länder Regionaler Standort Nordrhein-Westfalen, Juli 2013, Düsseldorf.
- FDZ der Statistischen Ämter des Bundes und der Länder (2013c), Umsatzsteuerstatistik – Voranmeldungen 2011, EVAS 73311, Metadaten für die On-Site-Nutzung, FDZ der Statistischen Ämter der Länder Regionaler Standort Hessen, 8. August 2013, Wiesbaden.
- Gerhards, Christian, Sabine Mohr, Anett Friedrich, Klaus Troltsch und Alexander Christ (2013), BIBB-Qualifizierungspanel – Version 1, BIBB-FDZ Daten- und Methodenbericht Nr. 2/2013, Bonn.
- Haisken-DeNew, John P. und Joachim R. Frick (2005), DTC – Desktop Companion to the German Socio-Economic Panel (SOEP), Datendokumentation des FDZ des SOEP am DIW, Version 8.0, Dezember 2005, Berlin.
- Huber, M. und Schmucker, A. (2012), Panel „WeLL“ - Arbeitnehmerbefragung für das Projekt „Berufliche Weiterbildung als Bestandteil Lebenslangen Lernens“ – Dokumentation für die Originaldaten Wellen 1-4, FDZ-Datenreport 03/2012, Nürnberg.
- IAB, RWI, ISG (2011), Evaluation bestehender gesetzlicher Mindestlohnregelungen - Branche: Bauhauptgewerbe, Abschlussbericht an das Bundesministerium für Arbeit und Soziales, 11. Oktober 2011, Nürnberg/Essen/Köln.
- IAQ (2011a), Evaluation bestehender gesetzlicher Mindestlohnregelungen - Branche: Gebäudereinigung, Abschlussbericht an das Bundesministerium für Arbeit und Soziales, 31. August 2011, Duisburg.
- IAQ (2011b), Evaluation bestehender gesetzlicher Mindestlohnregelungen - Branche: Wäschereidienstleistungen im Objektkundengeschäft, Abschlussbericht an das Bundesministerium für Arbeit und Soziales, 31. August 2011, Duisburg.
- IAW (2011a), Evaluation bestehender gesetzlicher Mindestlohnregelungen - Branche: Elektrohandwerk, Abschlussbericht an das Bundesministerium für Arbeit und Soziales, 31. August 2011, Tübingen.
- IAW (2011b), Evaluation bestehender gesetzlicher Mindestlohnregelungen - Branche: Maler- und Lackiererhandwerk, Abschlussbericht an das Bundesministerium für Arbeit und Soziales, 31. August 2011, Tübingen.
- IAW (2011c), Evaluation bestehender gesetzlicher Mindestlohnregelungen - Branche: Pflege, Abschlussbericht an das Bundesministerium für Arbeit und Soziales, 31. August 2011, Tübingen.

- ISG (2012), Dokumentation „Expertenworkshop zur Einleitung der zweiten Runde der Evaluation bestehender gesetzlicher Mindestlohnregelungen“, 6. und 7. November am BMAS, Berlin.
- Janik, Florian; Kohaut, Susanne (2012): Why don't they answer? Unit non-response in the IAB Establishment Panel. In: Quality and Quantity. International Journal of Methodology, Vol. 46, No. 3, S. 917-934.
- Lincke, Hans-Joachim; Martin Vomstein, Arianne Haug und Matthias Nübling (2013), Arbeitsplatzevaluierung psychischer Faktoren mit COPSQ, Österreichisches Forum Arbeitsmedizin, Nr. 01/13.
- Müller, Walter; Uwe Blien; Peter Knoche; Heike Wirth (1991), Die faktische Anonymität von Mikrodaten, Stuttgart: Metzler-Poeschel.
- Müller, Anne; Rebien, Martina; Vogler-Ludwig, Kurt (2013): IAB-Erhebung des gesamtwirtschaftlichen Stellenangebots. Betriebsbefragung zu Stellenangebot und Besetzungsprozessen. Wellen 2000 bis 2010 und Folgequartale ab 2006. FDZ-Datenreport, 03/2013 (de), Nürnberg.
- NEPS (2012), Das Nationale Bildungspanel – Notwendigkeit, Grundzüge und Analysepotenzial, Universität Bamberg.
- Nübling, Matthias; Martin Vomstein, Thomas Nübling, Ulrich Stöbel, Hans-Martin Hasselhorn und Friedrich Hofmann (2011), Erfassung psychischer Belastungen anhand eines erprobten Fragebogens-Aufbau der COPSQ-Datenbank, Schriftenreihe der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin, Forschungsbericht Fb2031, Freiburg/Dortmund.
- Paulus, W. und Matthes, B. (2013), Klassifikation der Berufe - Struktur, Codierung und Umsteigeschlüssel, FDZ Methodenreport Nr. 08/2013, Nürnberg.
- Rattenhuber, Pia (2011), Building the Minimum Wage in Germany. Germany's First Sectoral Minimum Wage and its Impact on Wages in the Construction Industry, DIW-Discussion paper 1111/2011, Berlin.
- Rohrbach-Schmidt, Daniela (2009), The BIBB/IAB- and BIBB/BAuA-Surveys on the Working Population on Qualification and Working Conditions in Germany, Version 1.1, BIBB-FDZ Daten- und Methodenbericht Nr. 1/2009, Bonn.
- Rohrbach-Schmidt, Daniela und Anja Hall (2013), BIBB/BAuA-Erwerbstätigenbefragung 2012, Version 1.1, BIBB-FDZ Daten- und Methodenbericht Nr. 1/2013, Bonn.
- Statistisches Bundesamt (1992), Klassifikation der Berufe 1992 (KldB 92) - Gliederungsstruktur bis zur 4-Steller-Ebene, Ausgabe 1992, Statistisches Bundesamt, Wiesbaden.
- Statistisches Bundesamt (2007), Statistik über ambulante Pflegeeinrichtungen (Pflegedienste) und Statistik über stationäre Pflegeeinrichtungen (Pflegeheime), Qualitätsbericht des Statistischen Bundesamtes, August 2007, Wiesbaden.
- Statistisches Bundesamt (2008), Absolute Anonymisierung der Steuerstatistik – CAMPUS-File der Einkommenssteuer 2001, Dokumentation des Statistischen Bundesamtes, 16. Januar 2008, Wiesbaden.
- Statistisches Bundesamt (2011), Unternehmensregister – System 95, Qualitätsbericht des Statistischen Bundesamtes, 23. September 2011, Wiesbaden.
- Statistisches Bundesamt (2012a), Einkommens- und Verbrauchsstichprobe 2008, Qualitätsbericht des Statistischen Bundesamtes, 12. Juni 2012, Wiesbaden.

- Statistisches Bundesamt (2012b), Lohn- und Einkommenssteuerstatistik, Qualitätsbericht des Statistischen Bundesamtes, 20. August 2012, Wiesbaden.
- Statistisches Bundesamt (2013a), Faktische Anonymisierung der Steuerstatistik (FAST) – Lohn- und Einkommenssteuer 2007, Dokumentation des Statistischen Bundesamtes, April 2013, Wiesbaden.
- Statistisches Bundesamt (2013b), Umsatzsteuerstatistik (Vorankündigungen), Qualitätsbericht des Statistischen Bundesamtes, 4. April 2013, Wiesbaden.
- Statistisches Bundesamt (2013c), Produzierendes Gewerbe – Kostenstruktur des Verarbeitenden Gewerbes sowie des Bergbaus und der Gewinnung von Steinen und Erden, Qualitätsbericht des Statistischen Bundesamtes, 4. Juni 2013, Wiesbaden.
- Statistisches Bundesamt (2013d), Verdienststrukturerhebung – Erhebung der Struktur der Arbeitsverdienste nach § 4 Verdienststatistikgesetz, Qualitätsbericht des Statistischen Bundesamtes, 2. Juli 2013, Wiesbaden.
- Statistisches Bundesamt (2013e), Strukturerhebung im Dienstleistungsbereich – Berichtsjahr 2011, Qualitätsbericht des Statistischen Bundesamtes, 15. Juli 2013, Wiesbaden.
- Statistisches Bundesamt (2013f), Verdienste und Arbeitskosten – Verdienststrukturen 2010, Fachserie 16, 2. August 2013, Wiesbaden.
- TNS Infratest Sozialforschung (2012), SOEP 2011 – Methodenbericht zum Befragungsjahr 2011 (Welle 28) des Sozio-Oekonomischen Panels, München.
- Uhly, Alexandra; Naomi Gericke und Nicole Lissek (2012), Erläuterungen zum „Datensystem Auszubildende“ (DAZUBI) – Auszubildenden-Daten der Berufsbildungsstatistik der statistischen Ämter des Bundes und der Länder (Erhebung zum 31.12.), Berufsmerkmale und Berechnungen des BIBB, Datenstand: 2011, Bundesinstitut für Berufsbildung, 27. November 2012, Bonn.
- WSI (2013), WSI Mindestlohndatenbank, Oktober 2013, Düsseldorf.
- ZEW, infas (2011a), Evaluation bestehender gesetzlicher Mindestlohnregelungen - Branche: Abfallwirtschaft, Abschlussbericht an das Bundesministerium für Arbeit und Soziales, 14. Oktober 2011, Mannheim.
- ZEW, infas (2011b), Evaluation bestehender gesetzlicher Mindestlohnregelungen - Branche: Dachdeckerhandwerk, Abschlussbericht an das Bundesministerium für Arbeit und Soziales, 14. Oktober 2011, Mannheim.
- ZEW (2012), Die Bereitstellung von Standardauswertungen zum Gründungsgeschehen in Deutschland für externe Datennutzer, Zentrum für Europäische Wirtschaftsforschung, Version 2012-12, Dezember 2012, Mannheim.

Anhang A.1: Leitfaden für die Expertengespräche

Die Gespräche unterteilen sich nach den folgenden Beobachtungseinheiten:

- a) Arbeitnehmer
- b) Betriebe
- c) Regionen (z.B. Kreise)
- d) Wirtschaftszweige

Zunächst soll also nach Datensätzen für Arbeitnehmer gefragt werden und für *jeden einzelnen Datensatz separat* die Liste der zu klärenden Punkte (s.u.) abgearbeitet werden. Danach nach Datensätzen für Betriebe, wobei wiederum für jeden Datensatz einzeln die gleiche Liste wieder von Anfang bis Ende durchgegangen wird usw.

Zu klärende Punkte:

1. Wie heißt der Datensatz?
Anm.: Neben dem Namen auch bereitstellende Organisation dokumentieren.
2. Handelt es sich bei dem Datensatz um administrative oder Befragungsdaten?
Anm.: Bei Befragungsdaten ist zudem relevant, ob die Erhebung einer Auskunftspflicht unterliegt bzw. ob diese für die komplette Erhebung oder nur für einige Merkmale gilt.
3. Wie ist die Grundgesamtheit definiert?
Anm.: Erwerbsfähige oder Erwerbstätige, alle Wirtschaftszweige oder Einschränkung auf bestimmte Wirtschaftszweige/Branchen
4. Handelt es sich um eine Stichprobe oder um eine Vollerhebung?
Anm.: Bei Stichproben ist auch noch das Stichprobendesign interessant, d.h. Stichprobengrundlage (bspw. IEB beim SIAB), Stichprobengröße absolut und in Relation zur Grundgesamtheit, Schichtungsmerkmale etc.
5. Wie hoch ist in etwa die Anzahl der Beobachtungseinheiten?
Anm.: Bei mehreren Wellen: Bandbreite. Bei Regionen: genaue Definition/Abgrenzung.
6. Liegen die Daten für alle Beobachtungseinheiten vor oder nur für ausgewählte?
Anm.: Das dürfte insb. bei Betriebsdaten häufiger der Fall sein (z.B. nur für Betriebe mit mehr als XX Beschäftigten).
7. Für welchen Zeitraum liegen die Daten vor? Gelten unterschiedliche Zeiträume für Teilpopulationen?
Anm.: Letzteres kann für bestimmte Teilgruppen der Beobachtungseinheiten z.B. Ost/West der Fall sein. Ggf. sind hier auch Interaktionen möglich, z.B. kleinere Betriebe in Westdeutschland o.ä.

8. Mit welcher Periodizität, d.h. zeitlichen Taktung liegen die Daten vor?
Anm.: Bei Jahresdaten interessieren auch die Erhebungszeiträume, d.h. von wann bis wann die Feldphase war.
9. Stehen bereits zukünftige Zeitpunkte fest oder sind geplant?
Anm.: Falls ja, dann ist auch wichtig, ob sich an obigen Angaben in Zukunft etwas ändern wird (z.B. erstmalige Hinzunahme bislang ausgeschlossener Beobachtungseinheiten).
10. Wie weit liegt der aktuelle Rand in der Vergangenheit?
Anm.: Dauer der Datenaufbereitung nach der Erhebung bis zur Verfügbarkeit.
11. Bei mehr als einer Welle: Handelt es sich um unverbundene Querschnitte, verbundene Querschnitte oder ein „echtes“ Panel?
Anm.: Falls es sich um ein „echtes“ Panel handelt, ist wichtig, ob dies im Rahmen der Standardbereitstellung der Daten auch nutzbar ist oder nur bei gesonderten Aufbereitungen.
12. Welche Themenbereiche und Merkmale werden erhoben?
Anm.: Ggf. vorhandene Fragebögen/Variablenlisten/Codebooks etc. zusenden lassen oder nach Downloadmöglichkeiten fragen.
13. In welcher Form sind die Wirtschaftszweige/Branchen in den Daten abgegrenzt?
Anm.: Ggf. auch Veränderungen im Zeitablauf erfassen.
14. Existieren Merkmale, die sich für eine Verknüpfung der Daten mit den BA-Prozessdaten eignen?
Anm.: Hier sind v.a. eindeutiger Identifier wie die SV-Nummer oder die BA-Betriebsnummer interessant. Als Alternative hierzu sind auch Informationen wie Name, Adresse, Geburtsdatum etc. relevant. Hier sollte differenziert werden nach dem Originaldatensatz und dem offiziell verfügbaren anonymisierten Datensatz.
15. Wie und in welcher Form stehen die Daten für Forschungszwecke zur Verfügung bzw. können sie verfügbar gemacht werden?
 - a. Wie erfolgt der Zugriff auf die Daten (SUF, DFV, GWP)?
 - b. Was wird für die Datenbeantragung benötigt (z.B. Datenantrag etc.)?
 - c. Liegen für die Zugangswege unterschiedliche Anforderungen an die Anonymisierung vor?
 - d. Wie lange dauert die Datenbereitstellung ab der Datenbeantragung?
 - e. Ist die Datenbeantragung mit Kosten verbunden?
16. Sind Ihnen weitere Datensätze bzw. -anbieter bekannt, die wir kontaktieren sollten?
Anm.: Nach Ansprechpartner fragen.

8.1.1.1 Relevante Branchen

1. Abfallwirtschaft
2. Bauhauptgewerbe
3. Dachdecker
4. Elektrohandwerk
5. Gebäudereiniger
6. Maler und Lackierer
7. Pflegebranche
8. Wäschereidienstleistungen im Objektkundengeschäft
9. Berufliche Aus- und Weiterbildung im SGB II oder SGB III
10. Sicherheitsdienstleistungen
11. Gerüstbauergewerbe
12. Steinmetze

8.1.1.2 Ergebnisindikatoren

- Individualebene:
 - Einkommen: Stundenlohn/Entgelte
 - Arbeitszeit: tariflich und/oder tatsächlich
 - Qualifikation/Tätigkeitsbereiche
 - Wirtschaftszweige/tarifvertragliche Regelungen
 - Qualität von Beschäftigungsverhältnissen (COPSOQ-/ERI-Ansatz)
 - Arbeitsmotivation/Produktivität
 - Arbeitsbeziehungen
 - Bildungsentscheidung junger Kohorten
 - Regionale und branchenbezogene Mobilität
- Betriebsebene
 - Umsatz
 - Beschäftigungsstruktur (Qualifikation, Alter, Anzahl Beschäftigter)
 - Marktein- und -austritte
 - Output und Investitionen
 - Preise
- Gesamtwirtschaft/Branche
 - Konsumausgaben/Ersparnis
 - Schattenwirtschaft
 - Steuereinnahmen (nach Steuerarten)
 - Sozialversicherungsbeiträge