

## Die IAB-Regionalstichprobe 1975-2001: IABS-R01

Hamann, Silke; Krug, Gerhard; Köhler, Markus; Ludwig-Mayerhofer, Wolfgang; Hacket, Anne

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

### Empfohlene Zitierung / Suggested Citation:

Hamann, S., Krug, G., Köhler, M., Ludwig-Mayerhofer, W., & Hacket, A. (2004). Die IAB-Regionalstichprobe 1975-2001: IABS-R01. *ZA-Information / Zentralarchiv für Empirische Sozialforschung*, 55, 34-59. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-198722>

### Nutzungsbedingungen:

*Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.*

*Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.*

### Terms of use:

*This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.*

*By using this particular document, you accept the above-stated conditions of use.*

## Die IAB-Regionalstichprobe 1975-2001: IABS-R01

von **Silke Hamann, Gerhard Krug, Markus Köhler, Wolfgang Ludwig-Mayerhofer und Anne Hacket**<sup>1</sup>

### *Zusammenfassung*

*Die Arbeit stellt die neue regionalisierte Stichprobe aus der Beschäftigten- und Leistungshistorik des Instituts für Arbeitsmarkt- und Berufsforschung vor. Ausgehend von der Datengrundlage und den Arbeiten der Datenaufbereitung und Anonymisierung werden vor allem die Neuerungen gegenüber früheren Regionalstichproben vorgestellt, so der größere Stichprobenumfang, die Möglichkeit, die Stichprobe zukünftig wesentlich schneller und einfacher zu aktualisieren, oder die Einbeziehung geringfügiger Beschäftigungsverhältnisse ab dem Jahr 1999. Auch ein Online-Informationsangebot für die Arbeit mit der Stichprobe steht zur Verfügung. Ein kurzes Beispiel verdeutlicht abschließend das Analysepotenzial der Daten.*

### *Abstract*

*The latest version of the IAB employment sample contains information on the course of employment and unemployment of more than 1 million individuals as well as information on the region where the respective employer is located. Together with basic information on the data from which the sample is drawn and details about data processing and anonymisation, we describe major innovations in the current sample, e.g. the larger sample size, better possibilities for future updates of the sample and the inclusion of “petty jobs” (“geringfügige Beschäftigung”) starting from 1999. The new release will be accompanied by an online information service. A short example demonstrates the potential for analysis offered by the data.*

---

<sup>1</sup> **Silke Hamann** (Dipl.-Sowi.), **Gerhard Krug** (Dipl.-Soz.) und **Markus Köhler** (Dipl.-Phys.) arbeiten im Institut für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit, Regensburger Straße 104, 90478 Nürnberg. Auskunft und Beratung zur IABS-R01 bietet das Forschungsdatenzentrum (FDZ) der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung, Regensburger Str. 104, 90478 Nürnberg, Telefon: 0911/179-1752, E-Mail: iab.fdz@iab.de (Ansprechpartnerin: Dr. **Christina Wübbeke**). Prof. Dr. **Wolfgang Ludwig-Mayerhofer** und **Anne Hacket** (Dipl.-Soz.) arbeiten am Fachbereich 1 der Universität Siegen, Adolf-Reichwein-Str. 2, 57068 Siegen, Telefon: 0271/740-4384, E-Mail: ludwig-mayerhofer@soziologie.uni-siegen.de.

## 1 Einleitung<sup>2</sup>

Neben dem Aspekt der zeitlichen Veränderlichkeit ist es vor allem der Aspekt der räumlichen Disparitäten, der vielen Gegenständen der empirischen Forschung stets neue Perspektiven und Fragestellungen liefert. Auch in der Arbeitsmarktforschung gelten Zeit und Raum längst als unverzichtbare Analysedimensionen (*Ludwig-Mayerhofer* 1990; *Blien, Hirschenauer* 1994). Grundvoraussetzung jeder empirischen Forschungsarbeit ist allerdings, dass eine Datenbasis zur Verfügung steht, die der bearbeiteten Fragestellung angemessen ist. Für die regionale Arbeitsmarktforschung bedeutet dies, dass neben Informationen auf individueller Ebene über Qualifikation, Erwerbsstatus, Einkommen, Sektorzugehörigkeit sowie Dauer von Beschäftigung und Arbeitslosigkeit auch eine Verknüpfung mit regionalen Kontextinformationen gegeben oder herstellbar sein muss. Da der Kontext auf verschiedenen Aggregationsebenen wirksam sein kann, bedarf es einer möglichst differenzierten und flexibel handhabbaren Gliederung des geographischen Raumes, dem eine Beobachtung entstammt. Mit der aktuellen IAB-Stichprobe regional 2001 (IABS-R01) des Instituts für Arbeitsmarkt und Berufsforschung (IAB) ist ein weiterer wichtiger Schritt in diese Richtung gelungen.<sup>3</sup> Gezogen aus der Beschäftigten- und Leistungsempfängerhistorik (BLH) des IAB, liefert sie genaue, zum Teil sogar tagesgenaue Angaben zum Erwerbsverlauf und Informationen zur Kreiszugehörigkeit des Beschäftigungsortes. Für eine große Vielzahl von Forschungsvorhaben liefert die Stichprobe eine geeignete Datenbasis mit Mikroinformationen. Angaben zu den Einkommen aus sozialversicherungspflichtiger Beschäftigung sind von hoher Genauigkeit und Reliabilität, da die Daten der Berechnung von Leistungsansprüchen dienen. Die Stichprobe ist zudem sehr groß und erlaubt dadurch, neben Vorhaben mit regionaler Fragestellung viele spezielle Themen aus einem weiten Bereich von Wirtschafts- und Sozialwissenschaften zu bearbeiten. Bereits mit Vorgängerversionen der aktuellen Stichprobe wurden viele wichtige Forschungsfragen untersucht (*Windzio* 2004a, 2004b, *Haas* 2003, *Büttner* und *Fitzenberger* 1998, *Baltagi* und *Blien* 1998, *Blien* und *Mederer* 1998, *Blien* 1995).

---

2 Das diesem Bericht zugrunde liegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 07RBS11 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autorinnen und Autoren.

3 Wir möchten uns bei den folgenden Personen herzlich bedanken: den Mitarbeitern des Servicebereichs IT und Informationsmanagement für die Bereitstellung der Datenbasis BLH und die Unterstützung nicht nur bei technischen Fragen, *Stefan Bender* und ganz besonders *Anette Haas* für die wertvolle Unterstützung durch ihr Fachwissen und ihre Erfahrungen aus vorherigen Stichprobenprojekten, *Annette Kohlmann* und *Christina Wübbeke* für die Beantwortung von Fragen zum FDZ, *Jasna Timpe* für die Unterstützung der Projektarbeiten, *Mark Josenhans* für die Programmierung der Datendeskriptionen und schließlich *Andreas Damelang* für die Unterstützung bei Anonymisierungsarbeiten und Datendeskriptionen.

Die IABS-R01 wurde im Rahmen eines durch das Bundesministerium für Bildung und Forschung (BMBF) finanzierten Verbundvorhabens „Aufbereitung der Beschäftigtenstatistik der Bundesagentur für Arbeit (früher: Bundesanstalt für Arbeit) als Datenquelle für problemorientierte Berichtssysteme“ gezogen und liegt nun als Scientific Use File vor. Das Projekt wurde in Zusammenarbeit zwischen **Wolfgang Ludwig-Mayerhofer** (zunächst an der Universität Leipzig, dann an der Universität Siegen) und dem IAB realisiert; die Projektleitung beim IAB lag in den Händen von **Uwe Blien**. Zur Sicherung der Anonymität bei gleichzeitiger Gewährleistung eines größtmöglichen Informationsgehalts waren aufwändige Datenarbeiten und Vorkehrungen erforderlich. Die dabei angewandten Regeln wurden weitgehend von früheren Stichproben übernommen. Allerdings gibt es in der aktuellen Version der IABS-R01 eine Reihe von Verbesserungen. Diese betreffen unter anderem:

- die Grundgesamtheit der Stichprobenziehung,
- die Erhöhung des Stichprobenumfangs von 1 auf 2 Prozent,
- die Entwicklung von Routinen zur Aktualisierung der Daten,
- der Wegfall der Längsschnittanonymisierung und
- die Aufnahme geringfügig Beschäftigter.

## 2 Datenzugang und Beratung

Auch hinsichtlich des Zugangs zur Stichprobe und bei der Bereitstellung von Beratungsleistungen haben sich Neuerungen ergeben, was vor allem mit dem neu gegründeten Forschungsdatenzentrum (FDZ) der BA im IAB<sup>4</sup> zusammenhängt.

Ab Anfang November 2004 wird die anonymisierte IABS-R01 vorerst vom FDZ, nach Regelung der Weitergabemodalitäten schließlich vom ZA unter der ZA-Studien-Nr. 4057 zur Verfügung gestellt.<sup>5</sup> Die Stichprobe wird dann auf Antrag an das ZA für konkrete und zeitlich befristete Projekte im Bereich der Arbeitsmarkt- und Berufsforschung weitergegeben. Dazu ist vom Antragsteller u.a. eine dezidierte Beschreibung des wissenschaftlichen Vorhabens mit Angabe des Verwendungszwecks, des Bezugs zur Arbeitsmarkt- und Berufsforschung, einer Liste der Projektmitarbeiter sowie der befristeten Dauer des Vorhabens zu geben. An Hand dieser Informationen entscheidet das FDZ darüber, ob die IABS-R01 einer wissen-

---

4 Zum Forschungsdatenzentrum (FDZ) siehe **Oertel** et al. (2004). Über aktuelle Serviceleistungen und Datenangebote des FDZ können sich Nutzerinnen und Nutzer auch im Internet unter <http://fdz.iab.de> informieren.

5 Forscher aus dem Ausland können die Beschäftigtenstichprobe über das FDZ erhalten.

schaftlichen Einrichtung übermittelt werden darf. Ist dies der Fall, schließt das ZA einen Nutzungsvertrag mit dem Datenempfänger.

Das FDZ bietet darüber hinaus Beratung zum Datenhandling, den Analysemöglichkeiten sowie der Reichweite und Gültigkeit der Daten an. Neben dem Informationsangebot im Internet vermittelt das FDZ in anwendungsorientierten Workshops und in Nutzerkonferenzen systematisch Informationen nicht nur über die hier vorgestellte Stichprobe. Für das Frühjahr 2005 ist eine Nutzerkonferenz geplant, auf der u.a. die Nutzerinnen und Nutzer der IAB-Beschäftigtenstichprobe Gelegenheit haben, ihre auf den Stichprobendaten beruhenden Forschungsarbeiten vorzustellen.

Obwohl es im Folgenden um die anonymisierte Version der IABS-R01 gehen wird, sei darauf hingewiesen, dass im Rahmen von Gastaufenthalten am neu entstandenen FDZ in Zukunft auch die Möglichkeit bestehen wird, mit einer schwach anonymisierten Version der Beschäftigtenstichprobe zu arbeiten. Diese enthält wie die IABS-R01 keine bzw. nur systemfreie Identifikatoren für Personen und Betriebe. Zusätzlich sind jedoch weitere Merkmale vorhanden und auch die anonymisierungsbedingten Aggregationen bei manchen Merkmalen (siehe unten) finden nicht statt.

### 3 Datengrundlage

Die IAB-Regionalstichprobe zeichnet für den Zeitraum von 1975–2001 die Erwerbs- und Leistungsempfangsverläufe von zwei Prozent aller sozialversicherungspflichtig Beschäftigten mit oft tagesgenauen und regional differenzierten Informationen nach, welche aus der sogenannten Beschäftigten- und Leistungsempfängerhistorik (BLH) stammen, die wiederum selbst aus zwei Datenquellen gespeist wird.

Die Hauptdatenquelle ist die *Beschäftigtenhistorik* des IAB. Ihre rechtliche Grundlage ist das mit Wirkung vom 01. Januar 1973 eingeführte und seit dem 01. Januar 1991 auf Ostdeutschland ausgedehnte integrierte Meldeverfahren zur Kranken-, Renten- und Arbeitslosenversicherung, das durch das Kürzel DEÜV (früher DEVO/DÜVO) bezeichnet ist (vgl. für weitere Details: **Bender** et al. 1996, S. 4 ff.; **Wermter** und **Cramer** 1988). Es verlangt von den Arbeitgebern Meldungen für alle sozialversicherungspflichtig beschäftigten Arbeitnehmer an die Sozialversicherungsträger. Die Beschäftigtenhistorik erfasst alle Arbeiter und Angestellten sowie alle Auszubildenden, soweit sie nicht von der Sozialversicherungspflicht befreit sind. Seit der Änderung des Meldeverfahrens zum 01. April 1999 werden auch geringfügig Beschäftigte und mithelfende Familienangehörige erfasst. Nicht in der Datenbasis enthalten sind u. a. Beamte, Selbständige und ordentliche Studierende (vgl. **Cramer** 1985). Jedes Jahr, in dem ein Beschäftigungsverhältnis für eine Per-

son existiert, ist durch mindestens eine Meldung abgebildet, welche die Information aus verschiedenen Merkmalen enthält. Die Daten werden von den Krankenkassen aufgenommen, von der Bundesagentur für Arbeit in einer laufenden Datei gesammelt und anschließend vom IAB in einer Historikdatei integriert.

Zu den Daten aus der Beschäftigtenhistorik wurden die entsprechenden Informationen aus der Leistungsempfängerhistorik des IAB hinzugespielt. Diese erfasst Zeiträume, in denen Personen Lohnersatzleistungen von der Bundesagentur für Arbeit beziehen. Die Leistungen beinhalten Arbeitslosengeld, Arbeitslosenhilfe und Unterhaltsgeld. Da der Leistungsanspruch von der Erfüllung gesetzlicher Voraussetzungen abhängt, werden Arbeitslosigkeitsperioden, in denen die Voraussetzungen nicht vorliegen (z.B. fehlende Bedürftigkeit im Falle der Arbeitslosenhilfe oder Nichterfüllung der Anwartschaftszeit beim Arbeitslosengeld) auch nicht berichtet.

Im Gegensatz zu Vorgängerstichproben wird die IABS-R01 nicht mehr direkt aus der Beschäftigtenhistorik gezogen und mit Daten der Leistungshistorik ergänzt, sondern aus der Beschäftigungs- und Leistungsempfängerhistorik (BLH) des IAB, in der die beiden Datenquellen bereits integriert und einige Datenbereinigungen durchgeführt worden sind. Die kontinuierliche Aufbereitung und Fortschreibung des Gesamtmaterials BLH diente vor allem auch dem Ziel, die Stichprobenziehung nicht wie bisher einmalig (oder in unregelmäßigen Abständen) durchzuführen, sondern Routinen der Datengewinnung für ein regelmäßiges Berichtssystem zu entwickeln. Diese Aktualisierbarkeit ist ein wesentlicher Fortschritt gegenüber früheren Versionen der IAB-Beschäftigtenstichprobe.

#### **4 Routinemäßige Aktualisierung der IABS-R01**

Die vorliegende IABS-R01 wurde Ende des Jahres 2003 aus dem bis dahin vorliegenden Datenmaterial (1975–2001) gezogen. Für den Zeitraum ab 2002 wird die Stichprobe dann jahresweise aktualisiert und fortgeschrieben werden. Dafür wurden verschiedene Stichprobenverfahren mit dem Ziel getestet, ein möglichst einfaches Datengenerierungsverfahren zu finden, das gleichwohl die Repräsentativität der Stichprobe gewährleistet. Anders als bei der Ziehung früherer Stichproben aus der Beschäftigtenhistorik (vgl. *Bender* et al. 1996) wurden dazu mehrere reine Zufallsverfahren getestet. Beibehalten wurden hingegen die Schichtungsmerkmale „Deutsche“/„Nicht-Deutsche“ und „Alte Bundesländer“/„Neue Bundesländer“.

Anhand der Sozialversicherungsnummer wurde zunächst eine 2-%-Personenstichprobe aus der BLH (Zeitraum 1975–2001) gezogen, wobei nur diejenigen Personen in die Auswahl kamen, die mindestens einen Beschäftigungsfall im Konto aufwie-

sen. Da dadurch Personen, die ausschließlich Episoden mit Leistungsbezügen aufwiesen, ausgeschlossen wurden, handelt es sich bei der IAB-R01 um eine *Beschäftigtenstichprobe*. Die Ziehung erfolgte getrennt nach den vier Schichten; aus jeder Schicht wurde eine 2-%-Zufallsstichprobe gezogen. Damit stellt die Stichprobe eine proportional geschichtete Zufallsstichprobe dar.

Das beschriebene Verfahren erleichtert die spätere Ergänzung und Aktualisierung der Beschäftigtenstichprobe, die jetzt zum ersten Mal vorgesehen ist. Bisher musste stets eine neue Stichprobe gezogen werden, wenn die Datenbasis verlängert werden sollte. Die Routinisierung und Erleichterung einer solchen zeitlichen Verlängerung ist eine wesentliche Innovation des neuen Projekts. Bei der jahresweisen Ergänzung der Stichprobe für den Zeitraum ab 2002 werden zum einen die Erwerbsverläufe der bereits in der Stichprobe enthaltenen Personen fortgeschrieben. Zum anderen werden aus den Erstmeldungen von sozialversicherungspflichtig Beschäftigten eines jeden Jahres nach 2001 wiederum 2-%-Zufallsstichproben gezogen und in die IAB-R01 integriert. Man erhält dadurch eine Stichprobe, deren (weiteres) Schichtungskriterium das Jahr ist. Für den Zeitraum 1975–2001 hat das Schichtungskriterium nur eine Ausprägung, danach treten jahresweise die weiteren Ausprägungen hinzu. Da proportional geschichtet wird, ist eine Gewichtung der einzelnen Schichten (Jahre) nicht nötig. Man erhält eine „Stichprobe mit Selbstgewichtung“ (*Cochran* 1972, S. 113).

Das Ziehungsverfahren gewährleistet damit Repräsentativität sowohl für Aussagen über den gesamten Zeitraum (Längsschnittrepräsentativität) als auch für Aussagen, die sich auf einzelne Jahre beziehen (Querschnittsrepräsentativität).<sup>6</sup> Da aus jedem zukünftigen Jahr der gleiche Prozentsatz an *neuen* sozialversicherungspflichtig Beschäftigten gezogen wird, besitzt jedes Element der Grundgesamtheit die gleiche Wahrscheinlichkeit, in die Stichprobe gezogen zu werden.

Um die Repräsentativität der gezogenen Stichprobe im Vergleich zur Grundgesamtheit, der Beschäftigten- und Leistungsempfängerhistorik (BLH), zu überprüfen, wurden Auszählungen zur Qualitätsprüfung vorgenommen. Ein Beispiel für diese Vergleiche befindet sich im Anhang (A1). Die Auswertungen belegen die Güte des gewählten Stichprobenverfahrens, da Abweichungen gering sind.

---

6 Lediglich hinsichtlich der Standardfehler sind bei inferenzstatistischen Aussagen geringe Abweichungen möglich. Für die Standardfehler könnte die Schicht als Cluster berücksichtigt werden, damit wäre die Effektivität der Schätzung für den Jahresvergleich größer. In diesem Fall bräuchten disproportionale Schichtungen sogar effektivere Stichproben. Angesichts des Stichprobenumfangs ist dieses Problem jedoch vollständig zu vernachlässigen.

## 5 Maßnahmen zur Aufbereitung der Daten

Trotz der Ziehung aus dem bereits aufbereiteten Datenmaterial der BLH war noch eine Reihe zusätzlicher Bereinigungs-schritte für die Regionalstichprobe durchzuführen. Diese betrafen

- die Bereinigung von Überschneidungsperioden, die auf Meldefehlern beruhen,
- die Schließung von Lücken im Erwerbsverlauf, die auf Meldefehlern beruhen,
- die Anonymisierung der Streikteilnehmer im Juni/Juli 1984 sowie
- den Umgang mit realen Überschneidungen zweier gleichzeitig möglicher Zustände.

Zur Korrektur zeitlicher Überlappungen einander ausschließender Meldungen<sup>7</sup> wurden folgende Routinen entwickelt: Bei Überschneidungen von regulären Beschäftigungen und Leistungsempfang wurden in der Regel die Beschäftigungsmeldungen beibehalten, da diese auf der genaueren Datengrundlage beruhen, und die Leistungsmeldung um den entsprechenden Zeitraum gekürzt. Bei gleichzeitigem Bezug mehrerer Lohnersatzleistungen wurde stets die mit der höchsten Priorität (Arbeitslosengeld vor Arbeitslosenhilfe vor Unterhaltsgeld) beibehalten. Bei Überschneidungen der Bezugszeiten wurde die zeitlich vorangehende Meldung gekürzt.

Lücken in Erwerbsverläufen, die auf fehlende Meldungen zurückgeführt werden konnten, wurden durch Ergänzungsmeldungen gefüllt. Wenn also z. B. im Kontenverlauf nach einer Jahresmeldung eine Meldelücke von genau einem Jahr auftrat, wurde eine Jahresmeldung ergänzt und die Lücke damit gefüllt. Bis 1999 wurde maximal um ein Jahr ergänzt, ab 1999 sind längere Zeiträume ergänzt worden. Dieses Vorgehen bereinigt nicht nur die Daten, sondern ermöglicht auch die Vergleichbarkeit mit der offiziellen Statistik, die bei ihrer Stichtagszählung ähnlich vorgeht.

Sowohl aus Gründen des Datenschutzes als auch aus inhaltlichen Gründen wurden die Beschäftigungsunterbrechungen bereinigt, die wegen der Aussperrung im Streik von Betrieben der Automobilindustrie in Baden-Württemberg und Hessen im Juni/Juli 1984 auftraten.

---

<sup>7</sup> An dieser Stelle ist eine begriffliche Klärung nötig. Im Folgenden sollen unter einem (Versicherten-)Konto alle zu einer Person vorliegende Meldungen verstanden werden. Als *Meldung* wiederum (synonym auch *Satz* oder *Episode*) werden die Informationen über Beschäftigungsperioden bezeichnet, so wie sie von den Sozialversicherungsträgern an die BA weitergeleitet werden. Davon ist die Bezeichnung *Spell* zu unterscheiden, die sich auf eine Zeile in der IABS-R01 bezieht, unabhängig davon, ob diese eine vollständige Meldung oder einen durch das Episodensplitting entstandenen Teil einer Meldung abbildet.



Da sich die erfassten Zustände in der IAB-Regionalstichprobe nicht immer gegenseitig ausschließen, sind auch zeitliche Überschneidungen von Beschäftigungsepisoden möglich (Mehrfachbeschäftigung). Ein Beispiel soll dies verdeutlichen. Eine Person nimmt während der Laufzeit ihrer regulären Hauptbeschäftigung eine geringfügige Beschäftigung auf. Nach einiger Zeit endet die Hauptbeschäftigung, die geringfügige Nebentätigkeit besteht weiter und wird nun zur Hauptbeschäftigung. Graphisch veranschaulicht dies die folgende Darstellung:



Damit auch diese Informationen für Analysen verwertbar sind, wurde für parallele Zustände ein Episodensplitting durchgeführt. Liegen zu einem Zeitraum mehrere Meldungen vor, werden diese so aufgeteilt, dass Anfangs- und Enddatum übereinstimmen. Im Ergebnis existieren einerseits Zeiträume (Spells), für die nur eine Meldung vorliegt. Im Beispiel ist dies zum einen der Zeitraum von der Aufnahme der Hauptbeschäftigung bis zur Aufnahme der geringfügigen Beschäftigung und zum anderen der Zeitraum vom Ende der Hauptbeschäftigung bis zum Ende der geringfügigen Beschäftigung. Dem stehen für die Zeiträume mit Überschneidungen Spells gegenüber, die alle dasselbe Beginn- und Enddatum besitzen.

Meldefolge nach Splitting:



Welchem der parallelen Spells Priorität zugewiesen wird, d.h. welche Beschäftigung z.B. als Haupt- und welche als Nebenbeschäftigung gilt, wird durch die Variable „level“ angezeigt. Die Zuweisung orientiert sich an folgenden Kriterien: Bei Mehrfachbeschäftigung werden sozialversicherungspflichtige Beschäftigungen gegenüber geringfügigen Beschäftigungen als Hauptbeschäftigung (level = 0) eingeordnet, ansonsten kann eine geringfügige Beschäftigung auch eine Hauptbeschäftigung sein. Innerhalb der beiden Beschäftigungsarten erfolgt die Entscheidung anhand des höheren Tagesentgeltes. Wie viele parallele Zustände eine Person im gleichen Zeitraum einnimmt, gibt das Merkmal „nlevel“ an. Zwischen Zeiträumen der Beschäftigung und des Leistungsbezugs können (mit wenigen, hier nicht relevanten

Ausnahmen) aus rechtlichen Gründen bis 1999 keine Überschneidungen bestehen. Seit 1999 kann neben dem Leistungsbezug allerdings auch eine geringfügige Beschäftigung bestehen, so dass auch hier Episodensplittings durchgeführt wurden. Mehrfachleistungsbezüge sind in der Beschäftigtenstichprobe im Gegensatz zu Mehrfachbeschäftigungen nicht möglich.

Im Datensatz ist nunmehr auch eine Kennung der ursprünglichen (systemfreien) Satznummer enthalten. Damit kann im Zweifelsfall genau nachvollzogen werden, aus welchen ursprünglichen Meldungen gesplittete Spells entstanden sind.

## 6 Die Merkmale in der IAB-Regionalstichprobe

Die IAB-Beschäftigtenstichprobe stellt eine zweiprozentige Zufallsauswahl aus allen seit 1975 in ein sozialversicherungspflichtiges Beschäftigungsverhältnis eingetretenen Personen dar. Für diese Personen ist eine Reihe von Informationen verfügbar, die für regionale Arbeitsmarkt- und Sozialstrukturanalysen von besonderer Bedeutung sind (eine genauere Beschreibung der Merkmale findet sich unter A3 im Anhang):

Personennummer	Stellung im Beruf
Geschlecht	Beruf (ausgeübte Tätigkeit)
Geburtsjahr	Betriebsnummernzähler
Staatsangehörigkeit (nur West)	Betriebsort (Kreisregion)
Beginn und Ende der Beschäftigung	Wirtschaftszweig
Personengruppe (z. B. geringfügige Beschäftigung)	sozialversicherungspflichtiges Bruttoentgelt (Tagesentgelt)
Rechtskreis Ost/West	Grund der Meldung
beschäftigt/arbeitslos	Beginn und Ende der Leistung
Rentenversicherungsträger	Leistungsart (ALG, ALHI, UHG)
Ausbildung	

Sowohl der zeitliche Aspekt als auch die regionale Differenziertheit der Daten sind besonders hervorzuheben. In zeitlicher Hinsicht bildet die IAB-Beschäftigtenstich-

probe die Erwerbsverläufe von sozialversicherungspflichtig Beschäftigten einschließlich Zeiten des Leistungsbezuges ab.

Aufgrund ihres Verwendungszweckes unterscheiden sich die erhobenen Merkmale allerdings in der zeitlichen Genauigkeit. Einige Merkmale wie der Krankenkassenwechsel, die Beschäftigungszeit und das versicherungspflichtige Entgelt dienen überwiegend versicherungsrechtlichen Zwecken. Sie sind bei jeder Veränderung meldepflichtig und weisen daher eine hohe Genauigkeit auf. Daneben enthält die Beschäftigtenhistorik auch rein statistische Angaben, deren Veränderungen nur zum Jahresende mit der Jahresmeldung angezeigt werden, z. B. der Wechsel des Berufes (*Cramer* 1985, für mögliche Konsequenzen in Analysen vgl. *Bender* 1997, *Bender, Haas, Klose* 1999).

In einer ersten Meldung wird das Datum des Beschäftigungsbeginns dokumentiert und, wenn die Beschäftigung nicht über den 31. 12. desselben Jahres hinaus andauert, das Datum der Beendigung. Besteht eine Beschäftigung dagegen über das Ende eines Jahres hinaus weiter, wird die Meldung am Jahresende gesplittet, d. h. es wird eine sog. Jahresmeldung mit dem Enddatum 31. 12. und eine nächste Meldung mit Anfangsdatum 1. 1. erzeugt. Ändert sich während der Beschäftigung ein versicherungsrelevanter Sachverhalt, wie die Betriebs- oder die Krankenkassenzugehörigkeit, dann wird dies ebenfalls durch eine Meldung dokumentiert.

Dem entsprechend wird auch der Beginn des Leistungsbezuges angegeben. Allerdings kommt es bei Leistungsbeziehern nicht zu der Splittung am Jahresende, so dass hier die berichteten Zeiträume ein Jahr überschreiten können. Auch für die Zeiten des Leistungsbezuges werden neben dem Ende des Bezuges noch andere Ereignisse als Meldung erfasst, so das Eintreten von Sperrzeiten oder der Wechsel der Leistungsart von Arbeitslosengeld auf Arbeitslosenhilfe.

Insgesamt sind in der IAB-Regionalstichprobe 21 041 596 Spells von 1 293 819 Personen enthalten; 18 501 706 Spells betreffen Beschäftigungszeiten, die übrigen 2 539 890 Zeiten des Leistungsbezuges.

Von zentraler Bedeutung für die Regionalforschung ist die Gliederung in regionale Einheiten. In der IAB-Regionalstichprobe wurde auf folgende Aspekte besonders Wert gelegt:

- möglichst kleinräumige Aufteilung,
- eine hohe Kontinuität im Zeitverlauf, sowie
- Anschlussfähigkeit an die amtliche Statistik für die Zuspiegelung von Kontextinformationen.

Die regionale Gliederung der Daten orientiert sich am Standort des Betriebes, in dem jemand beschäftigt ist. Dieser wird über die Betriebsnummer den sozialversicherungspflichtig Beschäftigten zugespielt, so dass alle in einem Betrieb arbeitenden Personen dem gleichen Ort zugeordnet sind. Entsprechendes gilt auch für den Wirtschaftszweig. Regionale Auswertungen erfolgen also nach dem Arbeitsort- und nicht nach dem Wohnortprinzip. Regionale Mobilität kann somit anhand regionaler Betriebswechsel im Sinne von Beschäftigungsmobilität untersucht werden. Ob diese Betriebswechsel auch mit einem Wohnortwechsel einhergehen, ist nicht erkennbar.

Die Fragestellungen der Regionalforschung können sich auf unterschiedliche Ebenen der Aggregation beziehen. Um diesbezüglich ein Höchstmaß an Flexibilität zu gewährleisten, liegen die Regionaleinheiten in möglichst kleinräumiger Aufteilung vor. Die Gemeindeebene scheidet als kleinste Einheit aus, da in diesem Fall die faktische Anonymität der Individualdaten nicht mehr sichergestellt werden kann. Die Kreise und kreisfreien Städte können jedoch mit wenigen Ausnahmen (siehe unten) als kleinste räumliche Einheit für ein regionales Scientific Use File verwendet werden (siehe weiter unten Abbildung 1).

Um bei der Analyse von Zeitverläufen keine Artefakte bezüglich der regionalen Mobilität zu erzeugen, musste wegen zahlreicher Gebietsstandsreformen vor allem in Ostdeutschland ein einheitlicher Gebietsstand gewählt werden. Im Datensatz geben die Kreiskennziffern den Gebietsstand von 2001 wieder. Ehemals gültige Kennziffern wurden auf aktuell gültige umgesetzt. Für den Osten wurden ca. 1 300 verschiedene ehemalige Gemeindekennziffern in den Gebietsstand von 2001 umkodiert. Die erzeugten Regionaleinheiten sind mit denen des früheren Regionalfiles zwar kompatibel, allerdings nicht völlig deckungsgleich. Anders als in Vorgängerstichproben ist die regionale Information nicht erst ab dem Meldedatum 1980, sondern für das gesamte Zeitfenster ab 1975 verfügbar.<sup>8</sup>

Die Kreisebene hat als administrative Einheit den Vorteil, dass sie für die Hinzuspielung von Kontextinformationen anschlussfähig ist, die für Kreise und kreisfreie Städte vorliegen. So können etwa die Beschäftigungsquote einer Region, Infrastrukturvariablen, Arbeitslosenquoten, Geburtenziffern, Altersstruktur, Preisniveau etc. leicht hinzugespielt werden. Dies erweitert den Analyserahmen beträchtlich, da auf diese Weise vielfältige inhaltliche Fragestellungen bearbeitbar sind.

---

<sup>8</sup> Aufgrund unzähliger Gemeindereformen in den siebziger Jahren waren die Umschlüsselungen auf den aktuellen Gebietsstand ungleich schwieriger, so dass Auswertungen auf regionaler Ebene für diesen Zeitraum gegebenenfalls weniger zuverlässig sind.

Bei Episoden des Leistungsbezuges ist weder die Angabe eines ausgeübten Berufes, noch die einer Stellung im Beruf vorhanden. Zudem kann auch keine Information zur Region und zum Wirtschaftszweig vorliegen, da beide Merkmale betriebsgebunden sind. Für die Zeiten des Leistungsbezuges wurden daher die Angaben zu den genannten Merkmalen jeweils aus der letzten vorangegangenen Episode der Erwerbstätigkeit übernommen.

## 7 Anonymisierung

Wie die Daten aus der Beschäftigtenstatistik und der Leistungsempfängerdatei unterliegt auch die daraus gezogene Stichprobe den Bestimmungen des Datenschutzes. Das IAB hat deshalb die Regionalstichprobe anonymisiert und sich dabei am Konzept der „faktischen Anonymisierung“ nach dem Bundes-Statistik-Gesetz und dem SGB X orientiert. Die faktische Anonymität der Daten ist dann gegeben, wenn ein Datenangreifer unverhältnismäßig viel Zeit, Kosten und Arbeitskraft aufbringen muss, um einen Datensatz zu deanonymisieren (§ 67 Abs. 8 SGB X). Die Operationalisierung der faktischen Anonymisierung erfolgte in Anlehnung an Tests und Verfahrensvorschläge, die von *Müller* et al. (1991) für den Mikrozensus entwickelt wurden. Die Vorschläge wurden zwischen den Datenschutzbeauftragten, den statistischen Ämtern und den Vertretern der Sozial- und Wirtschaftswissenschaften so abgestimmt, dass sie bei hohem Anonymisierungsschutz ein Maximum an wissenschaftlicher Auswertbarkeit garantieren.

In der aktuellen Stichprobe wurde – im Unterschied zu früheren Stichproben – neben der Anonymisierung der personen- und betriebsbezogenen Querschnittsinformationen keine Anonymisierung der Längsschnittinformation vorgenommen, die in einer Verschiebung des gesamten Erwerbsverlaufs jeder Person um einen zufälligen Faktor auf der Zeitachse bestanden hätte. Durch eine solche Längsschnittanonymisierung wurden Querschnittsbetrachtungen wie Auszählungen zum Stichtag verzerrt. Da die Erfahrungen mit bisherigen Stichproben gezeigt haben, dass die vorgenommenen Anonymisierungsmaßnahmen im Querschnitt ausreichenden Schutz vor Datenangreifern bieten, konnte auf die Längsschnittanonymisierung verzichtet werden, ohne damit den Datenschutz zu vernachlässigen. Die Verfahrensweise wurde mit dem in der BA für den Datenschutz zuständigen Referat abgestimmt.

In der Tabelle A2 im Anhang sind die Anonymisierungsmaßnahmen der IABS-R01 im Vergleich zur unmittelbaren Vorgängerstichprobe für die einzelnen Merkmale aufgeführt. Bei einigen Merkmalen waren keinerlei Anonymisierungsmaßnahmen nötig, während bei einer Reihe von Merkmalen aufgrund zu niedriger Fallzahlen einige Ausprägungen zusammengefasst wurden (z. B. Grund der Meldung). Dabei

wurden auch unterschiedlich starke Aggregationen für Ost- und Westdeutschland durchgeführt, um möglichst viele der Originalinformationen zu erhalten (z. B. Personengruppe und Staatsangehörigkeit).

Besondere Bedeutung kommt bei der Anonymisierung in der IAB-Regionalstichprobe der Frage der kleinsten identifizierbaren Raumeinheiten und deren Mindestgröße zu. Als Anforderungen an die Raumeinheiten definieren *Müller* et al. (1991), dass sie mindestens 100 000 Einwohner aufweisen müssen. Um die Anonymisierungsanforderung der Mindestgröße zu erfüllen, wurden entsprechend der Angaben des statistischen Bundesamtes über die Einwohnerzahlen zum 31. 12. 2001 kleinere Kreise zusammengefasst. Im Vergleich zur unmittelbaren Vorgängerstichprobe reg\_75–97 mussten vor allem wegen der Bevölkerungsabwanderungen in Ostdeutschland mehr Kreise aggregiert werden. Die Zusammenlegung benachbarter Kreise hatte dabei Vorrang vor der Forderung nach gleichem Kreistyp (vgl. Abbildung 1).

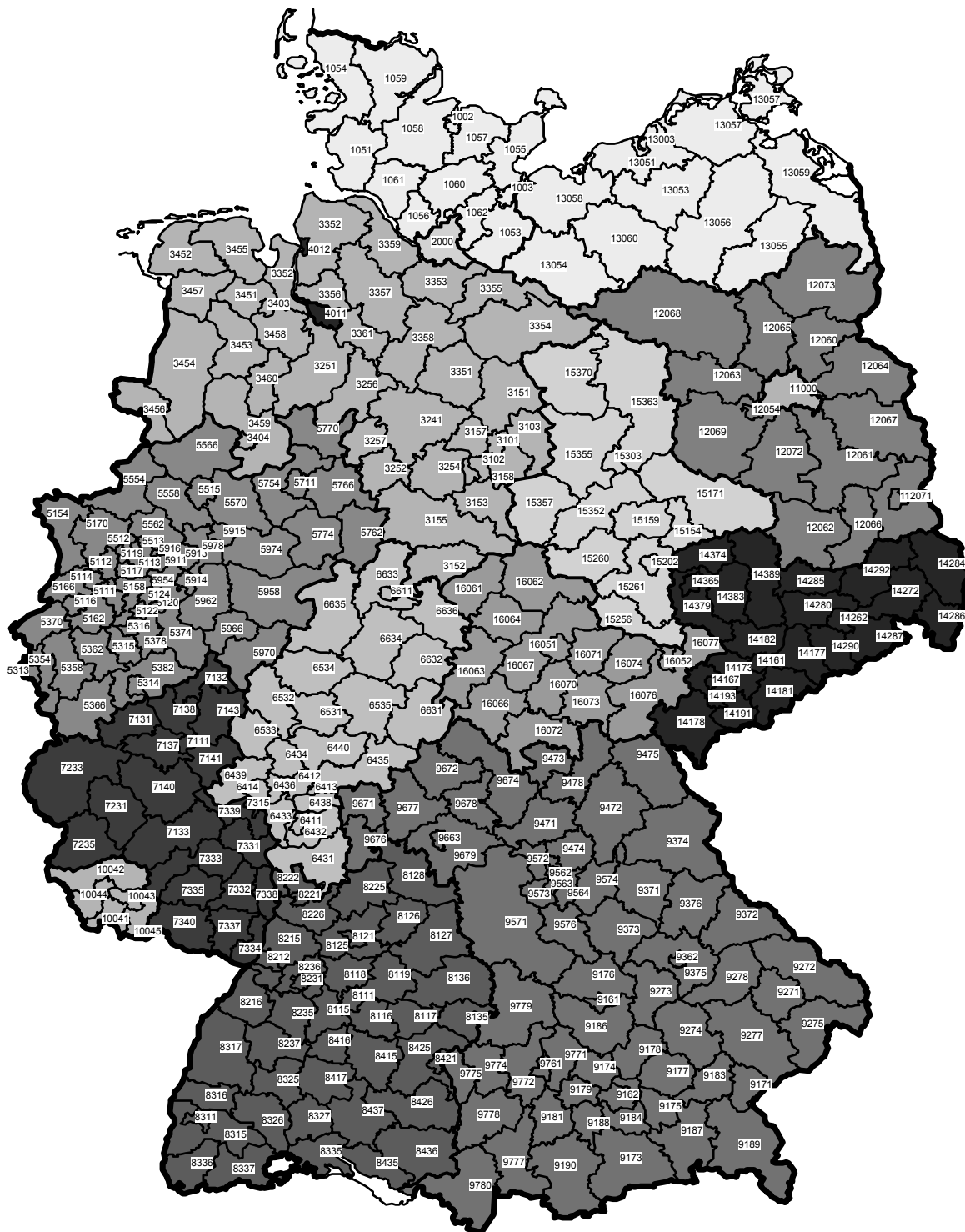
Die Aggregation betraf zum Beispiel Schwerin, Jerichower Land, den Saale-Orla-Kreis, aber auch Kaiserslautern oder Helmstedt.<sup>9</sup> Nicht mehr aggregiert werden mussten Saalfeld und Rudolstadt. Insgesamt konnten 343 (vorher 348) Regionen ausgewiesen werden, 266 in den alten und 76 in den neuen Bundesländern und Berlin.

Die Merkmale Beruf und der Wirtschaftszweig können vor allem durch die Erhöhung des Stichprobenumfangs auf zwei Prozent weniger stark zusammengefasst weitergegeben werden. Statt der bisher enthaltenen 117 enthält die Stichprobe nunmehr 130 Berufsaggregate. Beim Merkmal Wirtschaftszweig konnten statt vorher 13 Klassen nun 16 Klassen abgebildet werden. Differenziertere Analysen sind damit innerhalb der Investitionsgüterproduktion, des Handels und der gesellschaftsbezogenen Dienstleistungen möglich.

---

9 Unabhängig von der anonymisierungsbedingten Aggregation wurden im Rahmen einer Gebietsstandsänderung auch Hannover Stadt und Land zusammengelegt.

Abbildung 1 In der Stichprobe ausgewiesene Kreise und Kreisaggregate



## 8 Online-Infoangebot

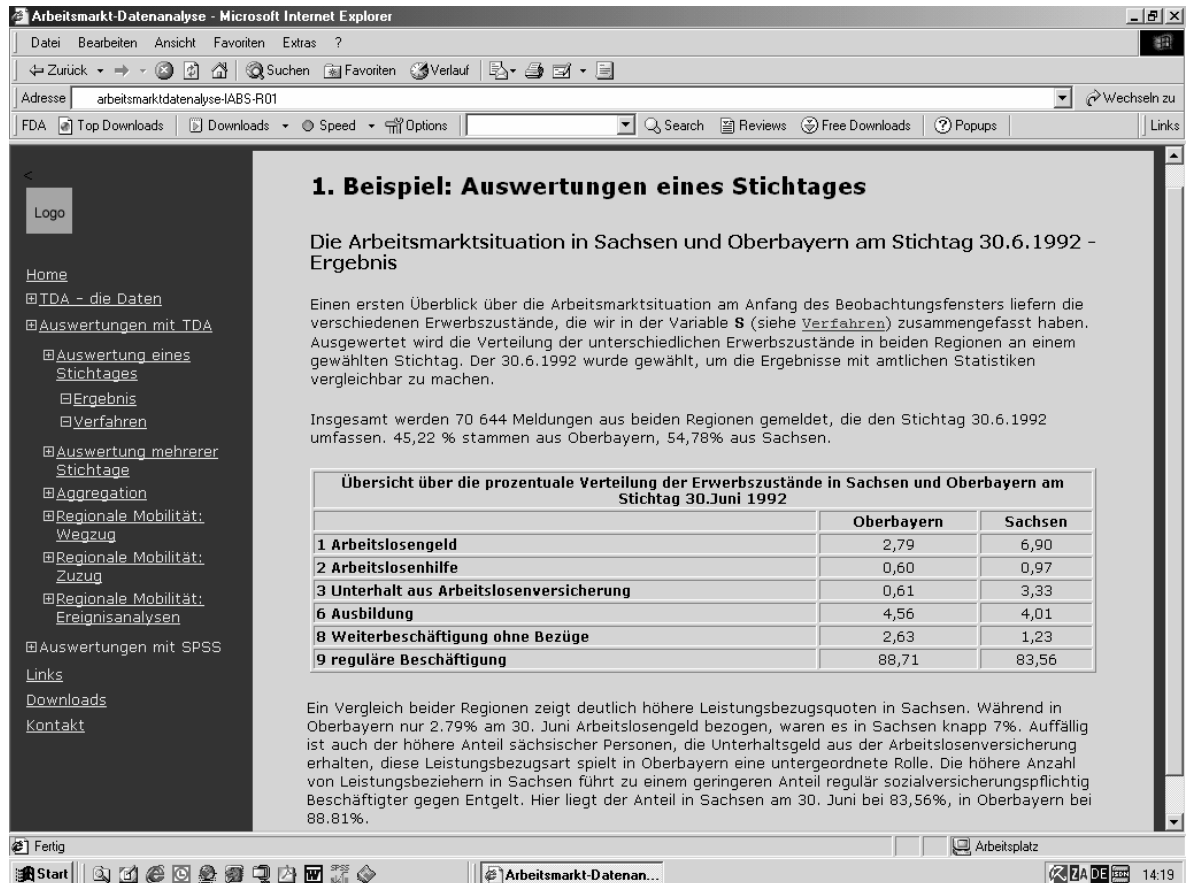
Als zentrale Informationsschnittstelle für die Nutzer der IABS-R01 wird aktuell eine Website eingerichtet. Mit dieser sollen die Informationsangebote des Zentralarchivs nicht ersetzt, aber (umfangreich) ergänzt werden. Das Onlineangebot hat im Vergleich zur bisherigen Praxis, die Beschäftigtenstichprobe mit einer Buchpublikation zu dokumentieren, einige Vorteile. Die Informationen sind von jedem Computerarbeitsplatz schnell zugänglich und können stets auf den neusten Stand gebracht werden. Zudem können die beispielhaften Auswertungen aufgrund der mitgelieferten SPSS- bzw. TDA-Programme leichter als Basis eigener Analysen herangezogen werden. Die Website zur IABS-R01 soll u. a. folgende Informationen bereitstellen:

- Die genaue Struktur des Datensatzes (Variablen und ihre Ausprägungen),
- Grundzüge der Generierung der Stichprobe und ihre Konsequenzen für mögliche Zielsetzungen und Reichweite empirischer Analysen,
- Downloadbare Syntax-Files, die (gegebenenfalls mit geringfügigen Modifikationen) unmittelbar von Nutzern für eigene Datenaufbereitungen und -analysen eingesetzt werden können,
- Grundauszählungen und einfache Kreuztabellierungen von Merkmalen, und
- nachvollziehbare Darstellung komplexerer Analysen und ihrer Ergebnisse.

Die folgenden Screenshots verdeutlichen den grundsätzlichen Aufbau der Website: Datenauswertungen werden anhand unterschiedlicher Beispiele erläutert. Bei jedem Beispiel können die Nutzer/-innen anhand der Seite „Ergebnis“ schnell erkennen, ob das angeführte Beispiel für sie überhaupt von Interesse ist (im hier vorgestellten Beispiel – siehe Abbildung 2 – die Auswertung zu einem Stichtag). Ist das der Fall, so können sie auf der Seite „Verfahren“ detaillierte Informationen über das Vorgehen bei der Analyse nachvollziehen (Abbildung 3).

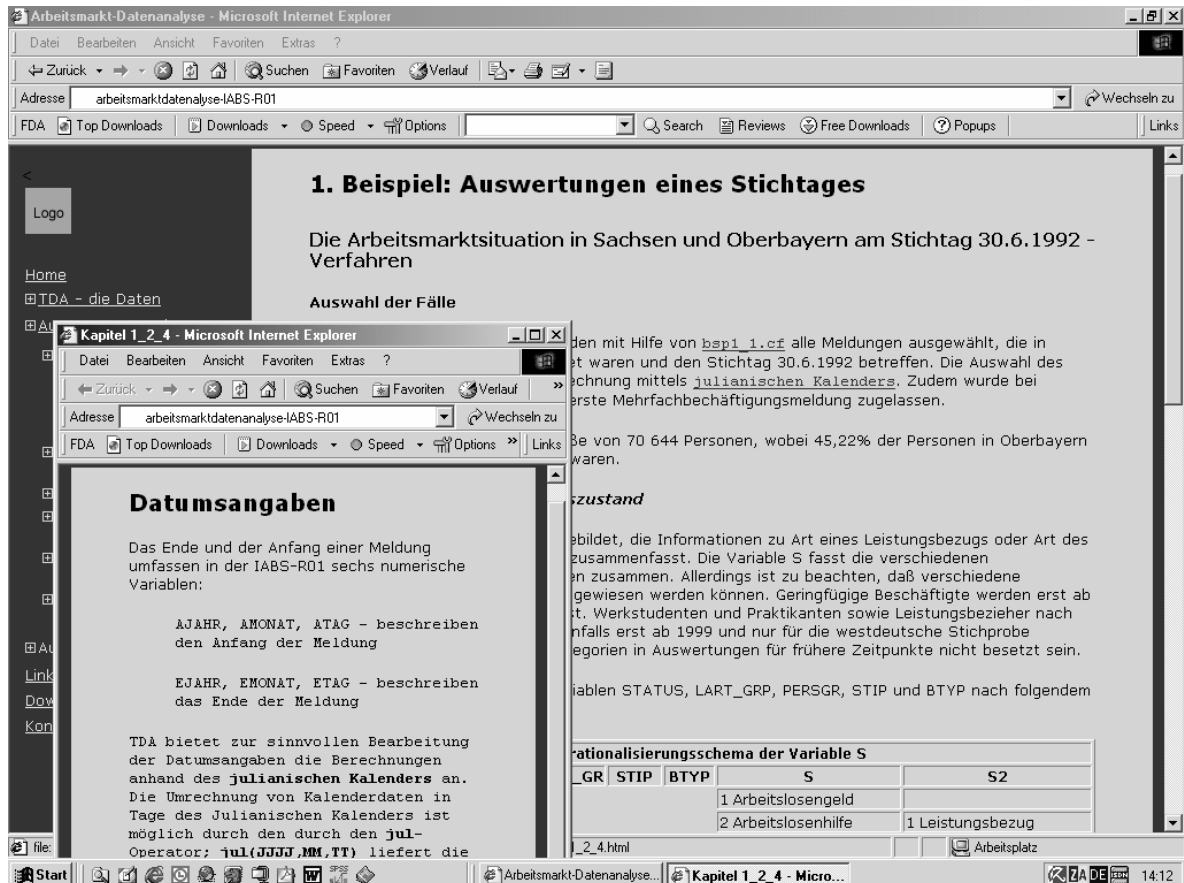


**Abbildung 2** Screenshot der Website: Ergebnis einer exemplarischen Auswertung



Die URL der Website steht derzeit noch nicht fest, sie wird aber auf jeden Fall über <http://www.fb1.uni-siegen.de/soziolog/personen/ludwigmayerhofer/rbs11d.html>, die Website des Projekts an der Universität Siegen, und über <http://fdz.iab.de>, die Homepage des Forschungsdatenzentrums, zugänglich sein.

**Abbildung 3** Screenshot der Website: Erläuterung des Verfahrens einer konkreten Auswertung



## 9 Grenzen und Besonderheiten der Daten

Trotz des einzigartigen Informationsgehalts der Stichprobe ist zu berücksichtigen, dass ihre Aussagekraft (wie die eines jeden Datensatzes) beschränkt ist. Wie schon oben herausgestellt, sind nur sozialversicherungspflichtige Beschäftigungsverhältnisse erfasst; über Beamte oder Selbstständige können damit keine Aussagen gemacht werden. Auch ist bei der Analyse der Einkommensangaben zu beachten, dass diese unterhalb der Geringfügigkeitsgrenze (bis 1998) und oberhalb der Beitragsbemessungsgrenze abgeschnitten sind; damit werden insbesondere die Einkommen hochqualifizierter Beschäftigter unterschätzt. Ferner darf bei der Arbeit mit den Einkommensangaben die Umstellung von DM auf Euro im Jahr 1999 (!) nicht übersehen werden. Bei Längsschnittanalysen sollte man nicht vergessen, dass die Erwerbsverläufe, die im Jahr 1975 (in Ostdeutschland: im Jahr 1992) beginnen, linkszensiert, und jene, die im Jahr 2001 enden, rechtszensiert sind.

Es ist ferner bekannt (und wird auch in den Daten sichtbar), dass nicht alle Merkmale mit der gleich hohen Validität und Genauigkeit vorliegen. Da beispielsweise die Angaben zur schulischen und beruflichen Bildung für den Erwerb sozialversicherungsrechtlicher Ansprüche keine Bedeutung haben, ist anzunehmen, dass sie von den Betrieben weniger genau gehandhabt werden. Auf jeden Fall finden sich im Datensatz Personen, die in nicht nachvollziehbarer Weise ihren Bildungsstatus im Zeitverlauf ändern. Gelegentlich gewinnt man den Eindruck, dass weniger die Qualifikation der Beschäftigten als die der Tätigkeit angegeben worden sein könnte. Doch ist auch die Größenordnung der Fehler nicht exakt quantifizierbar, so haben Analysen mit früheren Versionen der Daten keine Hinweise erbracht, dass unplausible Ergebnisse erzielt werden. Beispielsweise lassen sich die bekannten Einkommensunterschiede zwischen Bildungsgruppen sehr deutlich nachweisen. Das spricht dafür, dass Nutzer sich nicht von entsprechenden Analysen abhalten lassen sollten. In der neuen Version der Stichprobe sind jedoch im Unterschied zu früheren Versionen die Merkmale „Familienstand“ und „Kinderzahl“ nicht enthalten. Bei diesen Merkmalen hat sich gezeigt, dass die Validität zu gering ist. Seit 1999 entfallen sie zudem völlig, da sie dann auch nicht mehr von den Krankenkassen gemeldet werden.

Auch bei den Einkommensangaben ist mit minimalen Fehlern zu rechnen. In Einzelfällen liegen die gemeldeten Einkommen bei geringfügiger Beschäftigung über bzw. bei regulären Beschäftigungen unterhalb der Geringfügigkeitsgrenze. Das Merkmal Personengruppe (also z.B. die Angabe einer geringfügigen Beschäftigung) gilt als ‚hartes‘ Kriterium, so dass die Einkommensangaben in den genannten Konstellationen als problematisch erscheinen müssen. Die Probleme liegen jedoch insgesamt gesehen eher im Promillebereich.

Von ähnlich geringer Größenordnung ist ein weiteres Problem: Gelegentlich liegen Doppelmeldungen vor; dies kann z.B. vorkommen, wenn eine Korrektur einer früheren Meldung nicht als solche ausgewiesen oder ein Datensatz versehentlich doppelt gemeldet wurde. Solche Doppelmeldungen wurden bei der Datensatzbereinigung entfernt, wenn sämtliche Angaben zum Beschäftigungsverhältnis exakt übereinstimmten. Schon bei geringen Unterschieden zwischen den Angaben – etwa zum Entgelt – wurden jedoch beide Meldungen in den Datensatz aufgenommen. Diese Abweichungen sind z. T. im Datensatz nicht mehr zu erkennen, weil das Tagesentgelt auf ganze Zahlen gerundet wurde, was bei geringen Unterschieden im Jahresentgelt zu identischen Werten des Tagesentgeltes führen kann. Mit Hilfe von Plausibilitätsprüfungen lässt sich allerdings häufig feststellen, ob es sich um Datenfehler handeln könnte. (Ein typischer Fall: Eine Person weist einen stabilen Erwerbsverlauf bei einem Arbeitgeber mit allmählichen Gehaltssteigerungen auf. Für ein Jahr liegen zwei Meldungen vor, von denen jede einzelne ein Entgelt aufweist, das in die

bisherige Gehaltsentwicklung passt. Würde man diese aggregieren in der Annahme, dass es sich um zwei Beschäftigungsverhältnisse handelt, so ergäbe sich für das betreffende Jahr ein Entgelt, welches das Doppelte des erwartbaren Entgelts beträgt.)

## 10 Anwendungsbeispiel: Mobilität zwischen Ost- und Westdeutschland

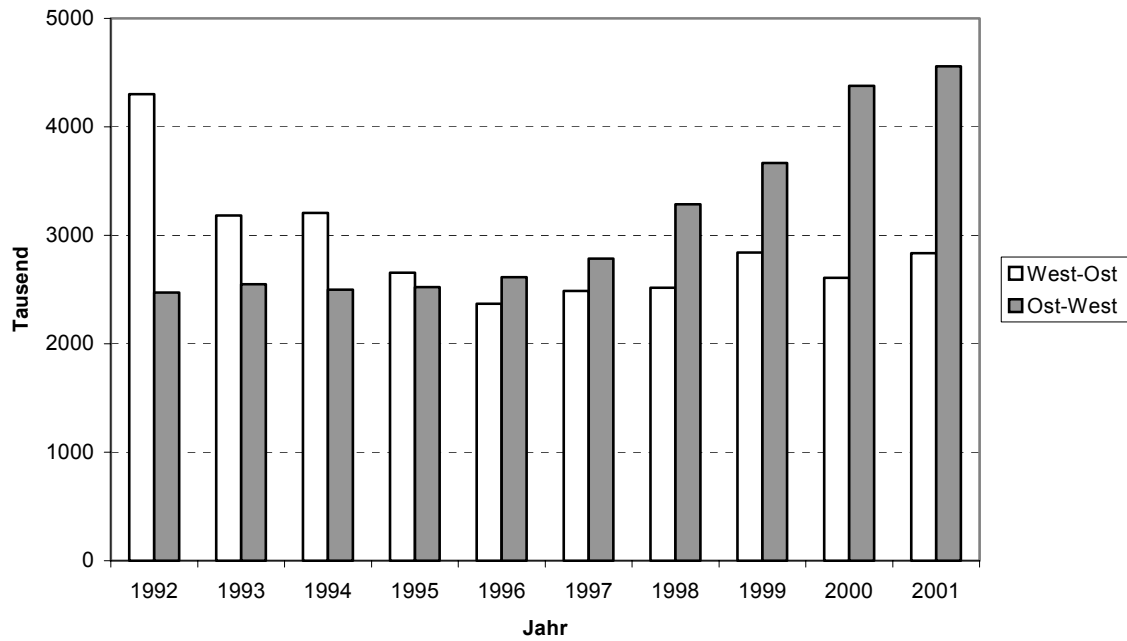
Durch den großen Stichprobenumfang liegen auch für fein differenzierte Ereignisse große Fallzahlen vor, die aussagekräftige Analysen ermöglichen. Zu verdeutlichen ist dies an einem einfachen Beispiel, nämlich Wechseln zwischen Ost- und Westdeutschland bzw. umgekehrt. Wie die nachfolgenden Graphiken zeigen, liegen für jedes Kalenderjahr einige Tausend solcher Wechsel vor (mögliche mehrfache Wechsel ein und derselben Person wurden jeweils neu mitgezählt).<sup>10</sup> Es lässt sich anhand dieser Daten gut zeigen, wie die kurz nach der Wende noch zahlreichen Wechsel westdeutscher Arbeitnehmer nach Ostdeutschland rasch abklingen, während die Mobilität in umgekehrter Richtung in absoluten Zahlen im Zeitverlauf zunimmt und deutlich über der West-Ost-Mobilität liegt (obwohl die Zahl der Erwerbspersonen in Ostdeutschland wesentlich geringer als in Westdeutschland ist) (Abbildung 4). Von Interesse ist es auch, die absoluten Zahlen auf die Zahl aller Wechsel zwischen Bundesländern zu beziehen (Abbildung 5).<sup>11</sup> Daran wird deutlich, dass Mobilität von West- nach Ostdeutschland nur einen kleinen – und abnehmenden – Anteil der regionalen Mobilität (ursprünglich) westdeutscher Beschäftigter ausmacht, während die Mobilität von Ost- nach Westdeutschland unter jenen regionalen Wechseln, die in Ostdeutschland ihren Ausgang nehmen, einen deutlich höheren und in den jüngsten Jahren steigenden Anteil aufweist.

---

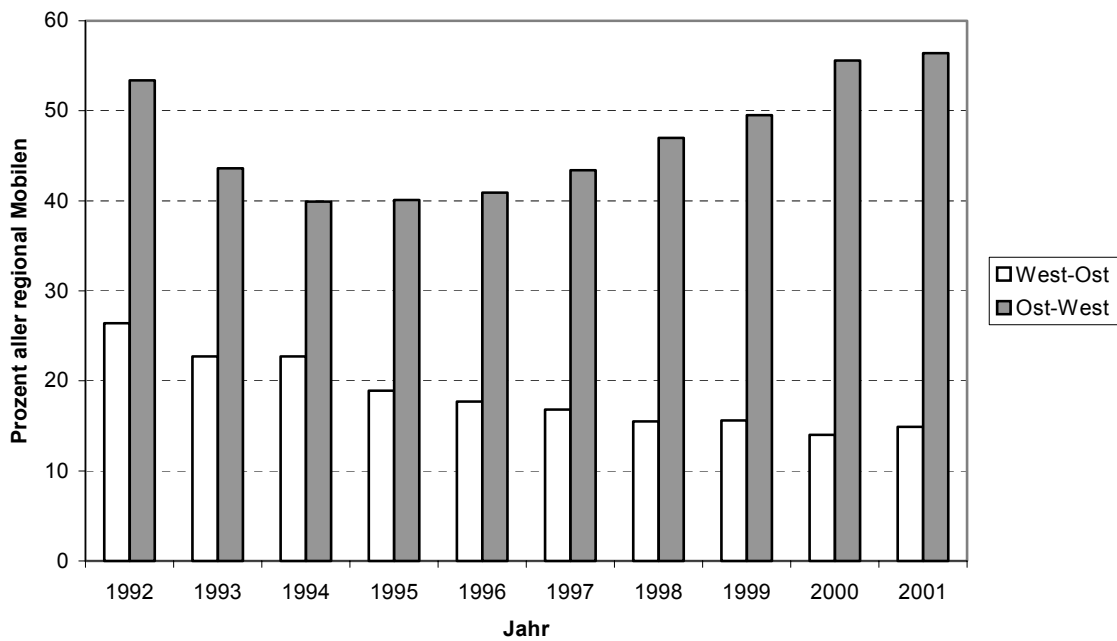
10 In die hier berichteten Angaben wurde gegebenenfalls nur das erste von mehreren gleichzeitig bestehenden Beschäftigungsverhältnissen einer Person einbezogen. Mobilität von West- nach Ost-Berlin wurde nicht als „Ost-/West-Mobilität“ gewertet (das gleiche gilt im Falle der umgekehrten Richtung).

11 Es ist klar, dass nicht jeder Wechsel von einem Bundesland in ein anderes mit regionaler Mobilität gleichzusetzen ist. Eine Person, die nahe an der Grenze zu einem anderen Bundesland wohnt und dort eine Erwerbstätigkeit aufnimmt, kann wesentlich immobil sein als eine Person, die innerhalb eines großen Flächenstaats den Arbeitgeber wechselt. Die hier vorgelegten Zahlen dienen nur einer einfachen Veranschaulichung möglicher Analysen, die noch weiterer Vertiefung bedürfen.

**Abbildung 4** Mobilität von West- nach Ostdeutschland und umgekehrt (absolute Zahlen)



**Abbildung 5** Mobilität von West- nach Ostdeutschland und umgekehrt (in Prozent aller Wechsel von einem Bundesland in ein anderes, die von Westdeutschland bzw. von Ostdeutschland aus vorgenommen werden)



## Literatur

- Baltagi, Badi H.** und **Uwe Blien** 1998: The German Wage Curve: Evidence from the IAB Employment Sample. In: *Economics Letters* 61: 135-142.
- Bender, Stefan** 1997: Die IAB-Beschäftigtenstichprobe 1975-1990: Analysemöglichkeiten der anonymisierten Stichprobe. In: **Hujer, R., Rendtel, U., Wagner, G.:** Wirtschafts- und sozialwissenschaftliche Panel-Studien – Datenstrukturen und Analyseverfahren. Sonderhefte zum Allgemeinen Statistischen Archiv 30: 103-126.
- Bender, Stefan, Anette Haas** und **Christoph Klose** 1999: Mobilität allein kann Arbeitsmarktprobleme nicht lösen. Die Entwicklung der beruflichen und betrieblichen Mobilität von 1985 bis 1995. IAB-Kurzbericht, Nr. 2/1999.
- Bender, Stefan, Jürgen Hilzendegen, Götz Rohwer** und **Helmut Rudolph** 1996: Die IAB-Beschäftigtenstichprobe 1975-1990. Beiträge zur Arbeitsmarkt- und Berufsforschung 197.
- Blien, Uwe** 1995: Die Lohnkurve von 1989. Eine Mehrebenenanalyse zum Zusammenhang von regionalen Durchschnittslöhnen und der regionalen Arbeitslosenquote. Mitteilungen aus der Arbeitsmarkt- und Berufsforschung 28/2, 155-170.
- Blien, Uwe** und **Franziska Hirschenauer** 1994: Die Entwicklung regionaler Disparitäten in Ostdeutschland. In: Mitteilungen aus der Arbeitsmarkt- und Berufsforschung 27: 323-337.
- Blien, Uwe** und **Anita Mederer** 1998: Die Regionaldimension geschlechtsspezifischer Entlohnung. In: Jahrbuch für Regionalwissenschaft 18: 37-54.
- Büttner, Thiess** und **Bernd Fitzenberger** 1998: Central Wage Bargaining and Local Wage Flexibility: Evidence from the Entire Wage Distribution. In: ZEW-Discussion Paper 98-39.
- Cramer, Ulrich** 1985: Probleme der Genauigkeit der Beschäftigtenstatistik. In: Allgemeines Statistisches Archiv 69: 56-68.
- Cochran, William G.** 1972: Stichprobenverfahren. Berlin/New York.
- Möller, Joachim** und **Anette Haas** 2003: The Agglomeration Wage Differential Reconsidered: An Investigation Using German Micro Data 1984-1997. In: **Johannes Bröcker** et al. (Hrsg.): Innovation Clusters and Interregional Competition. Heidelberg. S. 182-217.
- Ludwig-Mayerhofer, Wolfgang** 1990: Arbeitslosigkeit und Erwerbsverlauf. In: Zeitschrift für Soziologie 19: 345-359.
- Müller, Walter, Uwe Blien, Peter Knoche, Heike Wirth,** et al. 1991: Die faktische Anonymität von Mikrodaten. Forum der Bundesstatistik Nr. 19.
- Oertel, Martina, Jürgen Passenberger** und **Markus Janser** 2004: Datenservice: Forschungsdatenzentrum der BA geht an den Start. Externe Forscher werden bald besseren Zugang zu den Daten der Bundesagentur für Arbeit haben. Für mehr Wettbewerb in der Arbeitsmarktforschung. IAB-Kurzbericht Nr. 09/2004.
- Wermter, Winfried** und **Ulrich Cramer** 1988: Wie hoch war der Beschäftigtenanstieg seit 1983 – Ein Diskussionsbeitrag aus der Sicht der Beschäftigtenstatistik der Bundesanstalt für Arbeit. In: Mitteilungen der Arbeitsmarkt- und Berufsforschung 4: 468-482.
- Windzio, Michael** 2004a: Zwischen Nord- und Süddeutschland: Die Überwindung räumlicher Distanzen bei der Arbeitsmarktmobilität. In: Zeitschrift für Arbeitsmarktforschung 1: 29-44.
- Windzio, Michael** 2004b: Kann der regionale Kontext zur „Arbeitslosenfalle“ werden? In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 56: 257-278.



**A2: Übersicht der Anonymisierungsmaßnahmen in der neuen Regionalstichprobe**

Merkmale	Notwendige Anonymisierungsmaßnahmen im Vergleich zur Vorgängerstichprobe	
	IABS-REG 1975-1997	IABS-R01
Versicherungs-Nr.	Ersetzung durch systemfreie Personennummer	
Rentenvers.-träger	im Original vorhanden	
Geschlecht	im Original vorhanden	
Familienstand	im Original vorhanden	<b>entfällt, da nicht valide</b>
Ausbildung	im Original vorhanden	
Stellung im Beruf	im Original vorhanden	
Geburtsjahr	Wird für das gesamte Konto einer Person nicht ausgewiesen, falls zu irgendeinem Zeitpunkt die Zahl der Personen des selben Alters so gering ist, dass die Anonymität nicht gewährleistet ist	
Beginn und Ende der Beschäftigung	Längsschnittverschiebung des gesamten Erwerbsverlaufs einer Person um eine Zufallskonstante	<b>im Original vorhanden</b>
Grund der Meldung	Aggregation von Merkmalsausprägungen: „Abmeldung“ und „Tod“; „einmalig gezahltes Entgelt“; „Insolvenz des Arbeitgebers“ und „sonstige Gründe“	<i>vor 1999:</i> Aggregation der Abgabegründe ,2' bis ,5' zu ,1' = Ende der Beschäftigung. <i>nach 1999:</i> getrennte Ausweisung, ,2' = Abmeldung nach Krankenkassenwechsel, ,3' = Abmeldung nach Beitragsgruppenwechsel, ,4' = Abmeldung nach Unterbrechung größer ein Monat, ,5' = Abmeldung wegen Wechsel Entgeltabrechnungssystem und ,6' = gleichzeitige An- und Abmeldung.
Beruf (ausgeübte Tätigkeit)	Die insgesamt 335 Berufe werden zu <b>117</b> Berufsaggregaten zusammengefasst	Die insgesamt 335 Berufe werden zu <b>130</b> Berufsaggregaten zusammengefasst
Sozialversicherungspflichtiges Bruttoentgelt	Umrechnung auf durchschnittliches Bruttotagesentgelt und Rundung auf DM-Beträge, ab 1999 <b>Euro-Beträge</b> . Ausweisung bis zur jeweiligen Beitragsbemessungsgrenze der Rentenversicherung für Arbeiter und Angestellte	
Beginn und Ende der Leistung	Längsschnittverschiebung des gesamten Erwerbsverlaufs einer Person um eine Zufallskonstante	<b>im Original vorhanden</b>
Leistungsart	Arbeitslosengeld, Arbeitslosenhilfe und Unterhaltsgeld	
Staatsangehörigkeit	Ausweisung einer Nationalitätendummy „deutsch/ausländisch“ nur für die 270 Kreisregionen der alten Bundesländer	Ausweisung einer Nationalitätendummy „deutsch/ausländisch“ nur für die 266 Kreisregionen der alten Bundesländer
Kinderzahl	Bis zu sieben Kindern wird die Kinderzahl im Original ausgewiesen; ab acht Kindern wird in die Kategorie „acht und mehr Kinder“ zusammengefasst	<b>entfällt, da nicht valide</b>



Merkmale	Notwendige Anonymisierungsmaßnahmen im Vergleich zur Vorgängerstichprobe	
Personengruppe (Merkmal aus dem neuen Meldeverfahren, wird auch für zurückliegende Zeiträume konstruiert)	<b>neu:</b> Gibt an, ob eine Person sv-pfl beschäftigt o.bes. Merkmale, auszubildend, geringfügig entlohnt (ab 99), in Alters-TZ beschäftigt (nur West, ab 99), Praktikant (nur West, ab 99) oder Werkstudent (nur West, ab 99) ist	
Status	<b>neu:</b> Unterscheidet sozialversicherungspflichtig Beschäftigte, geringfügige Beschäftigte und Leistungsempfänger	
Betriebsnummernzähler	Gibt die Ordnungsnummer der unterschiedlichen Betriebe im Erwerbsverlauf an	
Betriebsort (Region)	<b>348</b> Kreisregionen auf Basis der Kreis-kennziffer ( <b>270</b> in den alten Bundesländern, <b>77</b> in den neuen Bundesländern und Berlin)	<b>343</b> Kreisregionen auf Basis der Kreis-kennziffer ( <b>266</b> in den alten Bundesländern, <b>76</b> in den neuen Bundesländern und Berlin)
Wirtschaftszweig	<b>13</b> Wirtschaftszweigklassen	<b>16</b> Wirtschaftszweigklassen

**A3: Die Merkmale und ihre Bedeutung**

- **persnr**  
Fortlaufender, systemfreier Zähler für die Person, ersetzt die Versicherungsnummer als Identifikator.
- **spell**  
Laufende Nummer der jeweiligen Datenzeile zu einer Person im Konto.
- **nspell**  
Gesamtzahl der Spells, die für die jeweilige Person im Konto enthalten sind.
- **satznr**  
Meldungen, die durch ein Episodensplitting in mehrere Episoden zerlegt wurden, können anhand gleicher Satznummern identifiziert werden. Allerdings haben auch ergänzte Meldungen die gleiche Satznummer wie die zeitlich vorangehende Meldung.
- **status**  
Angabe, ob es sich um sozialversicherungspflichtiges oder ein geringfügiges Beschäftigungsverhältnis oder eine Episode mit Leistungsbezug handelt.
- **btyp**  
Beschäftigungstyp bzw. Leistungsart; identifiziert erstens, ob es sich bei der Meldung um eine Beschäftigungs- oder Leistungsbezugsmeldung handelt und zweitens, ob es sich um eine originäre oder ergänzte Meldung handelt.

- ***lart\_grp***  
Art der Lohnersatzleistung (ALG, ALHi usw.).
- ***typ1***  
Rentenversicherungsträger.
- ***level***  
Bezeichnet Meldungen, die parallele Zeiträume mit anderen Meldungen aufweisen z. B. wegen Mehrfachbeschäftigung (die erste parallele Episode hat den Wert 0, die zweite den Wert 1, usw.).
- ***nlevel***  
Gibt die Anzahl paralleler Zustände in einem Zeitraum an.
- ***kom\_quel***  
Identifiziert parallele Zustände von Leistungsbezug und (geringfügiger) Beschäftigung.
- ***atag, amon, ajahr***  
Tag, Monat und Jahr des Beginns einer Episode.
- ***etag, emon, ejahr***  
Tag, Monat und Jahr des Endes einer Episode.
- ***ow\_knz***  
Rechtskreis Ost/West des aktuellen Betriebes bzw. bei Leistungsbezug des aktuell zuständigen Arbeitsamtes.
- ***ow\_kto***  
Gibt an, ob die erste Beschäftigung einer Person bei einem Betrieb in Ost- oder Westdeutschland stattfand. Obwohl wegen der geringen Qualität der Daten Sätze im Osten vor 1992 gelöscht wurden, weist das Merkmal *ow\_kto* den Rechtskreis der ersten Beschäftigung auch dann aus, wenn diese vor 1992 aufgenommen wurde. Daher werden in einigen Fällen Personen hier der Ausprägung „Ostdeutschland“ zugewiesen, deren erste Beschäftigung oder Leistungsepisode *im Datensatz* in Westdeutschland registriert ist.
- ***sex***  
Geschlecht.
- ***gebjahr***  
Geburtsjahr; wird für das gesamte Konto nicht ausgewiesen, wenn für eine Person zu irgendeinem Zeitpunkt (meist vor dem 16. oder nach dem 62. Lebensjahr) die Zahl der Personen des selben Alters so gering ist, dass die Anonymität nicht gewährleistet ist.
- ***beruf***  
nicht erlernter Beruf, sondern aktuell ausgeübte Tätigkeit; wird bei Leistungsbezug aus vorheriger Beschäftigungsmeldung übernommen.
- ***berufstg***  
Stellung im Beruf bzw. teilzeit- oder vollzeitbeschäftigt, wobei als „große“ Teilzeit eine Beschäftigung von wöchentlich 18 Stunden und mehr gilt, die aber keine Vollbeschäftigung ist. Als „kleine“ Teilzeit gilt dementsprechend eine Beschäftigung von wöchentlich unter 18 Stunden. Diese Grenzziehung besteht seit 1.1.1988. Bis 31.12.1978 war die Teilzeitgrenze 20 Wochenstunden, bis

31.12.1987 lag sie bei 15 Wochenstunden Arbeitszeit; bei Leistungsbezug findet wiederum eine Übernahme aus vorheriger Beschäftigungsmeldung statt.

- ***bild***

Schul- und Berufsausbildung (bei Leistungsbezug Übernahme aus vorheriger Beschäftigungsmeldung).

- ***entg\_tag***

Sozialversicherungspflichtiges Bruttoentgelt, als Entgelt pro Tag ausgewiesen. Das Tagesentgelt ist auf ganze Beträge gerundet und wird bis 1998 als DM-Betrag, ab 1999 als Euro-Betrag ausgewiesen. Es wurde für die Jahre 1975–1998 wie bisher unterhalb der Geringfügigkeitsgrenzen und oberhalb der Beitragsbemessungsgrenzen der jeweiligen Jahre gruppiert. Ab dem Jahr 1999 ist eine Vergrößerung am unteren Rand nicht mehr nötig, da durch die Erfassung der geringfügigen Beschäftigten auch die Entgeltkategorien unterhalb der Geringfügigkeitsgrenze ausreichend besetzt sind. Im Zeitverlauf wurden im Übrigen immer mehr Lohnzuschläge sozialversicherungspflichtig und somit Teil des in der Beschäftigtenstichprobe erfassten Entgelts.

- ***nation***

Nur für Westdeutschland: Staatsangehörigkeit „deutsch“ oder „andere“ (bei Leistungsbezug Übernahme aus vorheriger Beschäftigungsmeldung).

- ***pers\_gr***

Unterscheidet sozialversicherungspflichtige Beschäftigte, Auszubildende und (ab 1999) geringfügig Beschäftigte; im Westen und ab 1999 auch Personen in Altersteilzeit, PraktikantInnen und WerkstudentInnen.

- ***grund***

Grund der Abgabe einer Meldung; bei Beschäftigten z.B. Jahresmeldung oder Ende der Beschäftigung, bei Leistungsbeziehern z.B. Ablauf der Maßnahme oder Sperrzeit.

- ***bnn***

Die Ordnungszahl gibt an, in welchem der *unterschiedlichen* Betriebe im Erwerbsverlauf die aktuelle Beschäftigung stattfindet. Wenn eine Person bereits zwei Betriebswechsel hinter sich hat und nun wieder beschäftigt ist, dann besitzt bnn den Wert 3, sofern es sich um drei unterschiedliche Betriebe handelt. Wenn aber der zweite Wechsel wieder in den ursprünglichen Betrieb zurück erfolgt, dann nimmt bnn den Wert 1 an..

- ***region***

Region, in der der Beschäftigungsbetrieb liegt (bei Leistungsbezug Übernahme aus vorheriger Beschäftigungsmeldung).

- ***wzwg***

Klassifikation der Wirtschaftszweige des beschäftigenden Betriebes (bei Leistungsbezug Übernahme aus vorheriger Beschäftigungsmeldung).