SSOAR

Open Access Repository
www.ssoar.info

Das Beschäftigtenpanel der Bundesanstalt für Arbeit

Koch, Iris; Meinken, Holger

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:
Koch, I., & Meinken, H. (2003). Das Beschäftigtenpanel der Bundesanstalt für Arbeit. *ZA-Information / Zentralarchiv für Empirische Sozialforschung*, 52, 159-170. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-198980

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.

gesis Leibniz-Institut für Sozialwissenschaften

Mitglied der Leibniz-Gemeinschaft

Das Beschäftigtenpanel der Bundesanstalt für Arbeit[1]

von Iris Koch und Holger Meinken[2]

1 Einleitung

Die Nachfrage nach detaillierten Informationen zum Arbeitsmarkt hat in den letzten Jahren ständig zugenommen, und sie bezieht sich nicht mehr nur auf die regelmäßig von der Bundesanstalt für Arbeit (BA) erstellten statistischen Materialien und Sonderauswertungen. Vielmehr wird von Forschern ein erleichterter Zugang zu Einzeldaten gewünscht, der dann differenzierte eigene Analysen auf der Mikroebene ermöglicht. Im Bereich Statistik der BA-Hauptstelle wird deshalb seit 1999 ein Forschungsvorhaben bearbeitet, in dessen Rahmen die Bereitstellung von anonymisierten Mikrodaten zu sozialversicherungspflichtigen Beschäftigungen vorbereitet und durchgeführt wird. Dieses Vorhaben steht in der Entwicklungslinie der von der Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI) verfolgten Bestrebungen, den Zugang zu den Daten der statistischen Ämter und der sonstigen öffentlichen Datenproduzenten grundsätzlich zu verbessern (KVI 2001).

Da die fraglichen Einzeldaten der Beschäftigungsstatistik dem Sozialdatenschutz des Sozialgesetzbuches (SGB) unterliegen, dürfen nicht die "Originaldaten", sondern lediglich anonymisierte Datensätze an externe wissenschaftliche Einrichtungen übermittelt werden. Es waren daher zunächst entsprechende Anonymisierungsverfahren zu entwickeln und zu prüfen, bevor an eine Datenübermittlung gedacht werden konnte. Mittlerweile sind die Arbeiten jedoch so weit gediehen, dass eine Weitergabe der anonymisierten Daten über das Zentralarchiv erfolgen kann.

---

1 Das diesem Bericht zugrunde liegende Vorhaben wird mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 07SWF50 gefördert und trägt den Titel "Zusammenarbeit der Bundesanstalt für Arbeit mit externen Forschern auf dem Gebiet der Arbeitsmarktstatistik".

2 ***Iris Koch*** und ***Holger Meinken*** sind wissenschaftliche Mitarbeiter im Referat Beschäftigungsstatistik der Bundesanstalt für Arbeit in Nürnberg. E-Mail: iris.koch@arbeitsamt.de, holger.meinken@arbeitsamt.de

Der als Panel aufgebaute Datensatz zu sozialversicherungspflichtigen Beschäftigungen ist unter der ZA-Studien-Nr. 3887 für Zwecke der Arbeitsmarkt- und Berufsforschung verfügbar. Die Grundlage des Panels bilden Stichproben aus den Quartalsdaten der Beschäftigungsstatistik der Jahre 1998-2002. Die Einzelangaben zu den Beschäftigungsverhältnissen wurden anonymisiert und zu einem *Scientific Use File* aufbereitet. In diesem Datensatz sind einerseits alle wesentlichen meldepflichtigen Individualmerkmale der sozialversicherungspflichtig Beschäftigten und andererseits einige zusätzliche Merkmale zu den jeweiligen Beschäftigungsbetrieben enthalten. Beginnend mit dem 1. Quartal 1998 beinhaltet das Panel gegenwärtig 18 Wellen mit je ca. 600.000 Beschäftigten (einschließlich der geringfügig Beschäftigten). Das Beschäftigtenpanel wird zukünftig einmal jährlich aktualisiert bzw. ergänzt.

2 Datenquellen

Die vorrangigen Datenquellen des Panels sind die Quartalsdaten der Beschäftigungsstatistik der BA. Die Quartalsdaten resultieren aus Informationen, die das Meldeverfahren zur Sozialversicherung liefert. Im Meldeverfahren werden Angaben zu allen sozialversicherungspflichtigen Beschäftigungen im Bundesgebiet erhoben, unabhängig davon, ob eine Beschäftigung nur einen Tag oder mehrere Jahre dauert. Das Meldeverfahren besteht in seinen Grundzügen seit 1973 und wurde zuletzt 1999 geändert (***Neidert*** 1998).

Die Meldung der einzelnen Beschäftigungen erfolgt vom jeweiligen Arbeitgeber. Die Betriebe haben innerhalb bestimmter Fristen ihre Meldungen über die sozialversicherungspflichtig beschäftigten Personen an die Sozialversicherungsträger abzugeben, soweit für die Beschäftigten die Meldetatbestände vorliegen. Ein neu begründetes Beschäftigungsverhältnis muss durch den Arbeitgeber mit einer Anmeldung an die Träger der Sozialversicherung mitgeteilt werden. Sofern ein Beschäftigungsverhältnis beim gleichen Arbeitgeber ununterbrochen besteht, hat der Arbeitgeber nach Ablauf eines Kalenderjahres eine so genannte Jahresmeldung mit der Angabe des sozialversicherungspflichtigen Bruttojahresentgelts zu erstatten. Endet das Beschäftigungsverhältnis, hat der Arbeitgeber eine Abmeldung mit Angaben zum rechtlichen Ende des Beschäftigungsverhältnisses und dem bis dahin erzielten Bruttoentgelt vorzulegen. Damit ist ein Beschäftigungszeitraum mit dem darin erzielten Bruttoentgelt durch Jahresmeldungen lückenlos nachgewiesen, ein Beschäftigungsende kann aus der Abmeldung eindeutig abgeleitet werden. Diese Meldungen werden von der BA entgegengenommen und statistisch ausgewertet. So werden diese Meldungen beispielsweise ausgewertet, um den Bestand an Be-

schäftigungen zu den jeweiligen Quartalsstichtagen (31.3., 30.6., 30.9., 31.12.) zu ermitteln.

Zur Sicherstellung der Qualität des übermittelten Datenmaterials ist von den Trägern der Sozialversicherung ein gemeinsames Prüfungsprogramm entworfen worden, das sowohl von den Krankenkassen als auch von den Rententrägern bei Entgegennahme der Meldedaten eingesetzt wird. Die Datenprüfungen beschränken sich weitgehend auf die versicherungsrechtlich entscheidenden Angaben zur Versicherungsnummer, zu den Adressangaben, zu den angegebenen Beschäftigungszeiten und zum sozialversicherungspflichtigen Entgelt. Die darüber hinaus in der Meldung enthaltenen soziodemographischen Angaben werden einer knappen formalen Prüfung unterzogen.

Die Vorteile des Datenmaterials der Beschäftigtenstatistik der BA lassen sich recht gut durch einen Vergleich mit Umfragedaten verdeutlichen. Aufgrund des gesetzlich geregelten Meldeverfahrens weisen die prozessproduzierten Sozialversicherungsdaten nicht die Einschränkungen auf, die bei Umfragedaten üblicherweise vorkommen. So sind insbesondere die Angaben zum sozialversicherungspflichtigen Entgelt und zur Beschäftigungszeit als sehr zuverlässig anzusehen, weil jeder Beschäftigte eine Kopie jeder Meldung vom Arbeitgeber erhält. Damit ist im Verfahren selbst eine Kontrolle der Meldeangaben durch den Beschäftigten enthalten. Gerade bezüglich des Entgelts und der Beschäftigungszeiten sind die Angaben nicht von den typischen Mängeln der Umfragedaten zu diesem Themenbereich wie Erinnerungsfehlern, Antwortverweigerungen oder bewussten Falschangaben behaftet. Die bestehende Meldepflicht sorgt für einen zuverlässigen Datenfluss, so dass auch die Probleme wie z. B. Befragungsausfall und Antwortverweigerungen vernachlässigbar sind. Insbesondere für die Längsschnittperspektive besitzen die Daten den Vorzug, dass keine Beobachtungseinheiten wegen Teilnahmeverweigerung, Wohnortwechsel, zeitweiliger Abwesenheit o. ä. ausfallen. Die Ausfälle beschränken sich bei den vorliegenden Daten lediglich auf Beschäftigte, die vorübergehend oder endgültig aus einer Beschäftigung ausgeschieden sind. Dafür rücken allerdings auch neue Personen nach, nämlich jene, die erstmalig oder erneut eine sozialversicherungspflichtige Beschäftigung aufnehmen.

Durch die gesetzliche Grundlage des Meldeverfahrens fallen für ca. 75-80 % aller Erwerbstätigen der Bundesrepublik Deutschland innerhalb eines Jahres Meldedaten an, die in der BA für statistische Zwecke verarbeitet werden. Damit stellen die Beschäftigtendaten eine einzigartige Grundlage für Auswertungen zu erwerbsstatistischen Fragestellungen dar. Gleichzeitig ermöglichen die Meldedaten eine

Verknüpfung mit dem Datenmaterial aus anderen Geschäftsbereichen der Arbeitsverwaltung.

Der offenkundige Vorteil der Meldedaten ist darin begründet, dass es sich bei ihnen um eine Vollerhebung der sozialversicherungspflichtig Beschäftigten handelt. Damit können die Stichproben aus den Beschäftigtendaten, wie z.B. das Beschäftigtenpanel, hinsichtlich ihrer Abbildungstreue geprüft werden. Aufgrund der Kenntnis der Verteilung einzelner Merkmale in der Grundgesamtheit (aller sozialversicherungspflichtig Beschäftigten), lässt sich für etwaige Stichproben leicht prüfen, ob ihre Merkmalsverteilungen signifikant von denen der Grundgesamtheit abweichen.

Als Nachteile der Beschäftigtendaten wären der geringe Merkmalsumfang und der geringe Differenzierungsgrad innerhalb einzelner Merkmalsdimensionen zu nennen. Dies kann als spezifisches Charakteristikum für sekundärstatistisches Datenmaterial angesehen werden, das auf der Grundlage sehr großer Erhebungsgesamtheiten entstanden ist.

Ebenfalls als nachteilig einzuschätzen ist die meldetechnisch bedingte und unvermeidbare Verzögerung des Dateneingangs mit ihren negativen Auswirkungen für aktuelle Auswertungen. Aktuellste Auswertungen basieren auf einer Wartezeit von sechs Monaten bezogen auf den jeweiligen Quartalsstichtag. Wenngleich damit im Regelfall ca. 90 % der jeweiligen Meldungen erfasst werden und recht zuverlässige Ergebnisse liefern, bleibt für die amtlichen Daten doch ein Vorläufigkeitsbehalt bestehen. Kommen dann noch Verfahrensumstellungen - wie jüngst beim Meldeverfahren - oder nachträgliche Prüfbedarfe bei den zuständigen Versicherern hinzu, dann können auch längere Wartezeiten entstehen.

Zu den wesentlichen Neuerungen im Meldeverfahren ab 1999 zählen die Einführung der Meldepflicht für geringfügige Beschäftigungen, die verschiedenen Regelungen zu Altersteilzeit und Vorruhestand, die Neuregelung der dualen Berufsausbildung oder die mit der Einführung der Pflegeversicherung verbundenen Tätigkeiten. Im Meldeverfahren haben sich diese Veränderungen sowohl durch die Erweiterung der eine Meldung auslösenden Sachverhalte als auch durch die Veränderung der zu meldenden Merkmale niedergeschlagen.

Aus dem Datenmaterial der Meldungen wird der Bestand an sozialversicherungspflichtig und geringfügig Beschäftigten zu den vier Quartalsstichtagen eines Jahres ermittelt. Dazu wird durch ein umfangreiches Abfrageschema die letzte beschäftigungsanzeigende Meldung aus den Meldungen des einzelnen Versicherten gezogen und geprüft, ob die betreffende Person zum Quartalsstichtag sozialversicherungspflichtig beschäftigt war oder nicht. Aus den eigentlichen Meldedaten wird also eine

Querschnittinformation generiert. Den resultierenden Datensatz nennt man Quartalsdatei. Aus diesen Quartalsdateien werden dann die Stichproben des Beschäftigtenpanels gezogen.

Zur Erweiterung des Informationsgehaltes werden außerdem noch Merkmale der jeweiligen Beschäftigungsbetriebe zum Panel hinzugefügt. Über die Betriebsnummer kann aus der Betriebedatei zu jeder Beschäftigung die wirtschaftsfachliche und regionale Gliederung des Beschäftigungsbetriebes angegeben werden. Darüber hinaus werden aus der Grundgesamtheit aller Beschäftigten die Betriebsgrößen und die Beschäftigtenanteilswerte für jeden Quartalsstichtag ermittelt.

3 Datenbeschreibung

3.1 Stichprobenverfahren

Das Beschäftigtenpanel basiert auf einer Personenstichprobe anhand einer Geburtstagsauswahl. Das Geburtsdatum ist in der jeweiligen Sozialversicherungsnummer der Beschäftigten enthalten, die für alle Beschäftigungsmeldungen als eineindeutiger Identifikator verwendet wird. Bei erstmaliger Aufnahme einer sozialversicherungspflichtigen Beschäftigung erhält ein Beschäftigter seine Sozialversicherungsnummer zugewiesen, und diese ändert sich im Laufe des weiteren Erwerbslebens dann nicht mehr. Die Stichprobenziehung erfolgt dadurch, dass aus den Beschäftigungsmeldungen der einzelnen Quartalsstichtage (Quartalsdaten) nur diejenigen Beschäftigten ausgewählt werden, die an einem von sieben als konstant vorgegebenen Tagen des Jahres Geburtstag haben. Somit erhält man eine Stichprobe von 1,92 % aus den zum jeweiligen Stichtag beschäftigten Personen.

Für die meisten der durch ihr Geburtsdatum ausgewählten Beschäftigten, die langfristig in einem Beschäftigungsverhältnis verbleiben, ändert sich von Quartal zu Quartal nichts Wesentliches. Für Beschäftigte, die den Arbeitgeber wechseln, ändern sich die betreffenden Merkmale zur Art der Beschäftigung und zum Betrieb. Sie bleiben jedoch dauerhaft in den Stichproben, weil sie aufgrund ihrer Sozialversicherungsnummer auch zu den laufend neu hinzukommenden Quartalsstichtagen identifiziert werden können. Ist eine Person zu einem Quartalsstichtag nicht sozialversicherungspflichtig beschäftigt, so gibt es in der betreffenden Welle keine Datenzeile dieser Person. Erst bei erneuter Beschäftigung zu einem folgenden Quartalsstichtag würde die Person mit den Merkmalen zu dieser Beschäftigung wieder in den Datensatz aufgenommen.

Diese systematische Zufallsauswahl bietet eine Reihe von Vorteilen für die Stichprobenzusammensetzung. Sie sichert eine dauerhaft strukturtreue Abbildung der Grundgesamtheit durch die Stichprobe, und zwar sowohl im Querschnitt als auch im Längsschnitt. Stichprobenverluste wegen Panelmortalität werden automatisch ausgeglichen. Ältere Beschäftigte, die aus dem Erwerbsleben ausscheiden, werden durch junge Personen, die ihre erste Beschäftigung aufnehmen und eines der ausgewählten Geburtsdaten aufweisen, ersetzt. Die Besetzungen der Geburtskohorten sind immer proportional zu denen der Grundgesamtheit, womit die Stichprobe auch die Durchführung von Kohortenanalysen ermöglicht. Außerdem gewährleistet das Auswahlverfahren die Abbildung vorhandener Saison- und Konjunkturmuster, weil immer ein konstanter Anteil aller Beschäftigten ausgewählt wird. Gegenüber dem Mikrozensus mit seiner Flächenstichprobe hat die Stichprobe des Beschäftigtenpanels beispielsweise den Vorteil, dass die regionale Mobilität von Beschäftigten analysierbar ist und nicht zu Stichprobenausfällen führt.

Da die Stichproben der einzelnen Wellen proportional zur jeweiligen Grundgesamtheit sind, variieren ihre Umfänge mit dem konjunkturellen und saisonalen Verlauf der Beschäftigtenzahlen. Tabelle 1 gibt einen Überblick zu den Häufigkeiten von Stichprobe und Grundgesamtheit der letzten Welle. Bei einer gesamten Anzahl von ca. 32 Mio. Beschäftigten weisen die Stichproben rund 600.000 Fälle auf. Die ersten fünf Wellen des Panels kommen nur auf einen Stichprobenumfang von ca. 500.000 Beschäftigten, weil zu diesen Zeitpunkten die geringfügig entlohnten Beschäftigungen noch nicht in den Quartalsdaten enthalten waren.

Tabelle 1 Stichprobenumfang des BA-Beschäftigtenpanels

Quartalsdaten 30.6.2002	Westdeutschland		Ostdeutschland*		Deutschland	
	Stichprobe	Gesamt	Stichprobe	Gesamt	Stichprobe	Gesamt
Sozialversicherungspflichtig Beschäftigte	426.055	22.182.502	104.065	5.388.645	530.120	27.571.147
Geringfügig entlohnte Beschäftigte	68.942	3.599.798	10.768	569.368	79.710	4.169.166
Summe	494.997	25.782.300	114.833	5.958.031	609.830	31.740.313

* einschließlich Berlin

Um die Qualität der Stichprobe zu prüfen, wurden die univariaten Verteilungen aller Merkmale mit denen der Grundgesamtheit verglichen. Das Ergebnis war eine sehr gute Repräsentation der Grundgesamtheit durch die Stichprobe.

3.2 Merkmalsumfang

Der Datensatz umfasst 52 Merkmale, bei denen man zwischen den individuellen Merkmalen der Beschäftigten und den betrieblichen Merkmalen der jeweiligen Beschäftigungsbetriebe unterscheiden kann (Tabelle 2).

Tabelle 2 Merkmalsumfang des BA-Beschäftigtenpanels

Individualmerkmale	**Betriebliche Merkmale**
• Personennummer • Stichtagsdatum • Nummer der Panelwelle • Art der Meldung • Geschlecht • Alter in Jahren • Staatsangehörigkeit (23 Ausprägungen) • (Aus-)Bildungsabschluss (7 Ausprägungen) • Stellung im Beruf/Arbeitszeit (9 Auspräg.) • Rentenversicherungsträger (Arbeiter/ Angestellte) • sozialversicherungspflichtiges Entgelt • Personengruppe gemäß VDR (10 Auspräg.) • Beitragsgruppenschlüssel (6 Auspräg.) • Beruf / ausgeübte Tätigkeit (ca. 300 Ausprägungen.) • Betriebswechsel (5 Ausprägungen)	• Betriebsgrößenklasse (8 Ausprägungen) • Wirtschaftszweig (48 Ausprägungen) • Region (Ost-/Westdeutschland) • Anteil der Frauen • Anteil der Beschäftigten bestimmter Altersklassen (11 Klassen) • Anteil mit deutscher Staatsangehörigkeit • Anteil der Teilzeitbeschäftigten • Anteil der Beschäftigten mit bestimmten Bildungsabschlüssen (6 Klassen) • Anteil der Auszubildenden • Anteil der Arbeiter • Anteil der Facharbeiter • Anteil der Angestellten • Anteil Angestellten- /Arbeiterrenten-versicherung

3.3 Anonymisierungsverfahren

In der Bundesrepublik werden bereits seit einem guten Jahrzehnt in Einzelprojekten amtliche Mikrodaten anonymisiert, um sie als Scientific Use Files für Forschungszwecke leichter nutzbar zu machen. Die ersten Anonymisierungsverfahren wurden für den Mikrozensus entwickelt (***Müller*** et al. 1991) und in einem Pilotprojekt umgesetzt (***Köhler*** et al. 2000). In der Folge entstanden dann weitere anonymisierte Datensätze, die im Wesentlichen auf den gleichen oder ähnlichen Anonymisierungsverfahren aufbauten.

Allgemein lassen sich die bisher in der amtlichen Statistik eingesetzten Verfahren zur Anonymisierung von Mikrodaten folgendermaßen charakterisieren (***Köhler*** 1999):

- Verfremden der Daten (Überlagern mit Zufallsfehler, Erzeugung künstlicher Datensätze),
- Vergröbern der Daten (bei schwach oder extrem besetzten Merkmalsausprägungen),
- Weglassen von Untersuchungseinheiten oder Merkmalen (Stichprobenziehung, Entfernen von Identifikatoren),
- Systemfreie Sortierung der Daten.

Im Beschäftigtenpanel wurde auf die Verfremdung von Einzeldaten verzichtet, weil die Auswirkungen für die Auswertungen verfremdeter Daten und deren Transparenz bisher nicht hinreichend erforscht sind bzw. zum Teil eher kritisch eingeschätzt werden müssen (***Lechner/Pohlmeier*** 2003). Die konkreten Anonymisierungsschritte orientieren sich an denen, die bereits für den Mikrozensus entwickelt wurden, und sehen wie folgt aus:

(1) Ziehung einer Zufallsstichprobe aus der Grundgesamtheit mit einer Auswahlwahrscheinlichkeit von 1,92 %,

(2) Ersetzen der Sozialversicherungsnummer durch eine zufallsgenerierte, systemfreie Personennummer. Entfernen von sonstigen Identifikatoren und besonders sensiblen Merkmalen wie z.B. Geburtsdatum und Betriebsnummer,

(3) Systemfreie Sortierung nach der Personennummer,

(4) Vergröbern von Merkmalsausprägungen:

- Regionalangabe wird dichotomisiert (West- und Ostdeutschland).
- Bei der Staatsangehörigkeit sind jene Nationalitäten mit anderen zusammengefasst, die einzeln weniger als ca. 50.000 Personen in der Grundgesamtheit umfassen.
- Für alle anderen Merkmale wurden bei Bedarf die Merkmalsausprägungen so zusammengefasst, dass die univariaten Auszählungen in der Grundgesamtheit mindestens ca. 5.000 Personen umfassen. Ausnahmen bilden lediglich Residualkategorien oder bereits zusammengefasste Ausprägungen, deren Informationsgehalt ohnehin wenig spezifisch ist. Von dieser Anonymisierung sind im Wesentlichen die Merkmale Beruf und Wirtschaftszweig betroffen.

- Für das Merkmal Alter wurden die Randklassen (jünger als 14, älter als 70) zusammengefasst.

Mit diesen Anonymisierungen bleibt insgesamt ein sehr hohes Untersuchungspotenzial der Daten erhalten und auch die Integrität der Daten ist weitgehend gewahrt. Gravierende Einschränkungen sind lediglich für die Regionalangaben vorhanden. Es ist klar, dass die Beschränkung auf zwei Regionen (Ost- und Westdeutschland) für Forscher eine starke Restriktion darstellt, insbesondere wenn sie an Regionalanalysen interessiert sind. Allerdings ergäbe sich ein nicht vertretbares Risiko der Identifikation einzelner Beschäftigungsbetriebe, wenn man, zusammen mit der Information über die wirtschaftsfachliche Zuordnung eine tiefere regionale Gliederung ausweisen würde. Eine differenzierte Darstellung sowohl der wirtschaftsfachlichen als auch der regionalen Gliederung ist somit nicht möglich. Man kann nur für eines dieser beiden Merkmale eine tiefe Gliederung ausweisen. Als Alternative zu dem gewählten Vorgehen käme die Anonymisierung aller anderen betrieblichen Informationen in Frage, was jedoch wiederum zu Informationsreduktionen in den Daten führen würde. Einerseits sind Betriebsdaten weitreichender zu anonymisieren als Beschäftigtendaten und andererseits sind die Anonymisierungsverfahren für Betriebsdaten bisher nicht zufrieden stellend entwickelt (***Brand*** 1999).

Als Lösung des beschriebenen Dilemmas eingeschränkter Untersuchungsmöglichkeiten bietet sich für Forscher die Schalterstelle zum Beschäftigtenpanel an, in deren Rahmen dann z.B. auch differenzierte Regionalanalysen durchgeführt werden können. Grundsätzlich sollen jene Forschungsprojekte, die im Einzelfall auf differenziertere Merkmalsausprägungen angewiesen sind, als diese im anonymisierten Beschäftigtenpanel vorliegen, die Möglichkeit der Schalterstellenauswertung nutzen können. Das bedeutet, dass externe Forscher ihre Auswertungsprogramme (Programm-Dateien) per E-Mail an die Bundesanstalt für Arbeit senden, dort ausführen lassen und das Ergebnis (Output-Dateien) nach einer Datenschutzprüfung zurück bekommen. Als erfolgreiches Beispiel für die Praktikabilität eines solchen Verfahrens ist die Schalterstelle des IAB-Betriebspanels zu nennen (***Kölling*** 2001).

3.4 Unterschiede von BA-Beschäftigtenpanel und IAB-Beschäftigtenstichprobe

Neben dem Beschäftigtenpanel existieren bereits die anonymisierten Daten der IAB-Beschäftigtenstichprobe bzw. der Regionalstichprobe (***Bender*** et al. 1999, ***Haas*** 2001). Auch wenn diese Datensätze weitgehend auf der gleichen primären Datenquelle, den Beschäftigungsmeldungen, basieren, gibt es doch einige wesentliche Unterschiede.

Beim Beschäftigtenpanel erfolgt die Aufbereitung der Daten nach einzelnen Querschnitten (Quartalsstichtagen), die zu einem Panel zusammengefügt werden. Die Beschäftigtenstichprobe besteht dagegen aus Ereignisdaten, die die gesamten Beschäftigungsverläufe mit tagesgenauen Angaben zum Beginn und zum Ende von Beschäftigungen einzelner Personen beinhalten. Die Beschäftigtenstichprobe eignet sich daher besonders für Untersuchungen von Erwerbsverläufen. Das Beschäftigtenpanel beinhaltet zwar auch die individuellen Zustandswechsel der zugrunde liegenden Berufsverläufe, allerdings nur in einem gröberen Zeitraster (quartalsweise). Die Vorteile des Panels liegen für Forscher darin, dass die Daten einerseits die amtlichen Quartalsdaten sehr gut repräsentieren und für stichtagsorientierte Auswertungen direkt verwertbar sind, andererseits bieten sie auch die Möglichkeit zeitbezogener Untersuchungen, seien es Zeitreihen-, Panel- oder Kohortenanalysen.

Aus den Ereignisdaten der Beschäftigtenstichprobe sind die Querschnittinformationen des Beschäftigtenpanels nicht ohne weiteres rekonstruierbar, weil unterschiedliche Datenaufbereitungen zugrunde liegen. Die Quartalsdaten der BA, aus denen das Panel als Stichprobe gezogen wird, beinhalten spezielle Verfahren, mit denen z.B. jene Meldungen ausgeglichen werden, die nicht innerhalb der halbjährigen Wartefrist bei der BA eingegangen sind. Wenn also ohnehin geplant ist, querschnittbezogene Auswertungen wie beispielsweise Quartals- oder Jahresvergleiche vorzunehmen, erspart das Panel einem die aufwändige Überführung von Ereignisdaten in Querschnittdaten und erleichtert zudem Hochrechnungen auf die amtlichen Daten.

Eine Ableitung der Querschnitte aus den Ereignisdaten der Beschäftigtenstichprobe wird zusätzlich auch dadurch erschwert, dass diese Daten im Längsschnitt anonymisiert (verschoben) sind. Beim Beschäftigtenpanel wurde bislang auf eine Längsschnittanonymisierung verzichtet, weil der im Einzelfall rekonstruierbare Zeitverlauf mit Lücken von 3 Monaten eine verfahrensbedingte Unschärfe aufweist. Auftretende Berufs- und Betriebswechsel der Beschäftigten sind somit nur quartalsweise erfasst.

Das Zeitfenster der beiden Datensätze ist grundsätzlich verschieden. Während die IAB-Beschäftigtenstichprobe einen sehr langen Zeitraum (1975-1997) abdeckt, beziehen sich die Daten des Beschäftigtenpanels auf eine bislang kurze Zeitspanne (1998-2002). Der hauptsächliche Fokus des Panels ist der aktuelle zeitliche Rand. Die Bereitstellung von aktuelleren Daten ist deshalb möglich, weil die Quartalsdaten mit Wartezeiten von 6 Monaten erstellt werden. Im Gegensatz dazu ist bei der IAB-Beschäftigtenstichprobe eine deutlich längere Wartezeit (mindestens 18 Monate) erforderlich, um möglichst auch die verspätet eintreffenden Beschäftigungsmeldungen einbeziehen zu können.

Inhaltlich bietet die IAB-Beschäftigtenstichprobe den Vorteil, dass sie auch Informationen über Zeiten des Leistungsbezugs (Arbeitslosengeld, -hilfe) einbezieht. Diese Angaben sollen dem BA-Beschäftigtenpanel in der nächsten Version hinzugefügt werden.

Zusammenfassend kann man sagen, dass die Unterschiede von Beschäftigtenpanel und Beschäftigtenstichprobe in der verschiedenen Datenaufbereitung und den verschiedenen Auswertungsdesigns bestehen. Bei der Beschäftigtenstichprobe liegt der Schwerpunkt auf Verlaufsanalysen, beim Beschäftigtenpanel auf Untersuchungen von stichtagsorientierten Querschnitten und ggf. Veränderungen zwischen Stichtagen. Je nach Forschungsfrage haben beide Datensätze ihre Vor- und Nachteile.

3.5 Bedingungen der Datenübermittlung

Aufgrund des geltenden Datenschutzrechts für Sozialdaten müssen bei der Übermittlung des Beschäftigtenpanels einige Bedingungen beachtet werden. Die wissenschaftliche Einrichtung, die das Panel bestellen möchte, darf die Daten nur für den Zweck der Arbeitsmarkt- und Berufsforschung verwenden, und die Nutzung ist auf die Dauer des jeweiligen Forschungsvorhabens befristet. Zur Regelung der formellen Bedingungen wird ein *Nutzervertrag* abgeschlossen. Das Beschäftigtenpanel darf als *Scientific Use File* lediglich im Rahmen von Forschungsprojekten, und nicht für die (universitäre) Lehre verwendet werden.

Die Wellen des Beschäftigtenpanels werden als ein Paket in Rohdatenform (ASCII-Dateien) geliefert. Die bestehenden 18 Wellen benötigen unkomprimiert ca. 2,2 GB Plattenspeicher. Die Datenlieferung kostet 75 € und beinhaltet eine ausführliche Datenbeschreibung (Codebuch) einschließlich Auszählungen, sowie SAS-Setup und Testdaten.

4 Ausblick

Mit den bisherigen Arbeiten zum BA-Beschäftigtenpanel wurde ein weiterer Grundstein gelegt, um Verwaltungsdaten auch für Forschungszwecke zu erschließen und in breiterer Form zugänglich zu machen. Im laufenden Jahr sind noch eine Reihe an methodischen Verbesserungen und inhaltlichen Erweiterungen des Panels geplant.

Unter anderem sollen die Angaben zum sozialversicherungspflichtigen Bruttoentgelt der Beschäftigten genauer, nämlich retrospektiv, den Wellen des Panels zugewiesen werden. Gegenwärtig sind diese Angaben den betreffenden aktuellen Wellen

zugeordnet. Die rückwärtsgerichtete Zuweisung der Entgelte ist aus Gründen, die im Meldeverfahren selbst liegen, nicht ohne beträchtlichen Aufwand zu realisieren.

Eine andere Aufgabe wird die Integration von weiteren Datenquellen sein. Soweit möglich, sollen dabei die Lücken einzelner Beschäftigter, die in den Wellen auftreten können, mit Informationen über die Arbeitslosigkeit, den Leistungsbezug oder die Teilnahme an arbeitsmarktpolitischen Maßnahmen gefüllt werden. Diese zusätzlichen Angaben erweitern noch einmal das Analysepotenzial des Panels.

Schließlich ist noch darauf hinzuweisen, dass die weitere Pflege und Ergänzung des Panels über das Jahr 2003 hinaus im Rahmen des Forschungsdatenzentrums der BA erfolgen soll. Dort wird eine zukünftig jährliche Aktualisierung des Datensatzes angestrebt.

Literatur

Bender, S./ Haas, A./ Klose, C., 1999: Die IAB-Beschäftigtenstichprobe 1975-1995, S. 104-115 in ZA-Information 45.

Brand, R.; 1999: Anonymität von Betriebsdaten, Beiträge zur Arbeitsmarkt- und Berufsforschung Bd. 237, Nürnberg.

Bundesanstalt für Arbeit (Hrsg.), 2003: BA-Beschäftigtenpanel, 1. Quartal 1998 – 2. Quartal 2002, Codebuch, Nürnberg.

Haas, A., 2001: Die IAB-Regionalstichprobe 1975-1997, S. 128-139 in ZA-Information 48.

Köhler, S., 1999: Anonymisierung von Mikrodaten in der Bundesstatistik und ihre Nutzung – Ein Überblick, S. 133-149 in Statistisches Bundesamt (Hrsg.), Methoden zur Sicherung der statistischen Geheimhaltung, Schriftenreihe Forum der Bundesstatistik Bd. 31, Stuttgart.

Köhler, S./ Schimpl-Neimanns, B./ Schwarz, N., 2000: Pilotprojekt zur Erleichterung der Nutzungsmöglichkeiten von faktisch anonymisierten Mikrodaten, S.30-37 in Wirtschaft und Statistik 1/2000.

Kölling, A., 2001: Ein "Schalter" für die Forschung, IAB-Werkstattbericht Nr. 9, Nürnberg.

Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI) (Hrsg.), 2001: Wege zu einer besseren informationellen Infrastruktur, Baden-Baden.

Lechner, S./ Pohlmeier, W., 2003: Schätzung ökonometrischer Modelle auf der Grundlage anonymisierter Daten, Vortrag für die Tagung "Faktische Anonymisierung wirtschaftsstatistischer Einzeldaten" des Statistischen Bundesamtes und des IAW, 20./21. März 2003.

Müller, W./ Blien, U./ Knoche, P./ Wirth, H., 1991: Die faktische Anonymität von Mikrodaten, Schriftenreihe Forum der Bundesstatistik Bd. 19, hrsg. vom Statistischen Bundesamt, Stuttgart.

Neidert, A., 1998: Neues Meldeverfahren zur Sozialversicherung ab 1999, S. 315-330 in Deutsche Rentenversicherung 5/1998.