

Neue Datenangebote in den Forschungsdatenzentren

Betriebs- und Unternehmensdaten im Längsschnitt

*Maurice Brandt (FDZ des Statistischen Bundesamtes)
Dirk Oberschachtsiek (FDZ der BA im IAB)
Ramona Pohl (FDZ Standort Berlin im Amt für Statistik
Berlin-Brandenburg)*

Inhaltsverzeichnis

| | |
|--|-----------|
| 1. Einleitung | 3 |
| 2. Datenzugangswege | 3 |
| 3. Datenangebot | 4 |
| 3.1 <i>Betriebs- und Unternehmensdaten in den FDZ der Statistischen Ämter des Bundes und der Länder</i> | 4 |
| 3.2 <i>Betriebs- und Unternehmensdaten des FDZ der BA im IAB</i> | 7 |
| 4. Verbesserung des Datenangebots durch das Projekt „Wirtschaftsstatistische Paneldaten und faktische Anonymisierung“ | 9 |
| 4.1 <i>Anonymisierung von Wirtschaftsstatistiken</i> | 10 |
| 4.2 <i>Herausforderungen bei der Anonymisierung Wirtschaftsstatistischer Längsschnittdaten</i> | 11 |
| 5. Ausblick auf weitere Vorhaben | 11 |
| 5.1 <i>AFiD – Amtliche Firmendaten für Deutschland</i> | 12 |
| 5.2 <i>KombiFiD – Kombinierte Firmendaten für Deutschland</i> | 14 |
| 6. Schlussbetrachtung | 15 |

Danksagung:

Das Papier ist auf Grundlage dreier Vorträge im Rahmen der Nutzerkonferenz zum Projekt „Wirtschaftsstatistische Paneldaten und faktische Anonymisierung“ entstanden. Organisiert wurde diese Veranstaltung vom Institut für Angewandte Wirtschaftsforschung e.V. & den Forschungsdatenzentren der BA am IAB und der Statistischen Ämter des Bundes und der Länder am 15.06.07 in Tübingen. Wir möchten den Veranstaltern danken. Besonderer Dank gilt Martin Rosemann für wertvolle Hinweise und Anregungen bei der Erstellung des Papiers. Für den Inhalt bleiben die Autoren allein verantwortlich.

1. Einleitung

Die Bundesagentur für Arbeit und die Statistischen Ämtern des Bundes und der Länder erheben regelmäßig Betriebs- und Unternehmensdaten. Diese Daten bieten ein hohes Informationspotential, da sie sehr viele verschiedene inhaltliche Schwerpunkte abdecken, einen großen Stichprobenumfang aufweisen und aufgrund der bestehenden Auskunftspflicht keine Ausfallproblematik vorliegt. Für die wissenschaftliche Forschung sind diese Daten auf der Basis der bestehenden rechtlichen Grundlagen seit einigen Jahren über die Forschungsdatenzentren (FDZ) der jeweiligen Datenproduzenten zugänglich.

Die FDZ entwickeln das Datenangebot für die wissenschaftliche Forschung kontinuierlich auf der Basis des bestehenden Datenbedarfs und in Kooperation mit der Wissenschaft weiter. Der nachfolgende Beitrag gibt einen Überblick über den aktuellen Stand des Datenangebotes und beschreibt einige richtungweisende Projekte zu dessen Verbesserung. Der Beitrag ist wie folgt gegliedert: Abschnitt 2 beschreibt die jeweiligen Zugangsmöglichkeiten zu den Mikrodaten in den FDZ. Anschließend wird eine Auswahl von Betriebs- und Unternehmensdaten vorgestellt, die mittlerweile für wissenschaftliche Auswertungen zur Verfügung stehen. Abschnitt 4 informiert über das Projekt „Wirtschaftsstatistische Paneldaten und faktische Anonymisierung“ und dessen Beitrag zur Verbesserung des Datenangebots im Bereich der wissenschaftlich nutzbaren Paneldaten über Betriebe und Unternehmen, gefolgt von einem Ausblick auf zukünftige Projekte zur Zusammenführung von Betriebs- und Unternehmensdaten aus unterschiedlichen Quellen in Abschnitt 5. Abschließend wird eine kurze Schlussbetrachtung vorgenommen.

2. Datenzugangswege

Die FDZ machen die Mikrodaten der Betriebs- und Unternehmensstatistiken auf der Basis der geltenden rechtlichen Grundlagen für die wissenschaftliche Forschung zugänglich. Obwohl sich die rechtlichen Grundlagen für die Arbeit der einzelnen FDZ unterscheiden, haben alle FDZ vergleichbare Zugangswege eingerichtet: Als standardisierte Produkte für die wissenschaftliche Forschung bieten alle FDZ so genannte Scientific Use Files an, bei denen es sich um faktisch anonyme Daten handelt. Dies sind Einzelangaben, die nur mit einem unverhältnismäßig hohen Aufwand an Zeit, Kosten und Arbeitskraft einem Auskunftsgibenden zugeordnet werden können. Demnach wird die Möglichkeit einer Deanonymisierung zwar nicht ausgeschlossen, jedoch müssen die Kosten einer versuchten Deanonymisierung über dem Nutzen liegen. Für Forscher, die detailliertere Analysen durchführen wollen, bietet sich die Nutzung eines Gastwissenschaftlerarbeitsplatzes innerhalb der geschützten Räume der FDZ an. Zudem stehen die Mikrodaten der Forschung im Rahmen der kontrollierten Datenfernverarbeitung zur Verfügung. Hier hat der Wissenschaftler keinen direkten Datenzugang, sondern erstellt auf der Basis eines

Testdatensatzes ein Auswertungsprogramm, das von einem Mitarbeiter des zuständigen FDZs auf die formal anonymisierten Mikrodaten angewendet wird. Sowohl bei der Nutzung am Gastwissenschaftlerarbeitsplatz als auch bei der kontrollierten Datenfernverarbeitung werden die berechneten Ergebnisse auf Einhaltung der Geheimhaltung überprüft (Zühlke et al. 2003).¹ Die Nutzungsbedingungen können im Einzelnen in den Internetauftritten der FDZ nachgesehen werden (www.forschungsdatenzentrum.de; www.iab.de).

3. Datenangebot

3.1 Betriebs- und Unternehmensdaten in den FDZ der Statistischen Ämter des Bundes und der Länder

In diesen beiden FDZ können im Bereich der Wirtschaftsstatistiken Mikrodaten für Betriebe und Unternehmen aus ca. 20 Erhebungen genutzt werden. Hinzu kommen noch weitere Erhebungen aus dem Agrar- und Umweltbereich sowie einzelne Steuerstatistiken. Die Erhebungen bei Betrieben bzw. Unternehmen² erfolgen in der Regel mit Auskunftspflicht und getrennt nach Wirtschaftsbereichen. Nachfolgend wird eine Auswahl der in den FDZ verfügbaren Firmendaten vorgestellt.³ Zunächst werden die Wirtschaftsbereiche im Einzelnen betrachtet und anschließend wirtschaftsbereichsübergreifende Mikrodaten dargestellt.

Verarbeitendes Gewerbe

Das Erhebungssystem im Bereich des Verarbeitenden Gewerbes, Bergbau sowie Gewinnung von Steinen und Erden ist vergleichsweise gut ausgebaut. Hier werden umfassende Erhebungen durchgeführt und allgemeine Betriebs- und Unternehmensmerkmale wie Umsatz, Beschäftigung oder auch Kostenstrukturen, Investitionen und die Produkte erfasst.

Der Monatsbericht beinhaltet für Betriebe Informationen zu Beschäftigten, Löhnen, Gehältern, Arbeitsstunden und Umsatz, auch untergliedert nach In- und Auslandsumsatz. Daneben werden in der vierteljährlichen Produktionserhebung die Produkte mit den entsprechenden

¹ Eine weitere Form des Datenzugangs wird derzeit über Campus Files ermöglicht. Bei diesen handelt es sich um absolut anonymisierte Daten, mit denen Studierende Erfahrungen im Umgang mit den Mikrodaten der amtlichen Statistik sammeln können.

² In der amtlichen Statistik ist ein Betrieb als die kleinste örtliche Einheit definiert. Die kleinste rechtliche Einheit stellt das Unternehmen dar. In der Regel besteht die Möglichkeit Betriebsdaten auf Unternehmensebene zusammenzuspielen. Als Firmendaten werden in diesem Beitrag die amtlichen Betriebs- und Unternehmensdaten bezeichnet.

³ Detaillierte Informationen zu den verfügbaren Mikrodaten in den FDZ sind im Internet unter www.forschungsdatenzentrum.de zu finden. Es besteht zusätzlich die Möglichkeit, sich für den Newsletter anzumelden, um so über Neuheiten oder auch Veranstaltungen informiert zu werden.

Absatzproduktionswerten auf Betriebsebene erfasst. In der Investitionserhebung hingegen sind zum Beispiel die Investitionen für erworbene und selbsterstellte Sachanlagen, wie Grundstücke enthalten. Bei den genannten Erhebungen handelt es sich um Totalerhebungen mit Abschneidegrenzen, deren Berichtskreise sich ähneln.⁴ Ergänzt wird das Datenangebot durch die Kleinbetriebserhebung der Jahre 1995 bis 2002, mit der Informationen zur Anzahl der Beschäftigten und dem Umsatz von Betrieben mit 19 Beschäftigten und weniger vorliegen. Die Mikrodaten dieser einzelnen Statistiken stehen zum einen als Querschnitt, zum anderen verknüpft über die Zeit als Längsschnittdatensatz für die Jahre 1995 bis 2005 für Auswertungen zur Verfügung. Nähere Ausführungen zu den Forschungspotentialen dieser Mikrodaten sind in Wagner (2007) zu finden.

Die Mikrodaten der Kostenstrukturerhebung (KSE) im Verarbeitenden Gewerbe stehen ebenfalls für Auswertungen zur Verfügung. Die KSE liefert umfassende Informationen über die Produktionsergebnisse, die eingesetzten Produktionsfaktoren sowie über die Wertschöpfung von Unternehmen mit mindestens 20 Beschäftigten (Statistisches Bundesamt 2005d) und ist für vielfältige Strukturuntersuchungen geeignet (Fritsch u.a. 2004). Bei dieser Erhebung handelt es sich um eine Stichprobenerhebung, die in der Regel alle vier Jahre neu gezogen wird, um die auskunftspflichtigen Unternehmen zu entlasten. Die Daten stehen sowohl für Analysen im Querschnitt als auch für Auswertungen im Längsschnitt für den Zeitraum 1995 bis 2004 bereit.

Handel und Gastgewerbe

Im Einzelhandel und Gastgewerbe werden jährlich Strukturerhebungen durchgeführt, die wirtschaftspolitisch bedeutsame Informationen über die Unternehmensstruktur sowie zur Beurteilung der Rentabilität und Produktivität der befragten Unternehmen liefern.

In beiden Wirtschaftsbereichen werden von Unternehmen neben dem Jahresumsatz, den Investitionen, dem Wareneingang und den Lagerbeständen auch die Anzahl der Beschäftigten, die Bruttolohn- und -gehaltssumme sowie die Sozialabgaben erfasst. In der Strukturerhebung im Einzelhandel wird der Gesamtumsatz zusätzlich nach Arten der ausgeübten wirtschaftlichen Tätigkeiten aufgegliedert.

⁴ Einbezogen werden in der Regel Betriebe, die einem Unternehmen des Produzierenden Gewerbes angehören, das mehr als 20 tätige Personen aufweist. Liegt der Schwerpunkt des Unternehmens außerhalb des Produzierenden Gewerbes, werden die Betriebe dann einbezogen, wenn sie mindestens 20 Personen beschäftigen. Ausnahme bilden 14 Wirtschaftszweige mit 10 Beschäftigten und mehr. Nähere Informationen sind in den aktuellen Qualitätsberichten des Statistischen Bundesamtes (2005 a,b,c) zu finden.

Aufgrund des breiten Spektrums an Erhebungsmerkmalen eignen sich diese Stichprobenerhebungen beispielsweise besonders für Analysen der Umsatz-, Investitions- und Beschäftigungsstruktur sowie von Kapital- und Beschäftigungsrentabilität.

Dienstleistungsbereich

Im Dienstleistungsbereich wird seit dem Jahr 2000 eine Strukturerhebung bei den Unternehmen und Einrichtungen, die den Schwerpunkt ihrer wirtschaftlichen Tätigkeit in den Bereichen Verkehr und Nachrichtenübermittlung, Grundstücks- und Wohnungswesen, Vermietung beweglicher Sachen sowie unternehmensnahe Dienstleistungen haben, durchgeführt.

Im Rahmen dieser Stichprobenerhebung⁵ werden neben allgemeinen Angaben der befragten Unternehmen wie Sitz oder der Anzahl der Niederlassungen auch umfassende Informationen über tätige Personen, Löhne und Gehälter, Umsätze, Aufwendungen, die Bestände am Anfang und am Ende des Berichtsjahres sowie über Steuern, Subventionen und getätigte Investitionen erfasst.

Weitere Betriebs- und Unternehmensdaten

Die im Folgenden vorgestellten Erhebungen enthalten Betriebe bzw. Unternehmen mehrerer Wirtschaftszweige.

Die Gehalts- und Lohnstrukturerhebung (GLS) wird im Allgemeinen alle vier Jahre als Stichprobe in den Betrieben des Verarbeitenden Gewerbes und ausgewählter Dienstleistungsbereiche durchgeführt. Die Erhebung enthält ausführliche Informationen zur Person, zur Tätigkeit und zum Verdienst einzelner Arbeitnehmer/innen. Auf Betriebsebene existieren zusätzlich Angaben über die Beteiligung der öffentlichen Hand am Unternehmen sowie welche Tarifverträge gelten und Angaben zur Anzahl der Beschäftigten, jeweils differenziert nach Geschlecht und Status der Arbeitnehmer/innen (Hafner/Lenz 2007).

Bei der GLS handelt es sich um einen Linked Employer-Employee-Datensatz, d. h. die Angaben zu den Beschäftigten lassen sich mit den Merkmalen des jeweiligen Betriebes verknüpfen. Die Mikrodaten eignen sich beispielsweise zur Analyse geschlechtsspezifischer Lohnunterschiede sowie zur Untersuchung der Verdienstunterschiede in tarifgebundenen Betrieben im Vergleich zu solchen, die nach freier Vereinbarung vergütet (vg. Heinbach 2007).

⁵ Kleinunternehmen mit einem jährlichen Umsatz unter 16 620 € werden nicht befragt. Angaben zu sämtlichen genannten Positionen sind für alle Unternehmen mit einem jährlichen Umsatz von mehr als 250 000 € enthalten (vgl. Statistisches Bundesamt 2005e).

Als weitere wichtige Erhebung im Bereich der Betriebs- und Unternehmensdaten lässt sich die Arbeitskostenerhebung nennen. Diese Stichprobenerhebung wird alle vier Jahre in Unternehmen des Produzierenden Gewerbes und in ausgewählten Dienstleistungsbranchen durchgeführt (Statistisches Bundesamt 2007). Die Arbeitskostenerhebung gibt Aufschluss über die Höhe und Struktur der Kosten, welche den Unternehmen bei der Beschäftigung von Arbeitnehmern entstehen. So sind neben den Löhnen und Gehältern insgesamt u. a. Informationen über Sonderzahlungen, Vergütung für Feiertage, Pflichtbeiträge zur Rentenversicherung, Beiträge zur Berufsgenossenschaft, Pensionsrückstellungen und Familienunterstützungen enthalten.

Die Arbeitskostenerhebung eignet sich beispielsweise dazu, den Anteil der Lohnnebenkosten und der zusätzlichen Leistungen der Arbeitgeber differenziert nach Branchen zu untersuchen.

Steuerstatistiken

Eine der umfangreichsten Wirtschaftsstatistiken bezüglich der Anzahl der Untersuchungseinheiten ist die Umsatzsteuerstatistik. Sie beinhaltet alle Unternehmen, die im Statistikjahr eine Umsatzsteuer-Voranmeldung abgegeben haben und deren jährliche Lieferungen und Leistungen über 17 500 Euro betragen. An Merkmalen umfasst die Umsatzsteuerstatistik Informationen zu allen steuerbaren Umsätzen, Umsatzsteuer, Vorsteuer und Dauer der Steuerpflicht. Sie ist ebenfalls für Auswertungen im Quer- und Längsschnitt für den Zeitraum 2000 bis 2004 verfügbar.

3.2 Betriebs- und Unternehmensdaten des FDZ der BA im IAB

Das FDZ der BA im IAB bietet derzeit (Stand: 09/2007) Zugang zu drei unterschiedlichen Betriebsdatensätzen.⁶ Alle drei Datensätze können zum aktuellen Zeitpunkt lediglich über Gastaufenthalte und Datenfernverarbeitung genutzt werden, bei denen projektspezifische Datenanonymisierungen vorgenommen werden.

Das Betriebs-Historik-Panel (BHP) ist ein relativ neuer Datensatz des IAB und nutzt die Beschäftigtenmeldungen der Bundesagentur für Arbeit. Dabei werden Aggregate auf Grundlage der Betriebsnummer zum 30.06 eines Jahres berechnet, so dass betriebspezifische Querschnittsdatsätze entstehen (vgl. Dundler et al. 2006). Diese umfassen Betriebe, die zur Jahresmitte mindestens eine gültige Meldung für ein sozialversicherungspflichtiges Beschäftigungsverhältnis aufweisen (ab 1999 betrifft dies auch Betriebe mit geringfügig Beschäftigten).

⁶ Eine aktuelle Übersicht kann den Internetseiten des FDZ der BA im IAB (<http://fdz.iab.de>) entnommen werden. Aufgeführt sind auch bisherige Forschungsarbeiten mit den Datensätzen.

Die einzelnen Querschnittsdatensätze (aktuell von 1975 bis 2003) können zu einem Panel verbunden werden. Auf diese Weise ermöglicht der Datensatz differenzierte Analysen zur Entwicklung der betrieblichen Personalstruktur und betrieblichen Lohnverteilungen.

Die Besonderheit des BHP liegt darin, dass es Betriebsmerkmale auf Basis von aggregierten Beschäftigtenmeldungen enthält und zwar für alle Betriebe im deutschen Raum. Es enthält neben den Informationen zum Wirtschaftszweig und dem Standort, Angaben über die Anzahl der Beschäftigten insgesamt als auch aufgeteilt nach unterschiedlichen Teilgruppen (z.B. differenziert nach Qualifikation). Darüber hinaus sind die Quartile des Alters der Beschäftigten eines Betriebes und des Tagesentgelts der Vollzeitbeschäftigten sowie der Beschäftigten insgesamt ausgewiesen.⁷

Beim IAB-Betriebspanel handelt es sich dagegen um einen bereits stark bekannten und häufig genutzten Datensatz basierend auf einer repräsentativen Arbeitgeberbefragung.⁸ Im Unterschied zum BHP steht über das IAB-Betriebspanel eine große Bandbreite an betrieblichen Merkmalen zur Beschäftigungsentwicklung und zur Betriebsstruktur zur Verfügung (z.B. betriebliche Arbeitszeiten, Umsatz, Exportanteil, betriebliche Investitionen, Innovationen sowie Berufsausbildung sowie Fort- und Weiterbildung). Eine Besonderheit des Datensatzes ist darüber hinaus, dass Abgleiche mit Planungs- und Istgrößen vorgenommen werden können. Durch das Panel-Design lassen sich diese Merkmale auch in ihren dynamischen Abläufen untersuchen. Zugang zum IAB-Betriebspanel ist über Datenfernverarbeitung und Gastaufenthalte möglich. Ein faktisch anonymer Datensatz soll durch das Projekt „Wirtschaftsstatistische Paneldaten und faktische Anonymisierung“ (vgl. nächstes Kapitel) entstehen.

Die Linked-Employer-Employee-Datensätze (LIAB) des IAB verknüpfen die Daten des Betriebspanels mit den längsschnittaufbereiteten Meldungen sozialversicherungspflichtig Beschäftigter. Dieser steht der Forschung seit 2005 für das FDZ der BA am IAB zur Verfügung.

Durch die Verknüpfung von Betriebs- und Beschäftigteninformationen ermöglicht der Datensatz u.a. die simultane Analyse der Angebots- und Nachfrageseite des Arbeitsmarktes, d.h. Analysen zur Interaktion von Betrieben und Beschäftigten. Beispielsweise können Erwerbsverläufe direkt im Kontext betrieblicher Variablen analysiert werden.

Derzeit stehen die LIAB-Daten der Forschung in zwei unterschiedlichen Datenmodellen zur Verfügung: dem LIAB-Querschnittsmodell und dem LIAB-Längsschnittmodell. Im ersten Modell werden Personendaten, die zum 30.06 eines Jahres gültig sind, mit den Daten des IAB-Betriebspanels verknüpft. Damit liegen für alle Betriebe des Panels zusätzliche Informationen zu den in den Betrieben beschäftigten Personen vor. In dem zweiten Datenmodell werden da-

⁷ Derzeit kann das BHP über einen Gastaufenthalt am FDZ der BA am IAB genutzt werden

gegen die Betriebsdaten mit den Personeninformationen verknüpft. Ausgangspunkt ist in diesem Fall eine Teilstichprobe der Betriebe im IAB-Betriebspanel, zu der alle Personen identifiziert werden, die in einem dieser Betriebe beschäftigt waren. Maßgeblich wird diese Variante des LIAB für Analysen zu Lohndeterminanten und Lohndifferentialen verwendet. Eine differenzierte Darstellung der Datenmodelle kann Alda et al. (2005) und Jacobebbinghaus (2007) entnommen werden. Der LIAB kann in der Erstnutzung mittels Gastaufenthalt und einer späteren Datenfernverarbeitung genutzt werden.

4. Verbesserung des Datenangebots durch das Projekt „Wirtschaftsstatistische Paneldaten und faktische Anonymisierung“

Mit dem Projekt „Wirtschaftsstatistische Paneldaten und faktische Anonymisierung“ (FAWE-PANEL) soll die Dateninfrastruktur im Bereich der Betriebs- und Unternehmensdaten im Längsschnitt und der Zugang der Wissenschaft zu Wirtschaftsstatistiken der amtlichen Statistik des Bundes und der Länder, sowie zu den Daten der Bundesagentur für Arbeit deutlich verbessert werden.⁹ Gegenstand des Projektes ist der Ausbau des Datenangebotes durch Längsschnittverknüpfungen von Statistiken, die bisher überwiegend im Querschnitt genutzt worden sind. Im Vordergrund stehen hierbei die Kostenstrukturerhebung, die Umsatzsteuerstatistik und der Monatsbericht im Verarbeitenden Gewerbe, die im Rahmen des Projektes als Längsschnittdatensätze aufbereitet wurden und für die wissenschaftliche Forschung bereits verfügbar sind. Ein weiterer wichtiger Bestandteil des Projektes ist die Ergänzung der längsschnittverknüpften Daten mit Angaben aus dem Unternehmensregister (Konold 2007). Diese Arbeiten zielen insbesondere darauf, anhand des Unternehmensregisters Gründe für fehlende Angaben im Längsschnitt zu benennen und hierdurch das Analysepotential der Daten zu erhöhen. So kann zum Beispiel auf der Basis des Unternehmensregisters nachvollzogen werden, ob eine Berichtseinheit aufgrund eines Wirtschaftszweigwechsels, einer Änderung der Beschäftigtenzahl oder einer Schließung nicht mehr in die Erhebung einbezogen wurde. Bei der Umsatzsteuerstatistik konnten zudem die Umsatzinformationen mit Hilfe des Unternehmensregisters durch Beschäftigtenangaben für die Jahre 2000 bis 2004 ergänzt werden.

Des Weiteren werden im Rahmen des Projektes Anonymisierungsmethoden für Paneldaten entwickelt mit dem Ziel, die Daten als Scientific Use File anbieten zu können. In diesem Zusammenhang wird neben dem Datenschutz das Analysepotential der Daten untersucht und

⁸ Genauere Angaben zum Aufbau des IAB-Betriebspanels finden sich z.B. in Bellmann (2002).

⁹ Das Projekt wird gemeinsam vom Institut für Arbeitsmarkt- und Berufsforschung, dem Institut für Angewandte Wirtschaftsforschung (IAW), dem FDZ der Statistischen Landesämter mit den Standorten Berlin, Düsseldorf und Wiesbaden sowie dem FDZ des Statistischen Bundesamtes durchgeführt.

sichergestellt, dass durch die Anonymisierungsverfahren keine gravierenden Einschränkungen des Analysepotentials auftreten.

4.1 Anonymisierung von Wirtschaftsstatistiken

Die Anonymisierung von Wirtschaftsdaten ist in vielerlei Hinsicht eine Herausforderung, da zusätzliche Informationen für Unternehmen zahlreich vorhanden sind und die Grundgesamtheiten von Betriebs- und Unternehmensdaten kleiner sind als bei Personendaten. Teilpopulationen weisen daher oftmals nur eine geringe Anzahl an Untersuchungseinheiten auf. Zudem treten bei Unternehmensdaten häufig Konzentrationen auf, so dass ein Großteil des Umsatzes nur auf wenige Unternehmen entfällt oder dass einzelne Unternehmen einzigartige Fälle bilden. Der Nutzen, der aus einer erfolgreichen Reidentifizierung eines Unternehmens resultieren kann, ist darüber hinaus im Vergleich zu personenbezogenen Daten erheblich größer, da z.B. Kenntnisse über aktuelle Zahlen einem Konkurrenzunternehmen Wettbewerbsvorteile liefern könnten.

Die Grundlagen für die Anonymisierung wirtschaftsstatistischer Einzeldaten wurden bereits in dem Projekt „Faktische Anonymisierung wirtschaftsstatistischer Einzeldaten“ (FAWE) erarbeitet. Zu den zentralen Arbeiten gehörte hier die Operationalisierung faktischer Anonymität (Statistisches Bundesamt 2005f). Dabei stellte sich heraus, dass die Sicherstellung der faktischen Anonymität bei wirtschaftsstatistischen Einzeldaten in der Regel nur unter Verwendung so genannter datenverändernder Anonymisierungsmethoden möglich ist, bei denen die Merkmalswerte systematisch oder stochastisch verändert werden. Im Rahmen des Projektes „FAWE“ wurden deshalb die verschiedenen Methoden zur Anonymisierung von Wirtschaftsdaten systematisiert, weiterentwickelt und hinsichtlich ihrer Auswirkungen auf die Datensicherheit und das Analysepotenzial untersucht und bewertet (Rosemann 2006).

Im Ergebnis konnten erstmals wirtschaftsstatistische Daten so anonymisiert werden, dass sie einerseits den strengen Anforderungen des Datenschutzes gerecht werden und andererseits ausreichend Potential für wissenschaftliche Analysen bieten (siehe Lenz et al. 2006).¹⁰

¹⁰ Auf der Basis dieser Arbeiten wurden bisher Scientific Use Files der folgenden Statistiken bereitgestellt: KSE 1999, Umsatzsteuerstatistik 2000, Einzelhandelsstatistik 1999 und als Ergebnis eines einzelnen Anonymisierungsprojektes die GLS.

4.2 Herausforderungen bei der Anonymisierung Wirtschaftsstatistischer Längsschnittdaten

Die bisher erstellten Scientific Use Files beschränken sich auf Querschnittsdaten. Allerdings werden insbesondere Längsschnitt- und Paneldaten von Seiten der Wissenschaft verstärkt nachgefragt, da nur mit diesen die Dynamik, Veränderung und Prozessverläufe über die Zeit abzubilden sind. Ein weiterer Vorteil von Paneldaten besteht darin, dass unbeobachtbare Heterogenität kontrolliert und dieses entsprechend für die Analysen berücksichtigt werden kann. Die positiven Aspekte, die Längsschnittdaten für die Auswertungen von Forschungsfragen mit sich bringen, könnten sich jedoch gleichzeitig als zusätzliche Herausforderungen bei der Anonymisierung erweisen, da über mehrere Wellen eine Struktur in den Daten sichtbar werden kann, die einem potentiellen Datenangreifer Zusatzwissen bietet und ihm somit für einen Reidentifikationsversuch hilfreich sein kann (siehe Lenz 2006).

Mit Blick auf die Erhaltung des Analysepotenzials von Paneldaten muss darauf geachtet werden, dass zeitliche Entwicklungen auch mit Hilfe anonymisierter Daten adäquat untersucht werden können und panelökonometrische Verfahren weiterhin zu konsistenten Schätzungen führen (Biewen/Ronning/Rosemann 2007 (IAW-Report; Biewen 2007 (in diesem Heft)).

Aktuell beschäftigt sich das Projekt daher mit der Frage, inwiefern die für Querschnittsdaten entwickelten Anonymisierungsverfahren für die Anonymisierung von Paneldaten weiterentwickelt werden müssen (vgl. Beitrag von Höhne in diesem Heft) und welche Auswirkungen diese Verfahren auf den Datenschutz und das Analysepotenzial der wirtschaftsstatistischen Paneldaten haben (vgl. hierzu Beiträge von Lenz und Biewen in diesem Heft).

Am Ende des Projekts sollen für die Datensätze der Kostenstrukturerhebung im Verarbeitenden Gewerbe, für den Monatsbericht, die Investitions- sowie die Kleinbetriebserhebung im Verarbeitenden Gewerbe und die Umsatzsteuerstatistik faktisch anonymisierte Paneldatensätze zur Verfügung gestellt werden.

5. Ausblick auf weitere Vorhaben

Viele der amtlichen Mikrodaten aus den Bereichen der Wirtschaftstatistiken verfügen wie bereits erwähnt über einen geringen Merkmalsumfang. So sind in dem oben genannten Monatsbericht oder dem IAB-Betriebspanel lediglich die Informationen zu den Beschäftigten oder dem Umsatz enthalten. Die FDZ wollen mit neuen Projekten das Informationspotential der amtlichen Firmendaten über Verknüpfungen der vorhandenen Betriebs- und Unternehmensdaten erhöhen – auch über die Grenzen der Datenproduzenten hinweg. Das Zusammenführen

von Betriebs- und Unternehmensdaten ist für wissenschaftliche Forschungsprojekte von besonderem Interesse, da somit nicht nur weitere Informationen für die einzelnen Betriebe bzw. Unternehmen vorliegen (beispielsweise zu Investitionen, Angaben aus Umweltstatistiken etc.), sondern auch über Wirtschaftsbereiche hinweg, d.h. Wirtschaftsbereichswechsler können identifiziert oder Analysen für alle Unternehmen in Deutschland durchgeführt werden.

Die FDZ der Statistischen Ämter des Bundes und der Länder wollen zukünftig mit dem Projekt der „Integration von amtlichen wirtschafts- und umweltstatistischen Daten unter Berücksichtigung der Zeitdimension“ ihr Dienstleistungsangebot weiterentwickeln. Dieses Projekt unterteilt sich in zwei Teilprojekte: „AFiD – Amtliche Firmendaten für Deutschland“ und „KombiFiD – Kombinierte Firmendaten für Deutschland“, die nachfolgend näher erläutert werden.

5.1 AFiD – Amtliche Firmendaten für Deutschland

Mit der Novellierung des Bundesstatistikgesetzes (BstatG) im Jahre 2005 besteht nunmehr die Möglichkeit, Informationen aus verschiedenen Erhebungen der Wirtschafts- und Umweltstatistiken zusammenzuführen (siehe §13a BStatG). Die Zusammenführung der Betriebs- und Unternehmensdaten der amtlichen Statistik erfolgt im Teilprojekt AFiD, als eine der Aufgaben des FDZ der Statistischen Ämter der Länder. Im Vorfeld des Teilprojektes wurden bereits in einer ersten Untersuchungsphase Erfahrungen mit der Verknüpfung wirtschaftsstatistischer Daten gesammelt. Für einzelne Forschungsprojekte wurden beispielsweise die Kostenstruktur und die Produktionserhebung zusammengeführt (Görzig u.a. 2005). Neben den Statistiken aus dem Verarbeitenden Gewerbe bieten sich eine Reihe von weiteren Erhebungen aus den unterschiedlichen Wirtschaftsbereichen zur Integration an, so beispielsweise aus dem Dienstleistungsbereich, dem Baugewerbe, der Landwirtschaft oder dem Umweltbereich. Des Weiteren können auch Informationen aus den Steuerstatistiken für die Zusammenführung genutzt werden.

Die Auswahlgrundlage der betroffenen Erhebungen wird bis auf wenige Ausnahmen vom Unternehmensregister (URS)¹¹ gebildet. Dieses existiert seit 1998 und wurde durch die EU-Registerverordnung initiiert. Seit 2001 liegt es in einer Form vor, die es ermöglicht Datensätze auf Grundlage eines zentralen Unternehmensidentifikators valide miteinander zu verbinden (vgl. Sturm/Tümmler, 2006). Ziel des Projektes AFiD ist eine Zusammenführung des URS mit den Mikrodaten ausgewählter Erhebungen. Eine Auswertung des Unternehmensregisters im Rahmen der Belastungsstudie der Unternehmen (vgl. Stäglich/Pfeiffer 2006) zeigte, dass von den rund 3,5 Millionen Unternehmen im Jahr 2004 nur 15,2% berichtspflichtig waren. Für diese

¹¹ Vgl. Verordnung Nr 2186/93 (EU-Register-Verordnung)

berichtspflichtigen Unternehmen sollen die firmenspezifischen Informationen aus den gewählten Erhebungen an die bisher vorhandenen Informationen im URS angefügt werden. Das Produzierende Gewerbe stellt dabei die höchsten Anteile der berichtspflichtigen Unternehmen, wobei sich dieser zum einen durch die Historie des URS erklärt und zum anderen damit begründet, dass in diesen Wirtschaftsbereichen Vollerhebungen mit Abschneidegrenze durchgeführt werden. Die weniger hohen Anteile in den Bereichen Handel und Gastgewerbe ergeben sich dagegen durch die dort durchgeführten Stichprobenerhebungen. Ebenfalls zum Handel- und Gastgewerbe weist auch der Dienstleistungsbereich tendenziell höhere Anteile meldepflichtiger Unternehmen aus, die aus der jungen Dienstleistungsstatistik resultieren.

Die Zusammenführung der Mikrodaten im Projekt AFiD soll schrittweise erfolgen. Zunächst werden für einzelne Wirtschaftsbereiche sowie auch bereichsübergreifend Teilprodukte erstellt. Ein Beispiel für ein solches Teilprodukt wäre ein Betriebs- oder Unternehmensdatensatz für das Verarbeitende Gewerbe im Längsschnitt, der alle Informationen aus den entsprechenden Erhebungen (Monatsbericht, Investitionserhebung, Produktionserhebung, Kleinbetriebserhebung)¹² verknüpft über die Zeit enthält. Über derartige Verknüpfungen entstehen umfassende Datenbestände, deren Analysepotenzial noch zusätzlich erweitert werden kann. So besteht dann zukünftig die Möglichkeit, die Firmendaten nicht nur hinsichtlich demographischer Entwicklungen und diese im Zusammenhang mit Kostengrößen zu untersuchen, sondern auch im Zusammenhang mit Merkmalen aus dem Umweltbereich, wie beispielsweise dem Wasseraufkommen.

Je komplexer der Datenbestand ist, desto vielfältiger sind jedoch die methodischen Anforderungen, die bei der Zusammenführung der Daten beachtet und im Projekt bearbeitet werden müssen. Hierzu gehören unter anderem die unterschiedlichen Erhebungsdesigns in den Statistiken oder das (Nicht-)Vorhandensein einheitlicher Identifikationsnummern. Ferner soll der Informationsgehalt der einzelnen Analyseeinheiten, wie Unternehmen, Betriebe oder Produkte erhalten bleiben. Darüber hinaus ergeben sich zusätzliche Anforderungen an die fachliche Erarbeitung der Metadaten, da den Wissenschaftlern sowohl Informationen über die Zusammenführung der Daten und die damit verbundenen Selektionsprozesse als auch über die Anwendung von Gewichtungsverfahren zur Verfügung gestellt werden müssen.

¹² Insofern kann hier für das verarbeitende Gewerbe auf die Zwischenergebnisse des Projekts FAWE-Panel aufgebaut werden.

5.2 KombiFiD – Kombinierte Firmendaten für Deutschland

Aufgrund gesetzlicher Bestimmungen können in Deutschland bislang nur solche Daten verknüpft werden, die einer einheitlichen gesetzlichen Grundlage unterliegen. So ist es derzeit nicht möglich, Daten der Bundesagentur für Arbeit mit Daten der Statistischen Ämter des Bundes und der Länder zu kombinieren. Mit der Machbarkeitsstudie KombiFiD soll diese Lücke geschlossen werden, indem erstmals Unternehmensdatensätze unterschiedlicher Datenproduzenten zusammengeführt werden. Das Zusammenführen dieser Daten lässt detailliertere und umfassendere Untersuchungen wirtschaftlicher Prozesse zu als dies mit dem bislang vorhandenen Datenmaterial möglich ist. Im Rahmen von KombiFiD soll zudem eine Gesetzesänderung erwirkt werden, die ein regelmäßiges Zusammenspielen der Daten erlaubt. Dadurch könnte eine erhebliche Entlastung der Unternehmen bei der statistischen Auskunftspflicht erreicht werden. Realisiert wird das Projekt durch das FDZ der BA im IAB, das FDZ des Statistischen Bundesamtes sowie den Lehrstuhl für „Empirische Wirtschaftsforschung“ an der Leuphana Universität Lüneburg.

Ausgangspunkt des Zusammenführens der unterschiedlichen Datensätze ist das URS. Zu den zentralen Datensätzen für ein Zusammenführen gehören unter anderem die KSE, der Monatsbericht des Verarbeitenden Gewerbes und die Gehalts- und Lohnstrukturerhebung aus dem Projekt AFiD sowie das Betriebs-Historik-Panel.¹³ Ausgehend von den Daten der Stichprobenerhebungen sollen dann Ergänzungen mit weiteren über das URS verknüpfbaren Datensätzen vorgenommen werden. Maßgabe des Projektes ist es zunächst ein Höchstmaß an Informationen über die verknüpften Daten zusammenzutragen. Durch das Verbinden von Beschäftigtenmeldungen des FDZ der BA im IAB mit den Unternehmensdaten der statistischen Ämter wird es somit möglich, umfangreiche neue Betriebs- und Unternehmensdaten im Längsschnitt für die Forschung anzubieten. Erstmals können dadurch z.B. Beschäftigungseffekte von Veränderungen der Unternehmensbesteuerung auf Mikroebene simultan untersucht werden.

Um die rechtliche Voraussetzung für ein Zusammenspielen zu ermöglichen, werden ausgewählte Betriebe im Rahmen einer Befragung angeschrieben und um eine schriftliche Einwilligung für das Zusammenführen gebeten.¹⁴ Auswahlgrundlage sind Betriebe, die zwischen 1995 und 2003 Beschäftigtenmeldungen abgegeben haben. Die Zusammenführung der Mikrodaten des URSs mit denen des IAB erfolgt nur für diejenigen Unternehmen, die einer sol-

¹³ Die Auswahl der Datensätze soll hierbei folgende Kriterien erfüllen: Zum einen sollen die Daten möglichst viele inhaltliche Auswertungen zulassen und zum anderen sollten die Daten als Panel vorliegen.

chen Verknüpfung schriftlich zugestimmt haben. Nach Festlegung der aus wissenschaftlicher Sicht wichtigen Datensätze und Merkmale werden Konsistenz- und Validierungsanalysen durchgeführt. Diese sollen Aufschluss darüber geben, welche Art des Zusammenführens zu inhaltlich sinnvollen Ergebnissen führt. Geplant ist, zunächst nur ausgewählte Daten zur Verfügung zu stellen. Der Zugang zu den Daten soll über Gastaufenthalte an den beteiligten FDZ und – wenn datenschutzrechtlich möglich – über Scientific Use Files erfolgen. Zu den wichtigen Arbeiten in dieser Projektphase gehören daher die Anonymisierung der zusammengeführten Daten sowie die Dokumentation der Daten. Darüber hinaus soll mit der zweiten Phase des Projektes eine Gesetzesänderung für ein regelmäßiges Zusammenspielen der Daten initiiert werden.

6. Schlussbetrachtung

In den letzten Jahren haben sich erhebliche Verbesserungen im Datenzugang, Datenprodukten, Informationssystemen und im Wissen um die Grenzen des Analysepotentials amtlicher Daten ergeben. Mit den FDZ existiert bereits jetzt eine umfangreiche Infrastruktur um Mikrodaten für Forschungsprojekte nutzen zu können. Der Zugang zu diesen Mikrodaten ist durch die FDZ möglich, die sich als Schnittstelle zwischen den Datenproduzenten und der empirisch arbeitenden Wissenschaft verstehen. Gleichzeitig wurden zahlreiche Projekte zur Anonymisierung einzelner Datenquellen durchgeführt. Durch das Projekt „Faktische Anonymisierung wirtschaftsstatistischer Einzeldaten“ wurden die Grundlagen zur Erstellung von Scientific Use Files im Bereich der Betriebs- und Unternehmensdaten gelegt. Dadurch wurden erstmals in Deutschland faktisch anonymisierte Querschnittsdaten über Unternehmen und Betriebe der Wissenschaft als Scientific Use Files zur Verfügung gestellt. Durch das Projekt „Wirtschaftsstatistische Paneldaten und faktische Anonymisierung“ wurde das Angebot der FDZ im Bereich der Paneldaten über Unternehmen und Betriebe erweitert. Im Rahmen des Projekts werden aktuell die Möglichkeiten der Anonymisierung dieser Daten untersucht. In einem nächsten Schritt sollen auch sie als Scientific Use Files verfügbar gemacht werden.

In näherer Zukunft werden sich die FDZ gemeinsam mit der Zusammenführung von Informationen aus unterschiedlichen Datenquellen in den Projekten „AFiD – Amtliche Firmendaten für Deutschland“ und „KombiFiD – Kombinierte Firmendaten für Deutschland“ befassen. Damit könnten die Analysemöglichkeiten amtlicher Daten sehr stark ausgeweitet werden.

¹⁴ Bei einer schriftlichen Einwilligung der Unternehmen/Betriebe können wirtschaftsstatistische Einzeldaten für ein zeitlich befristetes Projekt auch über die Grenze des einzelnen Datenproduzenten hinaus zusammengeführt werden.

Literatur

- Alda, H. / Bender, S. / Gartner, H. (2005): The linked employer-employee dataset created from the IAB establishment panel and the process-produced data of the IAB (LIAB). In: Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften, Jg. 125, H. 2
- Bellmann, L. (2002): Das IAB-Betriebspanel. Konzeption und Anwendungsbereiche. In Allgemeines Statistisches Archiv, Bd. 86, H. 2, S. 177-188.
- Dundler, A. / Stamm, M. / Adler, S.(2006): Das Betriebs-Historik-Panel BHP 1.0.Handbuch-Version 1.0.0. Nürnberg.
- Biewen, E., Ronning, G. und Rosemann, M. (2007): Estimation of Linear Panel Models with Anonymised Business Data. In: IAW-Report 1/2007, S. 87-114
- Fritsch, M., Görzig, B., Hennchen, O. und Stephan, A. (2004): Cost Structure Surveys in Germany, Schmollers Jahrbuch / Journal of Applied Social Science Studies 124, 557-566.
- Görzig, B., Bömermann, H. und Pohl, R. (2005): Produktdiversifizierung und Unternehmenserfolg: Nutzung der Forschungsdatenzentren der Statistischen Ämter. Allgemeines statistisches Archiv 89: 337-54.
- Hafner, H.-P., Lenz, R. (2007): Die Gehalts- und Lohnstrukturerhebung: Methodik, Datenzugang und Forschungspotential, FDZ-Arbeitspapier Nr. 18.
- Heinbach, W.D. (2007): Wages in Wage-Setting Regimes with Opening Clauses. In: AStA - Wirtschafts- und Sozialstatistisches Archiv, 1(3), im Druck.
- Jacobebbinghaus, P./ Alda, H. (2007): LIAB-Datenhandbuch, Version 2.0. (FDZ Datenreport, 02/2007), Nürnberg.
- Konold, M. (2007): New possibilities for economic research through integration of establishment-level panel data of German official statistics, Schmollers Jahrbuch / Journal of Applied Social Science Studies 127 (in press).
- Lenz, R. (2006): Measuring the anonymity of longitudinal linked economic statistics microdata. Methods - Approaches - Developments 2, pp. 3-6.
- Lenz, R., Rosemann, M., Vorgrimler, D. und Sturm, R. (2006): Anonymising business micro data - results of a German project. Journal of Applied Social Science Studies (Schmollers Jahrbuch) 126 (4), pp. 635-651.
- Rosemann, M. (2006): Auswirkungen datenverändernder Anonymisierungsverfahren auf die Analyse von Mikrodaten. IAW-Forschungsbericht, Nr. 66, Tübingen.Rat für Sozial- und Wirtschaftsdaten (2006): Eine moderne Dateninfrastruktur für exzellente Forschung und
-

Politikberatung. Bericht über die Arbeit des Rates für Sozial- und Wirtschaftsdaten (RatSWD) während der ersten Berufsperiode vom 31.12.2006, Frontenhausen

Stäglin, R., Pfeiffer, I. (2006): Die Bedeutung der Belastung der Wirtschaft durch amtliche Statistiken, *Wirtschaft und Statistik*, Heft 11/2006.

Statistisches Bundesamt (2005a): Monatsbericht für Betriebe des Verarbeitenden Gewerbes sowie des Bergbaus und der Gewinnung von Steinen und Erden, Qualitätsbericht.

Statistisches Bundesamt (2005b): Produktionserhebungen, Qualitätsbericht.

Statistisches Bundesamt (2005c): Investitionserhebung bei Unternehmen und Betrieben des Verarbeitenden Gewerbes sowie des Bergbaus und der Gewinnung von Steinen und Erden, Qualitätsbericht.

Statistisches Bundesamt (2005d). Kostenstrukturerhebung im Verarbeitenden Gewerbe, im Bergbau sowie in der Gewinnung von Steinen und Erden, Qualitätsbericht.

Statistisches Bundesamt (2005e): Jährliche Strukturerhebung im Dienstleistungsbereich, Qualitätsbericht.

Statistisches Bundesamt (2005f): Statistik und Wissenschaft, Handbuch zur Anonymisierung wirtschaftsstatistischer Mikrodaten, Band 4.

Statistisches Bundesamt (2007): Arbeitskostenerhebung 2004, Qualitätsbericht.

Sturm, R. / Tümmler, T. (2006): Das statistische Unternehmensregister - Entwicklungsstand und Perspektiven. In: *Wirtschaft und Statistik* 10/2006, S. 1021-1036

Verordnung (EWG) Nr. 2186/93 des Rates vom 22. Juli 1993 über die innergemeinschaftliche Koordinierung des Aufbaus von Unternehmensregistern für statistische Verwendungszwecke

Wagner, J. (2007): Die Forschungspotentiale der Betriebspaneldaten des Monatsberichtes im Verarbeitenden Gewerbe, Working Paper, University of Universität Lüneburg No. 51.

Zühlke S., Zwick, M., Scharnhorst S., Wende, T. (2003): Die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder , in: *Wirtschaft und Statistik*, Heft 10/2003.

Imprint**FDZ *Methodenreport***

No. 7/2007

Publisher

The Research Data Centre (FDZ)
of the Federal Employment Service
in the Institute for Employment Research
Regensburger Str. 104
D-90478 Nuremberg

Editorial staff

Stefan Bender, Dagmar Herrlinger

Technical production

Dagmar Herrlinger

Copyright

Reproduction – also in parts – only with permission of the FDZ

Downloadhttp://doku.iab.de/fdz/reporte/2007/MR_07-07.pdf**Internet**<http://fdz.iab.de/>**Corresponding author**

Dirk Oberschachtsiek, Institute for Employment Research,
Regensburger Str. 104, D-90478 Nuremberg
Phone: +49-(0)911/179-1615
E-Mail: dirk.oberschachtsiek@iab.de