

Datenverfügbarkeit und Datenzugang am Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung

VON JUTTA ALLMENDINGER UND ANNETTE KOHLMANN

ZUSAMMENFASSUNG: Die Bundesagentur für Arbeit (BA) gehört zu den wichtigsten Produzenten statistischer Daten über den Arbeitsmarkt und die wirtschaftliche Entwicklung in Deutschland. Bislang bestand kein institutionalisierter Datenzugangsweg für Forschende außerhalb des Instituts für Arbeitsmarkt- und Berufsforschung (IAB). Das Forschungsdatenzentrum (FDZ) der BA im IAB wurde mit dem Ziel eingerichtet, den Zugang zu den Daten der BA und des IAB für externe Forscher zu erschließen und systematisch auszubauen. In dem Beitrag beschreiben wir die im FDZ für die Fachöffentlichkeit zugänglich gemachten Daten der BA und des IAB sowie die verschiedenen Zugangsmöglichkeiten der wissenschaftlichen Forschung zu diesen Daten.

SUMMARY: The Federal Agency of Labour (BA) is one of the most important producers of official data on the labour market and the economic development in Germany. Until recently, there has been no institutionalized way of data access for researchers outside the Institute of Employment Research (IAB). The research data centre (FDZ) of the BA in the IAB has been established in order to develop and systematically extend the data access for scientists. In this paper we describe the data of the BA and the IAB which will be made accessible via the FDZ as well as the different modes of access to those data.

KEYWORDS: Labor market data, official microdata, data access, research data centre. JEL C80, J00, K00.

1. EINLEITUNG

Schon lange besteht der Wunsch, externen Wissenschaftlern¹ Zugang zu den amtlichen Daten der Bundesagentur für Arbeit (BA) und des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) zu verschaffen. Das IAB war nicht untätig: Über Scientific Use Files und Möglichkeiten der Datenfernverarbeitung² wurde ein Datenzugang ermöglicht, welcher allerdings aufgrund von Finanzierungsproblemen unvollständig und unsystematisch bleiben musste (Bender, Haas und Klose, 1999). Auch datenschutzrechtliche Gründe haben viele Zugangswünsche vereitelt oder führten nach der notwendigen Anonymisierung zu Daten, die ihren Wert für die Forschung weitgehend verloren hatten, weil diese Verfahren mit zu hohen Informationsverlusten verbunden waren.

Als die Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI) im Jahre 2001 den amtlichen Datenproduzenten empfahl, Forschungsdatenzentren einzurichten (Kommis-

Eingereicht: 02.11.2004 / Revision: 28.02.2005

¹ Im Folgenden wird für Personenangaben aus Gründen der vereinfachten Darstellung ausschließlich die männliche Form verwendet. Selbstverständlich beziehen sich alle getroffenen Aussagen ebenso auf Frauen.

² Zu einer Beschreibung dieser Datenzugangswege vgl. Abschnitt 3.

sion zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI), 2001, S. 247ff.) und die Bundesministerien eine Anschubfinanzierung zur Verfügung stellten, änderte sich die Situation grundlegend. Das IAB konnte nun handeln, die Grundzüge eines Forschungsdatenzentrums (FDZ) konzipieren und den Finanzierungsantrag einreichen. Das FDZ der BA wurde aus datenschutzrechtlichen Gründen als besondere Stabsstelle des IAB beantragt und direkt der Leiterin des IAB unterstellt.

Der Antrag auf Anschubfinanzierung war erfolgreich und das Pilotprojekt ‚Forschungsdatenzentrum der BA‘ wird seit April 2004 aus Mitteln des Bundesministeriums für Bildung und Forschung (BMBF) bis Ende November 2006 gefördert, ergänzt durch Eigenmittel der BA.

Das zentrale Ziel des FDZ, die Gewährung und Erleichterung des Datenzugangs für externe Forscher, umfasst die folgenden Aufgaben:

- die Entwicklung transparenter und standardisierter Zugangsregelungen bei Gleichbehandlung von Anfragen externer Forscher unter Wahrung der geltenden Datenschutzregelungen,
- die Durchführung von Datenaufbereitungen, -aktualisierungen und -prüfungen,
- die genaue und umfassende Dokumentation der Daten,³ wobei auch die rechtlichen Aspekte der Datensicherheit zu berücksichtigen sind sowie
- die Durchführung individueller Beratungsleistungen für Forscher.

Informationen zu den Daten, zu den Zugangswegen zu den Daten und zu ihrer Dokumentation im FDZ sowie zu Veranstaltungen sind auf den Internetseiten des FDZ unter <http://fdz.iab.de> erhältlich. Darüber hinaus sind individuelle Beratungen (Kontakt unter e-mail: iab.fdz@iab.de) möglich.

Im folgenden Abschnitt 2 geben wir einen Überblick darüber, welche Datenbasen in der BA und im IAB vorliegen, welche Datenprodukte daraus für die wissenschaftliche Nutzung erstellt werden und vom FDZ für die Nutzung durch Externe betreut werden.

Dies beinhaltet die Angabe der jeweiligen Population, des erfassten Zeitraums und der Merkmale des jeweiligen Datensatzes sowie weiterführende Hinweise zu Publikationen und Analysemöglichkeiten mit den Daten. In Abschnitt 3 wird zunächst der gesetzliche Hintergrund der Bereitstellung der Daten aufgezeigt und danach der Zugang zu den in Abschnitt 2 dargestellten Stichproben für die Wissenschaft beschrieben. Im letzten Abschnitt beschäftigen wir uns in einem Ausblick mit den Herausforderungen, vor denen das FDZ bei dieser Datenbereitstellung steht.

³ Die Serviceleistungen des Forschungsdatenzentrums bei der Dokumentation der Daten und ihrer methodischen Spezifika sowie die Beratungsleistungen des FDZ werden hier aus Platzgründen nicht näher erläutert.

2. DATENVERFÜGBARKEIT AM FDZ DER BA IM IAB⁴

Der Datenbestand der BA und des IAB ist vielfältig was die Entstehung, den Inhalt und die Zeitdimension der Daten angeht. Im Hinblick auf die *Daten-genese* kann generell unterschieden werden zwischen Daten, die über das Meldeverfahren der Sozialversicherung erfasst bzw. durch BA-interne Verfahren gesammelt und Daten, die durch Befragungen des IAB erhoben werden. *Inhaltlich* umfassen die Daten der BA Informationen zu Beschäftigten, Betrieben, Arbeitsuchenden, arbeitsmarktpolitischen Maßnahmen und Leistungsempfängern. Die unterschiedlichen *Zeitdimensionen* beziehen sich auf stichtagsbezogene Statistiken, Längsschnittdaten auf Stichtagsbasis und zeitraumbezogene Datensätze (Historiken).

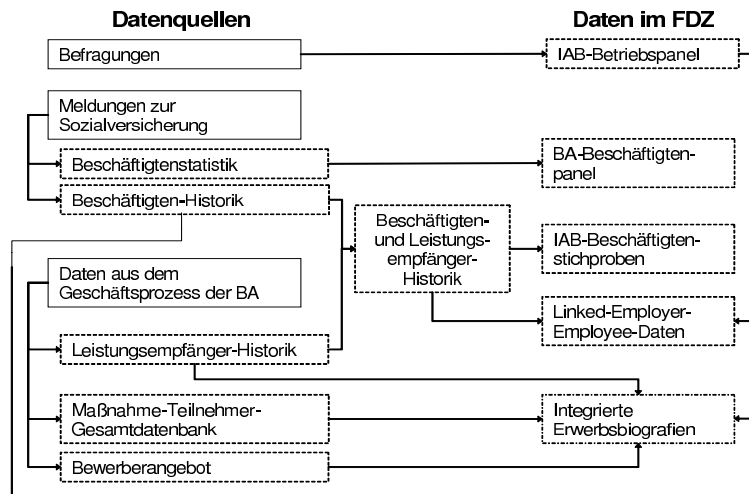


ABBILDUNG 1. Datenquellen, Datenwege und die Datenbereitstellung im FDZ

Die verschiedenen Einzeldaten werden im IAB zu integrierten Datenbanken aufbereitet, in denen mehrere Teilaspekte des Arbeitsmarktes miteinander verknüpft werden. Das FDZ zielt auf die sukzessive Bereitstellung dieser Daten unter Berücksichtigung des Datenschutzes ab. In Abbildung 1 verdeutlichen wir den Zusammenhang zwischen Datenquellen, Datenwegen und der Bereitstellung der Daten im FDZ.

Die Datenquellen und die Daten im FDZ werden nachfolgend beschrieben.

⁴ Im Folgenden gehen wir ausschließlich auf die Daten ein, die bis Mitte 2005 im FDZ zugänglich sind bzw. vom FDZ betreut werden. Es werden nicht alle vorhandenen Quelldaten beschrieben. Einen Überblick über alle Daten, die bis Ende der Projektlaufzeit im November 2006 durch das FDZ zur Verfügung gestellt werden sollen, geben Oertel, Passenberger und Janser (2004). Eine Priorisierung der Bereitstellung weiterer Daten erfolgt nach einer Nutzerbefragung, die Anfang 2005 stattfindet. Die Vorwegnahme der Ergebnisse der Nutzerbefragung ist zum Zeitpunkt der Drucklegung dieses Artikels nicht möglich.

2.1. BEFRAGUNGSDATEN: DAS IAB-BETRIEBSPANEL. Das IAB-Betriebspanel⁵ (Bellmann, 2002; Bellmann, Kohaut und Lahner, 2002) ist die größte Betriebsbefragung (repräsentative Zufallsstichprobe aller Wirtschaftszweige und Betriebsgrößenklassen) der Bundesrepublik Deutschland, die seit 1993 in Westdeutschland und seit 1996 in Ostdeutschland jährlich vom IAB-Forschungsbereich „Betriebe und Beschäftigung“ durchgeführt wird, der Bezugszeitpunkt ist der 30. Juni eines jeden Jahres. Die gegenwärtig aktuellste Welle liegt für das Jahr 2003 vor. Der Stichprobenumfang der Wellen wurde 1999 und 2000 aufgestockt, so dass zwischen 4.265 Betrieben (1993, nur Westdeutschland) und ca. 16.000 Betrieben (2003, 5.300 Betriebe davon in Ostdeutschland) für die Analyse zur Verfügung stehen. Eine vollständige Panelanalyse (1993-2003) ist für 1000 Betriebe in Westdeutschland und für 1.700 Betriebe in Ostdeutschland (1996-2003) möglich. Die Themen des IAB-Betriebspanels umfassen Fragenkomplexe zu Beschäftigung, Einstellungen und Entlassungen, Personalbedarf und -suche, zu Umsatz, Investitionen und Export, Innovationen und organisatorischen Änderungen, Entlohnung, Arbeits- und Betriebszeiten, Aus- und Weiterbildung sowie Angaben zur öffentlichen Förderung. Neben diesen jährlich wiederholten Themenkomplexen sind wechselnde Schwerpunktthemen enthalten, insgesamt sind ca. 300 Variablen auswertbar. Mit dem IAB-Betriebspanel sind repräsentative Auswertungen auch auf der Bundeslandebene durch Hochrechnung innerhalb der Zellen einer Matrix von zehn Betriebsgrößenklassen und zwanzig Wirtschaftszweigen möglich. Analysen auf der Basis des IAB-Betriebspanels konzentrierten sich bislang auf die betriebliche Beschäftigungsentwicklung, die betriebliche Aus- und Weiterbildung, die betriebliche Flexibilität, die Tarifbindung, die betriebliche Interessenvertretung und monetäre Anreizsysteme, die betriebliche Produktivitäts- und Innovationsforschung und verschiedene Branchenanalysen⁶. Das IAB-Betriebspanel eröffnet jedoch nicht nur Möglichkeiten für die separate Analyse von Betrieben. Da das Betriebspanel eine repräsentative Stichprobe aus den Betriebsnummern in der Beschäftigten-Historik (s.u.) ist, ist eine Verknüpfung mit Beschäftigtendaten (Linked-Employer-Employee-Daten, s.u.) und damit eine erhebliche Erweiterung der inhaltlichen Analysepotenziale möglich. Das IAB-Betriebspanel enthält ausschließlich Betriebe, die zum Zeitpunkt der Befragung mindestens eine sozialversicherungspflichtige Person beschäftigen.

2.2. MELDUNGEN ZUR SOZIALVERSICHERUNG. Grundlage der meisten Datenmengen der BA und des IAB sind die Meldungen der Arbeitgeber über ihre sozialversicherungspflichtig beschäftigten Arbeitnehmer an die Krankenkassen. Die gesetzliche Grundlage ist das mit Wirkung vom 1. Januar

⁵ Detaillierte Angaben zum IAB-Betriebspanel finden sich auf den Internetseiten des FDZ unter <http://fdz.iab.de/pageText.asp?PageID=24> und auf den Internetseiten des IAB-Forschungsbereichs „Betriebe und Beschäftigung“ unter <http://betriebspanel.iab.de>.

⁶ Eine Sammlung von Literaturangaben zu durchgeführten empirischen Analysen mit dem IAB-Betriebspanel findet sich unter <http://betriebspanel.iab.de/publikationen.htm>.

1973 eingeführte integrierte Meldeverfahren zur Kranken-, Renten- und Arbeitslosenversicherung (DEVO/DÜVO), das seit dem 1. Januar 1999 in einer aktualisierten Version (DEÜV) gilt.⁷

Damit liegt eine Vollerhebung aller sozialversicherungspflichtig Beschäftigten - seit 1999 auch der geringfügig Beschäftigten - in Deutschland vor, die gegenwärtig etwa 32 Millionen Beschäftigungsverhältnisse umfasst. Bei der Meldung der sozialversicherungspflichtig Beschäftigten durch die Arbeitgeber an die Krankenkassen wird auch die Betriebsnummer erfasst, die dem Arbeitgeber von der zuständigen Arbeitsagentur zugeteilt wird. Diese Betriebsnummern sind die Auswahlbasis für das IAB-Betriebspanel (s.o.).

Die Krankenkassen leiten die Meldungen an die Rentenversicherungsträger weiter, von wo aus sie dann der BA übermittelt werden, die diese Daten ihrem gesetzlichen Auftrag gemäß (§ 281 Sozialgesetzbuch III) registriert und in der Beschäftigtenstatistik auf Quartalsebene verwaltet. Der Eingang der Daten erfolgt verzögert, da Arbeitgebern Abgabefristen für die Meldungen eingeräumt werden. Diese Meldungen bilden die Grundlage für die Stichprobe des BA-Beschäftigtenpanels, das seit 2003 als Scientific Use File für die Wissenschaft zur Verfügung gestellt wird.

2.2.1. DAS BA-BESCHÄFTIGTENPANEL. Das BA - Beschäftigtenpanel⁸ (Koch und Meinken, 2004) besteht aus einer 1,92 %-Personenstichprobe (Geburtstagsauswahl) aus den Quartalsdaten der Beschäftigtenstatistik und enthält sozialversicherungspflichtig Beschäftigte und ab Juni 1999 auch geringfügig Beschäftigte. Es liegen 20 Wellen vor, beginnend mit dem 1. Quartal 1998 und endend mit dem 4. Quartal 2002, im FDZ werden Aktualisierungen in einjährigem Rhythmus durchgeführt. Jede Welle umfasst ca. 600.000 Fälle (500.000 in Westdeutschland, 100.000 in Ostdeutschland), im Quartal 4/2002 sind 530.000 sozialversicherungspflichtig Beschäftigte und 80.000 geringfügig Beschäftigte enthalten. 52 Individual- und betriebliche Merkmale stehen für die Analyse zur Verfügung. Die Individualmerkmale umfassen u.a. soziodemographische Angaben, Staatsangehörigkeit, (Aus-) Bildungsabschluss, Stellung im Beruf, Rentenversicherungsträger, sozialversicherungspflichtiges Entgelt, Betriebswechsel und Beruf. Die betrieblichen Merkmale enthalten die Betriebsgrößenklasse, den Wirtschaftszweig, die Region und Anteilswerte (Anteil der Frauen, der Auszubildenden, der Personen mit deutscher Staatsangehörigkeit etc.) für den Betrieb, in dem die Beschäftigten tätig sind.

⁷ DEVO (Datenerfassungsverordnung) - Verordnung über die Erfassung von Daten für die Träger der Sozialversicherung und für die Bundesanstalt für Arbeit vom 24. November 1972 (BGBl. I: 2159 ff.). DÜVO (Datenübertragungsverordnung) - Verordnung über die Datenermittlung auf maschinell verwertbaren Datenträgern im Bereich der Sozialversicherung und der Bundesanstalt für Arbeit vom 18. Dezember 1972 (BGBl. I: 2482 ff.). DEÜV (Datenerfassungs- und übermittlungsverordnung) - Verordnung über die Erfassung und Übermittlung von Daten für die Träger der Sozialversicherung vom 10. Februar 1998 (BGBl. I: 343 ff., Artikel 1).

⁸ Detaillierte Informationen zum BA-Beschäftigtenpanel finden sich auf den Internetseiten des FDZ unter <http://fdz.iab.de/pageText.asp?PageID=69>.

Die faktische Anonymisierung der Zufallsstichprobe erfolgt in Anlehnung an die für den Mikrozensus entwickelten Kriterien (Müller et al., 1991) und besteht in dem Entfernen bzw. Ersetzen von Identifikatoren, der systemfreien Sortierung und der Vergrößerung von Merkmalsausprägungen. Eine Längsschnittanonymisierung wurde nicht durchgeführt. Das Projekt der Erstellung des BA-Beschäftigtenpanels wurde vom BMBF zwischen 1999 und 2003 gefördert.

Das Scientific Use File des BA-Beschäftigtenpanels wurde bislang erst von wenigen Wissenschaftlern genutzt, entsprechend liegen noch keine intensiven Analyseerfahrungen vor. Generell als Vorteile des Datensatzes sind zu nennen die durch das Auswahlverfahren gewährleistete Analysierbarkeit regionaler Mobilität von Beschäftigten, die strukturtreue Abbildung von Saison- und Konjunkturmustern sowie ein vereinfachtes Datenhandling von komplexen prozessproduzierten Daten. Die Validität vor allem der versicherungsrechtlich relevanten Merkmale ist sehr hoch, die Zuverlässigkeit der darüber hinausgehenden Variablen ist dagegen weniger gut geprüft. Zu den Problemen des Scientific Use Files ist der geringe Merkmalsumfang, die durch das Meldeverfahren bedingte Zeitverzögerung bei der Datenbereitstellung und die geringe regionale Differenzierung (Ost-/Westdeutschland) zu zählen.

Der Nachteil der geringen regionalen Differenzierung ist ein Spezifikum des Scientific Use Files: In der Originalstichprobe sind Angaben auf Kreis- und Arbeitsamtebene verfügbar. Die Weitergabe von gleichzeitig wirtschaftsfachlich und regional tief gegliederten Personenmerkmalen und Betriebsdaten ist jedoch aus Gründen des Datenschutzes nicht möglich (Müller et al., 1991). Um dennoch solche differenzierten Analysen zu ermöglichen, steht die Originalstichprobe schwach anonymisiert per Datenfernverarbeitung und für Gastaufenthalte im FDZ zur Verfügung (vgl. Abschnitt 3).

2.3. MELDUNGEN ZUR SOZIALVERSICHERUNG UND LEISTUNGSEMPFANGSDATEN. Die von den Rentenversicherungsträgern eingehenden Meldungen zur Sozialversicherung werden im IAB archiviert und als Historikdatensatz aufbereitet, der *Beschäftigten-Historik* (*BeH*). Sie umfasst – im Gegensatz zu der quartalsweisen Aufbereitung der Daten in der Beschäftigtenstatistik (s.o.) – die gesamten Beschäftigungsverläufe mit tagesgenauen Angaben zum Beginn und zum Ende von Beschäftigungen einzelner Personen. Die Beschäftigten-Historik enthält personen-, betriebs- und beschäftigungsbezogene Merkmale sowie Merkmale zum Jahreskonto des/der Versicherten.

In gleicher Weise wie bei der Beschäftigten-Historik werden im IAB auch Informationen zum Leistungsempfang als Ereignisdaten aufbereitet. Die *Leistungsempfänger-Historik* (*LeH*) speist sich aus Daten aus dem Verfahren zur Auszahlung von Leistungen (*coLei*), einem internen Verfahren der BA. Diese Daten umfassen neben personenbezogenen Informationen Informationen über den Leistungsbezug (Arbeitslosengeld, Arbeitslosenhilfe,

Unterhaltsgeld, Kranken-/Pflegeversicherungsbeiträge) und Merkmale, die sich auf das letzte Beschäftigungsverhältnis beziehen. Die Daten werden aus den Arbeitsagenturen täglich an die Krankenkassen im Rahmen des integrierten Meldeverfahrens übermittelt.

Die Beschäftigten-Historik und die Leistungsempfänger-Historik können anhand der in beiden Datenbasen enthaltenen Sozialversicherungsnummern der Beschäftigten in dem Verfahren der *Beschäftigten- und Leistungsempfänger-Historik (BLH)* miteinander verknüpft werden. Dieses Verfahren ist die Ausgangsbasis für die IAB-Beschäftigtenstichproben und die Linked-Employer-Employee-Daten des IAB (LIAB, s.u.).

2.3.1. DIE IAB - BESCHÄFTIGTENSTICHPROBEN. Die IAB - Beschäftigtenstichproben (IABS)⁹ liegen mittlerweile seit mehreren Jahren als Scientific Use Files der Fachöffentlichkeit zur Nutzung vor. Es lassen sich zwei Varianten der Scientific Use Files unterscheiden¹⁰:

Das *Basisfile* der IABS (Bender und Haas, 2002; Bender, Haas und Klose, 2000a) ist eine zweistufige systematische 1%-Zufallsauswahl (getrennte Ziehung für Deutsche/Ausländer und West/Ostdeutschland) aus allen sozialversicherungspflichtig Beschäftigten zwischen 1975 und 1995 (Westdeutschland, 1992-1995 für Ostdeutschland). Diese Informationen stammen aus der Beschäftigten-Historik und wurden um Leistungsempfangszeiten aus der Leistungsempfänger-Historik und um Betriebsinformationen ergänzt. Die Daten umfassen die Meldungen von ca. 483.000 Personen (6.7 Mio Meldungen) in Westdeutschland und ca. 76.000 Personen in Ostdeutschland (ca. 354.000 Meldungen). Die Merkmale enthalten soziodemographische Informationen, Informationen zum Beschäftigungsverhältnis (u.a. Beginn und Ende der Beschäftigung, tagesgenaues sozialversicherungspflichtiges Bruttoentgelt, Beruf (3-Steller), Stellung im Beruf etc.), Informationen zum Betrieb (2-stelliger Wirtschaftszweig, Größe des Betriebes etc.) und Merkmale zum Leistungsbezug (Beginn und Ende der Periode, Art der Leistung etc.). Die regionale Gliederung dieses Datensatzes ist sehr grob, es wird lediglich zwischen West- und Ostdeutschland anhand des Betriebsorts unterschieden. Die Querschnittanonymisierung der Daten erfolgte wie bei dem BA-Beschäftigtenpanel nach den Kriterien, die für den Mikrozensus entwickelt wurden. Darüber hinaus wurden die Daten längsschnittanonymisiert (Verschiebung des gesamten Erwerbsverlaufs jeder Person auf der Zeitachse¹¹). Die Anonymisierung des Basisfiles der IAB-Beschäftigtenstichprobe erfolgte im Rahmen eines von der DFG geförderten Projekts in Kooperation zwischen dem IAB und der Universität Rostock (Prof. Dr. Peter A. Berger).

⁹ Detaillierte Informationen zu den IAB-Beschäftigtenstichproben finden sich im Internet-Angebot des FDZ unter <http://fdz.iab.de/pageText.asp?PageID=26>.

¹⁰ Im Folgenden wird ausschließlich auf die aktuellen Stichproben eingegangen, die Vorgänger dieser Scientific Use Files werden nicht beschrieben.

¹¹ Zu Auswirkungen der Längsschnittanonymisierung vgl. Bender, Haas und Klose (2000b).

Das *Regionalfile* der IABS (Hamann et al., 2004) ist eine einstufige reine 2%-Zufallsauswahl (getrennte Ziehung für Deutsche/Ausländer und West-/Ostdeutschland) aus allen sozialversicherungspflichtig Beschäftigten (ab Juni 1999 auch geringfügig Beschäftigte) zwischen 1975 und 2001. Die Stichprobe wurde erstmals aus der Beschäftigten- und Leistungsempfänger-Historik gezogen und umfasst Angaben zu 1,1 Mio Personen in Westdeutschland (18,9 Mio Meldungen) und 181.000 Personen in Ostdeutschland (2,1 Mio Meldungen). Der Variablenkatalog ähnelt dem des Basisfiles, der Schwerpunkt liegt jedoch auf dem regionalen Aspekt: In diesem Scientific Use File ist die räumliche Gliederung differenzierter (343 Mikrozensusregionen auf Basis der Kreiskennziffer), die wirtschaftsfachliche Gliederung (16 Wirtschaftszweige, 130 Berufsaggregate) jedoch gröber. Darüber hinaus sind einige Merkmale nicht mehr enthalten (z.B. Kinderzahl), die Gründe für die Abgabe von Meldungen zur Sozialversicherung sind stärker differenziert. Die Querschnittanonymisierung der Stichprobe erfolgte nach den Mikrozensus-Kriterien, eine Längsschnittanonymisierung wurde nicht vorgenommen.

Das Regionalfile wurde in einem DFG-geförderten Projekt in Kooperation zwischen dem IAB und der Universität Siegen (Prof. W. Ludwig-Mayerhofer) erstellt.

Eine Aktualisierung des Regionalfiles wird im zweijährigen Rhythmus im FDZ vorgenommen. Eine Aktualisierung des Basisfiles ist im Kontext der Erstellung von Stichproben aus dem Verfahren der Integrierten Erwerbsbiographien (s.u.) im FDZ geplant. Einen Überblick über die Analysemöglichkeiten und -grenzen sowie über Anwendungen mit den Scientific Use Files der IAB-Beschäftigtenstichprobe geben Bender und Haas (2002) und Bender, Haas und Klose (2000a).

Wie bei dem BA-Beschäftigtenpanel gilt auch bei den IAB-Beschäftigtenstichproben (Basisfile und Regionalfile) die datenschutzrechtlich bedingte Einschränkung, dass mit den Scientific Use Files nicht gleichzeitig wirtschaftsfachlich und regional tief gegliederte Informationen über Individuen und Betriebe zur Verfügung stehen dürfen. Diese Informationen sind ausschließlich in der Original-Stichprobe enthalten, die im Rahmen von Gastaufenthalten (vgl. Abschnitt 3) Wissenschaftlern zur Verfügung steht. Das Originalfile enthält Regionalangaben auf Kreis- und Arbeitsamtebene (Kreiskennziffern), Wirtschaftszweige (3- und 4-Steller), ausgeübte Tätigkeiten (3- und 4-Steller) sowie den vollständigen Umfang der Abgabegründe der Beschäftigtenmeldung und der Gründe für das Ende des Leistungsbezugs. Die Daten werden schwach anonymisiert zur Verfügung gestellt.

2.4. DIE VERKNÜPFUNG VON BESCHÄFTIGTEN- UND BETRIEBSDATEN.

Die personenbezogenen Daten des Verfahrens der Beschäftigten-Leistungsempfänger-Historik können neben der IAB-Beschäftigtenstichprobe auch zu einer Verknüpfung mit den IAB-Betriebspaneldaten herangezogen werden. Diese Verknüpfung von Beschäftigten- und Betriebsinformationen ist vor

allem dann von Interesse, wenn Interaktionen der Angebots- und Nachfrageseite auf dem Arbeitsmarkt untersucht werden sollen.

2.4.1. LINKED-EMPLOYER-EMPLOYEE-DATEN DES IAB (LIAB). Die Integration von Personendaten aus dem Verfahren der Beschäftigten-Leistungsempfänger-Historik mit Betriebsdaten aus dem IAB-Betriebspanel resultiert in den *Linked-Employer-Employee-Daten des IAB (LIAB)*¹² (Alda 2005a, 2005b, 2005c, 2005d), die seit 2005 Forschern im Rahmen von Gastaufenthalten im FDZ für Analysen zur Verfügung gestellt werden.

In Abbildung 2 zeigen wir das Grundprinzip der Verknüpfung zwischen den Personendaten und den Betriebsdaten. Informationen zur Verknüpfungsqualität der Daten werden in Alda (2005b) gegeben.

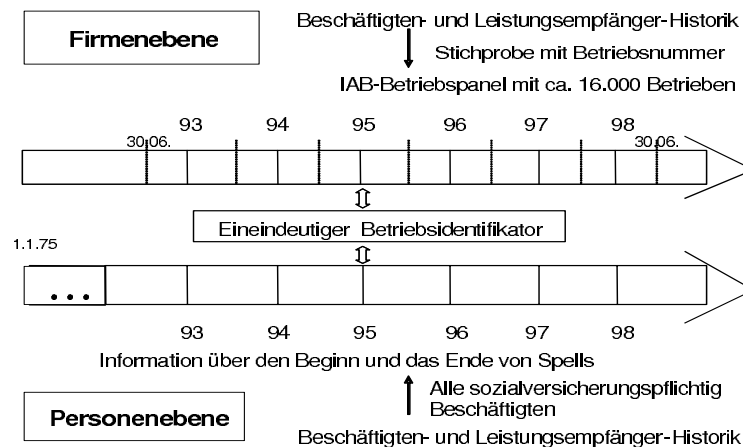


ABBILDUNG 2. Verknüpfung von Betriebs- und Beschäftigteninformationen

Wie oben bereits angesprochen, werden bei der Meldung der sozialversicherungspflichtig Beschäftigten durch die Arbeitgeber an die Krankenkassen auch die Betriebsnummern erfasst, die den Arbeitgebern von der zuständigen Arbeitsagentur zugeteilt werden und die die Auswahlgrundlage für das IAB-Betriebspanel sind. Die in der Befragung erfassten Informationen beziehen sich in der Regel auf den Stand zum 30. Juni eines Jahres.

Informationen aus dem IAB-Betriebspanel zum 30. Juni eines Jahres können mit fortlaufenden Informationen über die Beschäftigten aus dem Verfahren der Beschäftigten- und Leistungsempfänger-Historik mit Hilfe der

¹² Detaillierte Informationen zu den Linked-Employer-Employee-Daten des IAB (LIAB) finden sich auf den Internet-Seiten des FDZ unter <http://fdz.iab.de/pageText.asp?PageID=18>.

Betriebsnummer als eindeutigen Identifikator verknüpft werden. Diese Verknüpfung ist gegenwärtig möglich für die Jahre 1993 bis 2001. Grundsätzlich können alle Betriebe aus dem IAB-Betriebspanel in eine solche Verknüpfung mit einbezogen werden und die entsprechenden Individualinformationen hinzugespielt werden. Die Handhabung der Verknüpfung dieser Datenmengen ist jedoch aufgrund technischer Restriktionen kaum möglich.

Daher¹³ hat das FDZ in Kooperation mit der Arbeitsgruppe Linked-Employer-Employee-Daten des IAB (LIAB) und dem IAB-Forschungsbereich „Betriebe und Beschäftigung“ zwei Datenmodelle für Versionen des LIAB entwickelt, die eine ausreichende Anzahl an Beobachtungen bei gleichzeitiger Berücksichtigung der verfügbaren Prozessorleistungen und der Größe von Original- und Arbeitsdatensätzen enthalten.

Die beiden Datenmodelle sind keine real existierenden Datensätze, sondern sie beschreiben die verschiedenen Aufbereitungen der Personendaten (Verfahren der Beschäftigten-Leistungsempfänger-Historik) für eine Verknüpfung mit den Betriebsdaten des IAB-Betriebspanels. Die Verknüpfung dieser aufbereiteten Personendaten mit den Informationen aus dem IAB-Betriebspanel wird durch die Gastforscher im FDZ selbst vorgenommen.

Die Aufbereitung der Personendaten in den Modellen variiert im Hinblick auf die Verknüpfungslogik: es kann ein Querschnittmodell von einem Längsschnittmodell unterschieden werden, beide sind seit Jahresbeginn 2005 im Rahmen von Gastaufenthalten nutzbar.

Im *Querschnittmodell* (Alda, 2005c) des LIAB sind in so genannten „Jahresscheiben“ alle Meldungen von Personen enthalten, die zum 30. Juni eines Jahres (Referenzzeitpunkt der IAB-Betriebspanelwellen) zwischen 1993 und 2001 in einem befragten Betrieb sozialversicherungspflichtig beschäftigt waren. Diese einzelnen Querschnitte können über Betriebs- und Personenidentifikatoren miteinander verbunden werden, so dass Erwerbsverläufe von Personen in Panelbetrieben verfolgt werden können. Zu Personen, die während dieses Zeitraums einen im IAB-Betriebspanel erfassten Betrieb verlassen und nicht in einen anderen befragten Betrieb wechseln, sind nach dem Wechselereignis keine Informationen mehr verfügbar. Dieses Datenmodell entspricht am ehesten der LIAB-Variante, die in Bellmann, Bender und Kölling (2002) dokumentiert wurde. Das Querschnittsmodell umfasst im Zeitraum 1993 bis 2001 pro Jahr zwischen 1,9 und 2,7 Mio Beschäftigte in 3.900 bis max. 15.000 Betrieben. Die aufbereiteten Personendaten des Querschnittmodells enthalten 24 Variablen. Neben systemfreien und umgeschlüsselten Organisationsvariablen zur Verknüpfung zwischen den Querschnitten und zwischen Betrieben und Beschäftigten liegen als inhaltliche Variablen u.a. die Berufskennziffer (3-Steller), das sozialversicherungspflichtige Tagesentgelt, der Abgabegrund der Beschäftigtenmeldung, die Personengruppe

¹³ Neben technischen Restriktionen sind auch datenschutzrechtliche Aspekte ausschlaggebend für die Formulierung von Datenmodellen. Die Standardisierung von Datenanfragen (vgl. Abschnitt 3) hat bei der datenschutzrechtlichen Genehmigung eine deutliche Verkürzung des Wegs von der Antragstellung bis zum tatsächlichen Arbeiten mit den Daten zur Folge.

(ermöglicht die Identifikation von geringfügig Beschäftigten ab 1999) und die Stellung im Beruf vor. Insofern die regionalen und wirtschaftsfachlichen Informationen des IAB-Betriebspanels für die gewünschten Analysen nicht hinreichend differenziert sind, können Forscher – begründet – die Nutzung „sensibler“ Merkmale zu regionalen Aspekten (Arbeitsamtbezirke und Kreiskennziffern von Wohnort und Arbeitsort) und Wirtschaftszweigen (3-Steller der BA-Klassifikation 1973 und 5-Steller der BA-Klassifikation 1993) für ihren Gastaufenthalt beantragen.¹⁴ Auf der Betriebsseite sind prinzipiell alle Variablen des IAB-Beschäftigtenpanels mit den o.g. Personeninformationen verknüpfbar. Es erfolgt ausschließlich eine Umschlüsselung der Betriebsidentifikatoren in systemfreie Nummern, die identisch sind mit den in den Personendaten umgeschlüsselten systemfreien Betriebsnummern.

Im Laufe des Jahres 2005 werden Erweiterungen der Querschnittaufbereitung der Personendaten vorgenommen. Um die Analysemöglichkeiten von Mobilitätsprozessen zu erweitern, werden die Vor- und Nachgeschichte der Informationen einer Jahresscheibe bereitgestellt. Dies ist möglich, insofern eine Person im Vorjahr bzw. im Folgejahr des jeweiligen Querschnitts sozialversicherungspflichtig beschäftigt war. Darüber hinaus werden Indikatoren zur Historisierung des Erwerbsverlaufs (z.B. erstes Eintrittsdatum der Beschäftigten in den Betrieb und in das Erwerbsleben) in die Personendaten integriert.

Neben der Aufbereitung der Personendaten im Querschnitt wurde im FDZ ein *Längsschnittmodell* (Alda, 2005d) des LIAB entwickelt, das seit Jahresbeginn 2005 im Rahmen von Gastaufenthalten genutzt werden kann. Hierbei wurden je 2.100 Betriebe in Ost- und Westdeutschland ausgewählt¹⁵, die mindestens 1999 bis 2001 am IAB-Betriebspanel teilgenommen hatten. Für jede Person, die in einem dieser Betriebe mindestens einen Tag sozialversicherungspflichtig beschäftigt war, sind alle Beschäftigungs- und Leistungsmeldungen zwischen 1991 (Westdeutschland) bzw. 1992 (Ostdeutschland) und 2001 in den Personendaten enthalten. Im Unterschied zu dem Querschnittsmodell können hierbei also komplette Erwerbs- und Leistungsbezugsbiographien für Personen zwischen 1991 bzw. 1992 und 2001 untersucht werden, unabhängig davon, ob die Personen vor 1999 in einem befragten Betrieb beschäftigt waren oder nicht. Es ergeben sich ca. 25,5 Mio Meldungen (14,5 Mio Meldungen in Westdeutschland, 11 Mio Meldungen in

¹⁴ Diese Restriktion ergibt sich aus dem Umstand, dass bei der Hinzuziehung dieser stärker detaillierten Informationen aus den Personendaten ein höheres Reidentifikationsrisiko der Betriebe entsteht. Die Einbeziehung der sensiblen Variablen bedeutet einen größeren Aufwand bei der datenschutzrechtlichen Überprüfung der in Gastaufenthalten gewonnenen Ergebnisse und damit eine zeitliche Verzögerung der Übermittlung der geprüften Resultate (vgl. Abschnitt 3).

¹⁵ Die Auswahl dieser Betriebe beruht auf den Ergebnissen von Ausfallanalysen zu dem Problem, dass bei manchen Betrieben trotz existierender Betriebsnummer keine Personendaten gefunden werden konnten oder dass die im IAB-Betriebspanel erfassten Angaben im Hinblick auf die Anzahl sozialversicherungspflichtig Beschäftigter von den Informationen in den Personendaten erheblich abweichen (vgl. Alda, 2005b). Die ausgewählten 2.100 Betriebe liegen im Jahr 2001 innerhalb des in Alda (2005b) definierten Toleranzintervalls.

Ostdeutschland) von 1,8 Mio Personen (1,1 Mio Personen in Westdeutschland, 730.000 Personen in Ostdeutschland). Der Variablenkanon auf der Personenebene ist identisch mit den Personeninformationen im Querschnittmodell. Darüber hinaus sind Merkmale, die sich auf den Leistungsbezug durch die BA beziehen, enthalten¹⁶, u.a. die Leistungsart, der Bewilligungsgrund für den Leistungsbezug und der Grund der Beendigung des Leistungsbezugs. Auch bei der Nutzung der im Längsschnitt aufbereiteten Personendaten für die Verknüpfung mit den Informationen des IAB-Betriebspanels gelten die o.g. Bestimmungen bezüglich der sensiblen Merkmale und der Verknüpfbarkeit aller inhaltlichen Variablen des IAB-Betriebspanels mit der Längsschnittaufbereitung der Personendaten.

Weiterentwicklungen des Längsschnittmodells der LIAB-Daten im Jahr 2005 konzentrieren sich auf die Ausweitung der Jahre (1993-2001), in denen die Betriebe an der Betriebspanel-Befragung teilgenommen haben müssen. Diese Variante beschränkt sich damit auf Westdeutschland¹⁷.

Auf dem im November 2004 vom FDZ in Kooperation mit dem IAB-Forschungsbereich „Betriebe und Beschäftigung“, der IAB-Arbeitsgruppe „Linked-Employer-Employee-Daten des IAB“ und dem IAB-Servicebereich „Informationstechnologie und Informationsmanagement“ durchgeführten Workshop zu den Datenmodellen des LIAB wurden die o.g. Modelle von der scientific community positiv aufgenommen.

Die durchgeführte Befragung der Workshopteilnehmer zeigte u.a., dass von 45 Personen, die an der Befragung teilnahmen, 32 Personen die erstellten Modelle für die von ihnen geplanten Analysen verwenden können und nur eine Person die Daten als nicht hilfreich empfand¹⁸, 17 Personen planten dann auch konkret einen Gastaufenthalt. Weiterhin wurden auf dem Workshop auch Anregungen für die weitere Entwicklung der Quer- und Längsschnittmodelle gegeben. Der Einschluss von aggregierten Angaben zur Erwerbsbiographie im Querschnittmodell (s.o.) ist z.B. ein Resultat dieser Anregungen.¹⁹

Datenmodelle tragen in mehrfacher Weise dazu bei, dass der Datenzugang für Forscher schneller und effizienter möglich wird, indem die Aufbereitung der komplexen Daten reduziert wird und gleichermaßen die Bearbeitungsdauer des Antrags auf Datenzugang im FDZ durch das BMWA (vgl.

¹⁶ Aufgrund der Bedingung, dass die Personen lediglich einen Tag zwischen 1999 und 2001 in einem sozialversicherungspflichtigen Beschäftigungsverhältnis mit einem befragten Betrieb gestanden haben müssen, ergibt sich die Möglichkeit, dass diese Personen auch Zeiten des Leistungsbezugs in ihrer Biographie aufweisen. Dies ist bei dem Querschnittsmodell aufgrund der Selektionsvorgabe (s.o.) nicht möglich.

¹⁷ Ostdeutsche Betriebe wurden erst ab 1996 im IAB-Betriebspanel befragt (s.o.).

¹⁸ 10 Personen waren sich noch nicht sicher, ob die Daten für ihre Analysezwecke geeignet wären, von drei Personen wurden keine Angaben gemacht.

¹⁹ Das Vorgehen bei den LIAB-Modellen ist ein Beispiel für weitere Datenmodelle bzw. die Ziehung von Stichproben durch das FDZ: Es werden Vorschläge durch das FDZ entwickelt, die mit der scientific community auf Workshops diskutiert werden. Veränderungs- und Ergänzungsvorschläge von der scientific community werden dann – insofern technisch und datenschutzrechtlich möglich – in die Daten eingearbeitet.

Abschnitt 3) maßgeblich reduziert wird. Auch für den Betreuungsaufwand bei einem Gastaufenthalt ist mit einer größeren Effizienz zu rechnen, die sich wiederum in einer effizienteren Nutzung der Daten durch Gastforscher niederschlägt. Darüber hinaus bietet die Verwendung standardisierter Datenmodelle auch Synergieeffekte für die Wissenschaft: Einerseits besteht eine erhöhte Wahrscheinlichkeit für die sukzessive Aufdeckung von Mängeln in der Datenqualität und andererseits eine höhere Chance für die Entwicklung von Lösungen dieser Probleme.

Gleichermaßen stellt die Aufbereitung der Personendaten in Quer- oder Längsschnittmodellen nur in geringem Ausmaß eine Einschränkung dar, sondern dient eher als Orientierungsrahmen, innerhalb dessen weitestgehende Freiräume für die Erstellung eigener Panel- und Längsschnittdatensätze gegeben sind. Die grundsätzliche Entscheidung, Gastforschern im FDZ selbst die Möglichkeit zu geben, die für sie relevanten Betriebsmerkmale zu selektieren, spezifisch aufzubereiten und letztendlich mit den auf der Basis verschiedener Datenmodelle vorbereiteten Personendaten zu verknüpfen, trägt der Maxime Rechnung, ein größtmögliches Analysepotenzial der Daten zu erhalten. Die verschiedenen Schwerpunktthemen in den einzelnen Wellen des IAB-Betriebspanels und die jährliche Aktualisierung der Personendatenbasis erweitern dieses Analysepotenzial darüber hinaus.

Die Formulierung und die Umsetzung des Querschnitt- und des Längsschnittmodells verbessert die Nutzungspotenziale von verknüpften Betriebs- und Individualinformationen in den Daten des IAB entscheidend. Dies gilt sowohl für den Zugang externer Wissenschaftler wie aber auch für IAB-Forscher selbst. Die bislang im IAB vorhandenen LIAB-Versionen bestanden ausschließlich in Querschnitten oder in projektspezifisch aufbereiteten Daten. Ein Überblick über Forschungsarbeiten mit diesen älteren LIAB-Versionen wird in Alda, Bender und Gartner (2005) gegeben.

2.5. AUSBLICK: MELDUNGEN ZUR SOZIALVERSICHERUNG UND DATEN AUS DEN VERFAHREN DER BA. Neben der Leistungsempfänger-Historik (LeH, vgl. Abschnitt 2.3) können weitere Verfahren der Arbeitsverwaltung der BA mit Informationen zu sozialversicherungspflichtigen Beschäftigungen aus der Beschäftigten-Historik (BeH, vgl. Abschnitt 2.1) verknüpft werden. Hierzu gehören Informationen zu Maßnahmen der aktiven Arbeitsmarktpolitik und Daten zum Bewerberangebot.

Die *Maßnahme-Teilnehmergesamtdatenbank (MTG)* wird im IAB auf der Basis der BA-Verfahren zur Arbeitsvermittlung (coArb) und zur Verwaltung von Maßnahmeteilnahmen (coSach) sowie auf Basis der zentralen Kundendatenbank der Arbeitsagenturen (CORA) erstellt. Sie enthält personenbezogene Merkmale (demographische Angaben, Staatsangehörigkeit, Einreisestatus, Familienstand, gesundheitliche Einschränkungen etc.), erwerbsbiographische Merkmale (Schulbildung, Berufsbildung, letzter Beruf, Berufserfahrung etc.), BA-dienstleistungsbezogene Merkmale (Leistungsbezug, Arbeitslosigkeit etc.) und Informationen zu Maßnahmen (Art, Beginn,

Ende, Ziel, Praktikum, Vollzeit/Teilzeit, Förderart etc.) und Maßnahmeteilnahme (Eintritt, Austritt, Förderungsart, Erfolg etc.).²⁰

Daten zum *Bewerberangebot* (*BewA*) umfassen Informationen zu allen Arbeitsuchenden, die sich in den Arbeitsagenturen als solche melden, unabhängig ob Leistungen bezogen werden oder nicht. Damit geben diese Daten einen umfassenderen Einblick in das Arbeitskräfte-reservoir als die Leistungsempfängerhistorik, die ausschließlich solche Personen erfasst, die Leistungen beziehen. Neben personenbezogenen Merkmalen (demographische Angaben, Staatsangehörigkeit, Einreisedatum, Kinderzahl, Behinderungsgrad etc.) sind erwerbsbiographische Variablen (Schulbildung, Berufsbildung, gewünschte Tätigkeit, letzte Erwerbstätigkeit etc.) und BA-dienstleistungsbezogene Merkmale (Beratung, Leistungsbezug, Förderung etc.) enthalten. Informationen zum Bewerberangebot stammen aus dem BA-internen Verfahren der Arbeitsvermittlung (coArb).

Die Informationen der BeH, LeH, MTG und BewA erstrecken sich nur in begrenztem Ausmaß über einen gemeinsamen Zeitraum: Während die Daten der BeH aufgrund des gesetzlichen Verfahrens immer einen zeitlichen Verzug von ca. zwei Jahren aufweisen (vgl. Abschnitt 2.2), allerdings bis 1975 zurückreichen, liegen Informationen aus der MTG erst seit 2000 und Merkmale des BewA seit 1997 bis zum aktuellen Rand vor. Lediglich die LeH weist lange zurückreichende Meldungen (1975) bis zum aktuellen Rand aus. Die Daten können mit Hilfe der Sozialversicherungsnummern und der Kundennummern integriert werden. Für den Zeitraum zwischen 1990 und 2004 sind insgesamt ca. 637 Mio Beobachtungen von ca. 60 Mio Personen vorhanden. Diese Datenmengen dürfen aus datenschutzrechtlichen Gründen ausschließlich für die Ziehung von Stichproben²¹ zusammengeführt werden, eine Vorhaltung des gesamten Datenmaterials in integrierter Form ist nicht möglich.

Das FDZ wird themenspezifische Stichproben als Integrierte Erwerbsbiographien (IEB) generieren, die - vorbehaltlich der datenschutzrechtlichen Genehmigung - im Laufe des Jahres 2005 im Rahmen von Gastaufenthalten im FDZ für die Analyse zur Verfügung stehen sollen. Weiterhin ist im FDZ die faktische Anonymisierung dieser Stichproben und die Bereitstellung als Scientific Use Files - soweit möglich - geplant. Hierzu fanden erste Beratungen in einer Arbeitsgruppe bestehend aus IAB- und externen²² Wissenschaftlern sowie dem Bundesbeauftragten für Datenschutz statt.

²⁰ Die Daten zu Maßnahmen und Maßnahmeteilnahme beinhalten die Bereiche der beruflichen Fortbildung, Weiterbildung, Trainingsmaßnahmen, Deutschlehrgänge, die Freie Förderung, Arbeitsbeschaffungsmaßnahmen, Strukturanpassungsmaßnahmen, Lohnkostenzuschuss-Altfälle, Eingliederungszuschüsse, Eingliederungsvertrag, Mobilitätshilfen, Überbrückungsgeld für Existenzgründer, Beschäftigungshilfen für Langzeitarbeitslose etc.

²¹ Verschiedene Stichproben dieser Datenmengen werden gegenwärtig von durch das BMWA geförderten Projekten zur Hartz-Evaluation bearbeitet.

²² Beteiligte Institutionen sind IWH, IZA, RWI, ZUMA und die Universität St. Gallen.

3. DATENZUGANG AM FDZ

Der Datenzugang zu amtlichen Daten hat in Deutschland bereits einige Tradition, wenn auch im Vorfeld der institutionalisierten Lösung der Forschungsdatenzentren dieser Datenzugang unsystematisch war. Dies gilt insbesondere für Daten zum Arbeitsmarkt in Deutschland. Auf dem Weg zu einer Institutionalisierung von Zugangswegen zu diesen Daten sind die Erfahrungen der FDZ der Statistischen Ämter (Zühlke et al., 2003) und des Servicezentrums für Mikrodaten der Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen (GESIS) bei ZUMA (Lüttinger et al., 2004) bei der Bereitstellung von amtlichen Daten von großer Bedeutung. Wenn auch die rechtlichen Grundlagen (vgl. Abschnitt 3.1) dieser Bereitstellung und die Genese der Daten unterschiedlich sind, so sind die Zugangswege zu den Daten der BA und des IAB (vgl. Abschnitt 3.2) ebenso wie die Zugangswege zu den Daten der Rentenversicherung im FDZ der Rentenversicherung (Bütefisch, 2004; Rehfeld, 2004) prinzipiell dieselben wie die Zugangswege zu den Daten der Statistischen Ämter.²³

Im Folgenden beschreiben wir den datenschutzrechtlichen Hintergrund des Zugangs zu den Daten der BA und des IAB. Danach erläutern wir die verschiedenen Zugangswege zu den in Abschnitt 2 beschriebenen Daten, die den Zugangswegen in anderen FDZ gleichen.

3.1. GESETZLICHE GRUNDLAGEN DER DATENÜBERMITTLUNG NACH DEM SOZIALGESETZBUCH (SGB)²⁴. Bei den Daten, die durch das FDZ Forschern zur Verfügung gestellt werden und deren Nutzung Beratungsgegenstand ist, handelt es sich in der Regel um Sozialdaten gemäß § 67 SGB X bzw. um Daten, die Sozialdaten gleichgestellt sind. Sozialdaten sind Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person, die von der BA im Hinblick auf ihre Aufgaben nach dem SGB III erhoben und verarbeitet und dann im IAB für die Forschung aufbereitet werden. Sozialdaten sind damit besonders sensible Daten, die nach § 35 Abs. 1 SGB I als Sozialgeheimnis zu schützen sind.

Dieser besondere Schutz spiegelt sich in den Regelungen zur Datenweitergabe für die Forschung und Planung wider. Der Zugang zu *nicht anony-*

²³ Einen Überblick über den Zugang zu amtlichen Daten im internationalen Vergleich gibt KVI (2001). Bei diesem Vergleich wird deutlich, dass einerseits bislang die Kooperation zwischen Wissenschaft und Statistik in Deutschland sehr großen Nachholbedarf hat, dass andererseits aber auch in den meisten Ländern Befragungsdaten vergleichsweise einfacher zugänglich sind als prozessproduzierte administrative Daten. Vorbildfunktion in Bezug auf den Zugang zu letzteren besitzt vor allem Norwegen.

²⁴ Das Sozialgesetzbuch (SGB) steht im Internet zur Verfügung unter http://www.sozialgesetzbuch-bundessozialhilfegesetz.de/_buch/sgb_i.htm.

*misierten, schwach anonymisierten oder pseudonymisierten Sozialdaten*²⁵ durch die Forschung und Planung wird in §75 SGB X geregelt.

Als Bedingungen für eine solche Nutzung sind festgelegt:

- Die Daten müssen für ein Projekt aus dem Sozialleistungsbereich verwendet werden,
- das öffentliche Interesse muss das Geheimhaltungsinteresse des Betroffenen erheblich überwiegen,
- das Einholen der Einwilligung des Betroffenen muss unzumutbar sein und
- der Zweck des Vorhabens darf nicht auf andere Weise erreichbar sein (z.B. durch die Analyse von anderen Daten, zu denen ein einfacherer Zugang möglich ist).

Die Übermittlung von Sozialdaten²⁶ muss beim BMWA beantragt und von ihm genehmigt werden. Es werden umfangreiche Maßnahmen zur Gewährleistung von Datenschutz und Datensicherheit bei den Datenempfängern vorausgesetzt. Grundsätzlich ist eine Nutzung der Daten für kommerzielle Interessen ausgeschlossen.

Insofern Sozialdaten dahingehend verändert werden, dass Einzelangaben nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft reidentifiziert werden können (§16 Abs. 6 Bundesstatistikgesetz, §67 Abs. 8 SGB X), gelten die Daten als faktisch anonymisiert. In diesem Fall handelt es sich nicht mehr um Sozialdaten. Faktisch anonymisierte Daten der BA und des IAB können an wissenschaftliche Einrichtungen auf Antrag übermittelt werden, wenn es sich bei dem Forschungsvorhaben um ein Projekt aus der Arbeitsmarkt- und Berufsforschung handelt (§282 Abs. 7 SGB III), kommerzielle Forschungsinteressen sind auch hier ausgeschlossen.

3.2. ZUGANGSMÖGLICHKEITEN ZU DEN DATEN DER BA UND DES IAB. Entsprechend den in Abschnitt 3.1 beschriebenen gesetzlichen Regelungen unterscheidet sich der Zugang zu den in Abschnitt 2 beschriebenen Daten je nach Anonymisierungsgrad und nach Sensibilität der Daten. In Tabelle 1 geben wir einen Überblick über die verschiedenen Zugangswege, die wir nachfolgend erläutern.

²⁵ Unter ‚schwacher‘ Anonymisierung wird die systemfreie Umschlüsselung von Identifikatoren im Datensatz bei Erhalt des vollständigen Merkmalsumfangs verstanden. Bei der Pseudonymisierung werden Identifikatoren systematisch umgeschlüsselt bei Erhalt des vollständigen Merkmalsumfangs.

²⁶ Die externe Übermittlung von Sozialdaten der BA und des IAB nach § 75 SGB X kann beim Datenzentrum der Statistik der BA (e-mail: Service-Haus.Statistik-Datenzentrum@arbeitsagentur.de) beantragt werden. Die dabei entstehenden Kosten werden dem Antragsteller in Rechnung gestellt. Das Datenzentrum der Statistik erstellt auch kostenpflichtige Sonderauswertungen aus den amtlichen Statistiken der BA zur Lage und Entwicklung auf dem Arbeits- und Ausbildungsmarkt, zur Beschäftigung, zur Arbeitsförderung einschließlich der Eingliederungsbilanz sowie zu den Lohnersatz- und sonstigen Leistungen.

Datenzugangsweg	Scientific Use File	Datenfernverarbeitung	Gastaufenthalt
Anonymisierungsgrad	(faktisch anonymisierte Stichprobe)	(schwach anonymisierte Originalstichprobe)	(schwach anonymisierte Originalstichprobe)
Daten im FDZ			
IAB-Betriebspanel		X	X
BA-Beschäftigtenpanel	X	X	X
IAB-Beschäftigtenstichproben (Basisfile/Regionalfile)	X		X
Linked-Employer-Employee-Daten des IAB			X
Integrierte Erwerbsbiographien	X		X

TABELLE 1. Datenzugangswegen zu den Daten des FDZ

Drei standardisierte Zugangswege – Scientific Use Files, Datenfernverarbeitung und Gastaufenthalte im FDZ – sind hierbei zu unterscheiden.

Scientific Use Files bezeichnen faktisch anonymisierte Stichproben aus Originaldatensätzen. Diese Art des Datenzugangs wird gewählt, wenn durch die Anonymisierung von Sozialdaten keine großen Einbußen des Analysepotenzials entstehen. Die anonymisierten Stichproben werden Forschern für die Durchführung wissenschaftlicher Projekte im Bereich der Arbeitsmarkt- und Berufsforschung (vgl. Abschnitt 3.1) zeitlich befristet und thematisch gebunden zur Verfügung gestellt. Gegenwärtig sind die IAB-Beschäftigtenstichproben (ASCII-Format, TDA, STATA, SPSS) sowie das BA-Beschäftigtenpanel (ASCII-Format, SAS-Einleseroutine) als faktisch anonymisierte 1%- bzw. 2%-Stichproben verfügbar.

Der Zugang zu den Scientific Use Files der BA und des IAB erfolgt über eine Antragstellung an das Zentralarchiv für empirische Sozialforschung (ZA) in Köln²⁷. Dem Antrag muss eine ca. zweiseitige Projektbeschreibung beigelegt sein, die Projekttitel, -laufzeit, zugangsberechtigte Mitarbeiter und Maßnahmen zur Datensicherheit benennt. Auf der Basis dieser Angaben entscheidet das FDZ über die Überlassung der Daten. Nach dem positiven Bescheid durch das FDZ schließt das ZA gegen eine Bearbeitungsgebühr von 50 € einen Nutzungsvertrag mit dem Datenempfänger. In dem Nutzungsvertrag verpflichtet sich der Datenempfänger zur Unterlassung jeglicher Handlungen zum Zwecke der Deanonymisierung oder der Zusammenführung mit anderen Daten, zur ausschließlichen Nutzung der Daten in den Räumen der Forschungseinrichtung und zu Vorkehrungen zum Schutz der

²⁷ <http://www.gesis.org/ZA/index.htm>

Daten. Der Zeitraum zwischen Beantragung und erfolgter Zusendung der Daten beträgt ca. zwei Wochen.

Nach Ablauf der beantragten Nutzungsdauer sind alle Kopien und Auszüge der Daten zu löschen, die Originaldaten sind an das ZA zurückzuschicken. Eine erneute Beantragung der Nutzung ist möglich. Veröffentlichungen auf der Basis des Scientific Use Files müssen mit der Quellenangabe zur Datenbasis versehen werden, dem ZA und dem FDZ müssen je zwei Belegexemplare zugesendet werden. Die Nutzung der Scientific Use Files der BA und des IAB ist auch für Antragsteller aus EU-Ländern beim ZA möglich, die Antragstellung erfolgt analog.

Bei der *kontrollierten Datenfernverarbeitung* (Schalterstelle) wird Forschern kein direkter Zugang zu den schwach anonymisierten Sozialdaten gewährt.²⁸ Diese Form des Datenzugangs wird praktiziert, wenn eine Anonymisierung der Daten nicht oder nur mit einem unverhältnismäßig großen Informationsverlust verbunden ist.²⁹ Der Zugang von externen Wissenschaftlern zu den Daten des IAB-Betriebspanels mittels kontrollierter Datenfernverarbeitung erfolgte seit 1999 in einem vom BMBF geförderten Projekt im IAB-Forschungsbereich „Betriebe und Beschäftigung“, seit Mai 2004 übernimmt diese Aufgabe das FDZ. Seit Frühjahr 2005 betreut das FDZ auch die kontrollierte Datenfernverarbeitung für das BA-Beschäftigtenpanel.

Interessierte Forscher übersenden dem FDZ ein ausgefülltes Formular³⁰, mit dem die Zugangsvoraussetzungen geprüft werden. Dieses enthält den Namen der beantragenden Institution und der beteiligten Forscher und weist den Verwendungszweck der Analyse³¹ sowie die inhaltlichen Auswertungsziele aus. Insofern einer Auswertung nichts im Wege steht, senden die Forscher die Auswertungsprogramme ein, die in den Statistikprogrammen SPSS, STATA oder SAS verfasst sind.³² Im FDZ wird die Auswertungssyntax dann anhand der Originaldaten ausgeführt. Nach der Auswertung erfolgt im FDZ eine Datenschutzprüfung und die – absolut anonymisierten – Resultate werden den Forschenden übermittelt. Bei weiteren Analysen innerhalb desselben Projekts muss kein neuer Nutzungsantrag gestellt werden. Sofern die Ergebnisse veröffentlicht werden, übersenden die Forscher dem FDZ alsbald ein Belegexemplar. Die eingesendeten Auswertungsprogramme werden sechs Monate für Nachfolgeanalysen auf den Rechnern des FDZ ge-

²⁸ Betriebsdaten sind Sozialdaten in ihrer Schutzwürdigkeit gleichgestellte Daten, vgl. §35 Abs. 4, SGB I.

²⁹ Dies betrifft z.B. das IAB-Betriebspanel. Zu den Möglichkeiten der Anonymisierung dieser Daten vgl. Brand, Bender und Kohaut (1999). Gegenwärtig ist das IAB an einem Projekt des Statistischen Bundesamtes beteiligt, in dem u.a. auch die Möglichkeiten einer adäquaten Anonymisierung des IAB-Betriebspanels erforscht werden.

³⁰ Die Formulare zur Nutzung der Daten über kontrollierte Ferndatenverarbeitung sind auf den Internet-Seiten des FDZ zu den entsprechenden Datensätzen zugänglich, z.B. unter <http://doku.iab.de/fdz/iabb/Anfrageformular.pdf>.

³¹ Auch hier sind kommerzielle Forschungsinteressen von der Nutzung der Daten ausgeschlossen.

³² Als Standard sind diese Programme vorgesehen, die Verwendung anderer, spezieller Programme bedarf der Voranmeldung und ist nicht prinzipiell ausgeschlossen.

speichert. Anschließend werden die Dateien aus Platzgründen auf externe Datenträger ausgelagert. Der Rückgriff auf ältere Auswertungsprogramme ist möglich.

Da die Forscher keinen unmittelbaren Zugang zum Datenmaterial haben, benötigen sie Arbeitshilfen, um korrekte Auswertungsprogramme formulieren zu können. Für das IAB-Betriebspanel wurden für jede Befragungswelle Testdaten entwickelt, welche die Originalstruktur des Datensatzes abbilden. Die Testdaten³³ wurden so verfälscht, dass mit ihnen keine inhaltlichen Auswertungen möglich sind. Zusätzlich wurden einige Betriebe erfunden. Auf diese Weise wird sichergestellt, dass nur lauffähige Auswertungsprogramme an das FDZ übersandt und gleichzeitig die Datenschutzbestimmungen eingehalten werden.

Da das BA-Beschäftigtenpanel auch in einer faktisch anonymisierten Stichprobe für Externe zur Verfügung gestellt wird, wurde bislang auf die Erstellung von Testdaten zu diesem File verzichtet.

Die Bearbeitung der Anfragen über die kontrollierte Datenfernverarbeitung im FDZ erfolgt kostenfrei, die Prüfung der Zugangsvoraussetzungen nach Einsendung des Formulars erfolgt innerhalb von zwei Werktagen. Der Zeitraum zwischen der Einsendung von Analyseprogrammen und der Übermittlung der datenschutzrechtlich geprüften Ergebnisse der Analyse ist abhängig vom Umfang der durchgeführten Analysen. Er beträgt maximal fünf Werktage, im Durchschnitt ergeben sich bislang für die Bearbeitung von Anfragen zum IAB-Betriebspanel zwei Werktage.

Gastaufenthalte sind im Hinblick auf die durch Gastforscher zu erfüllenden Bedingungen der restriktivste, im Hinblick auf die Analysemöglichkeiten gleichzeitig jedoch auch der ergiebigste Zugangsweg für externe Forscher und Forscherinnen. Dieser Zugang wird dann gewählt, wenn das Datenmaterial nicht ohne erheblichen Informationsverlust faktisch anonymisierbar ist, gleichermaßen aber so sensibel und die Bearbeitung so aufwändig ist, dass Auswertungen nur am Ort der Datenhaltung durchgeführt werden können. Im FDZ liegen die Originaldaten der Stichproben schwach anonymisiert oder pseudonymisiert (vgl. Abschnitt 3.1) vor, der Zugang zu diesen Daten im FDZ ist für Gastwissenschaftler an vier abgeschotteten³⁴ Arbeitsplätzen möglich.

In Abschnitt 3.1 wurde bereits darauf hingewiesen, dass der Zugang zu Sozialdaten an bestimmte Voraussetzungen gebunden ist. Der bei Gastaufenthalten im FDZ gewährte Zugang zu Sozialdaten stellt eine Übermittlung im Sinne des §75 SGB X dar. Im Unterschied zu der externen Übermittlung

³³ Die Testdaten zum IAB-Betriebspanel und weitere Arbeitshilfen (Codebuch, Fragebögen, Programmierungs- und Rekodierungsbeispiele) zur Erstellung fehlerfreier Syntaxen zur Analyse des IAB-Betriebspanels sind zugänglich auf den Internet-Seiten des FDZ unter <http://fdz.iab.de/pageText.asp?PageID=64>.

³⁴ Als ‚abgeschottet‘ wird ein Arbeitsplatz bezeichnet, von dem aus weder ein Zugriff auf die Server des IAB oder der BA möglich ist, noch ein Zugang zu Internet und E-Mail besteht. Von den Arbeitsplätzen der Gastwissenschaftler im FDZ kann ausschließlich auf spezifische Verzeichnisse auf dem FDZ-Server zugegriffen werden, Zugänge für externe Speichermedien (Disketten- und CD-Rom-Laufwerke, USB-Anschlüsse) sind gesperrt.

von Sozialdaten, die bei dem Datenzentrum der Statistik der BA zu beantragen ist (vgl. Abschnitt 3.1) bedürfen Gastaufenthalte im FDZ jedoch der Erfüllung weitaus geringerer Bedingungen. Durch die Datenhaltung im FDZ³⁵ und die Formulierung von Datenmodellen entfällt für Forscher die Einrichtung umfangreicher Sicherheitsvorkehrungen an ihrem Forschungsinstitut, das beim BMWa zu durchlaufende Genehmigungsverfahren kann erheblich verkürzt werden und dem Antragsteller entstehen keine Kosten für die Datenerstellung. Neben der Beschleunigung des Datenzugangs im Vergleich zur externen Übermittlung der Sozialdaten ist vor allem bei komplexen Daten die individuelle Betreuung der Gastwissenschaftler durch das FDZ von Vorteil. Gastaufenthalte im FDZ sind seit Anfang 2005 für die LIAB-Modelle und seit Frühjahr 2005 für die Originalstichproben der IAB-Beschäftigtenstichprobe und das IAB-Betriebspanel möglich.

Der Datenzugang im Rahmen von Gastaufenthalten ist wie folgt gestaltet: Interessenten stellen auf der Basis des kommentierten Antragsformulars³⁶ einen Antrag auf den Zugang zu den schwach anonymisierten bzw. pseudonymisierten Daten der Originalstichprobe³⁷ nach § 75 SGB X.

Dieser Antrag enthält neben den in Abschnitt 3.1 dargestellten Anforderungen für die Übermittlung von Sozialdaten nach §75 SGB X die Spezifizierung des benötigten Datenmaterials, die Namen der Personen, die Zugang zu den Daten erhalten sollen (dies können auch mehrere Personen sein) und eine Beschreibung des Forschungsdesigns.³⁸

Der gestellte Antrag wird im FDZ in Kooperation mit dem BMWa auf die Einhaltung der Bedingungen (vgl. Abschnitt 3.1) überprüft. Nach Antragsbewilligung wird zwischen dem Antragsteller und dem FDZ ein Vertrag geschlossen, der zum Zugang zu den Sozialdaten berechtigt. Anschließend wird ein Termin für einen Gastaufenthalt von maximal zwei Wochen vereinbart.³⁹ Der Nutzer erstellt im Vorfeld des Gastaufenthalts auf der Basis von

³⁵ Das FDZ klärt im Vorfeld des Gastaufenthalts mit den für Datenschutz zuständigen Stellen in der BA und am IAB sowie dem BMWa die rechtliche Zulässigkeit der Datenhaltung, der Zugangsregelungen und Zugriffsmöglichkeiten wie auch die nachfolgenden Datenschutzprüfungen. Dieses Verfahren ist für jeden Datenbestand, der im FDZ für die Analyse durch Externe zur Verfügung gestellt wird, durchzuführen.

³⁶ Antragsformulare sind datensatzspezifisch auf den jeweiligen Internet-Seiten des FDZ zu finden, für den LIAB z.B. unter http://doku.iab.de/fdz/liab/LIAB_Antrag.pdf.

³⁷ Im Fall der LIAB-Modelle wird hiermit auch gleichzeitig der Antrag auf Übermittlung der Testdaten (Personendaten) und auf die Übermittlung der detaillierten Datendokumentationen gestellt. Diese werden dem Antragsteller nach Abschluss des Vertrags zugesendet.

³⁸ Bei den LIAB-Modellen sind zudem die sensiblen Merkmale auf der Personenebene (insofern benötigt, vgl. Abschnitt 2.4) zu benennen und der Bedarf zu begründen.

³⁹ Die Angabe von zwei Wochen bezieht sich auf einen Richtwert für jeden durchgeführten Gastaufenthalt, diese sind prinzipiell fortführbar bzw. wiederholbar. Die realisierte Aufenthaltszeit im FDZ ist von verschiedenen Faktoren abhängig: Bei geringen Vorkenntnissen und komplexen Datensätzen (z.B. LIAB) ist von längeren Aufenthaltszeiten auszugehen, bei größeren Vorkenntnissen (z.B. durch die Datenfernverarbeitung oder Scientific Use Files) und einfacheren Datenstrukturen dagegen von einem kürzeren Gastaufenthalt.

Testdaten⁴⁰ bzw. Scientific Use Files und Datendokumentationen lauffähige und dokumentierte Auswertungsprogramme in SPSS, STATA oder in SAS und übersendet diese an das FDZ. Mit diesem Vorgehen wird eine effizientere Nutzung des Zugangs zu den Daten im FDZ angestrebt. Der Gastaufenthalt wird unter Einhaltung der Richtlinien für Gastaufenthalte im FDZ durchgeführt⁴¹. Die Verknüpfung der im FDZ zugänglich gemachten Sozialdaten mit anderen, externen Mikrodaten ist nicht zulässig. Die Verknüpfung der Sozialdaten mit aggregierten Kenngrößen aus externen Quellen bedarf der vorherigen Klärung mit dem FDZ. Die Analyseergebnisse werden durch das FDZ datenschutzrechtlich geprüft und spätestens fünf Werktage nach Beendigung des Aufenthalts an den Nutzer übersandt. Sofern die Analyseergebnisse publiziert werden, senden die Forscher zwei Belegexemplare der Veröffentlichung an das FDZ.

Der Zugang zu den Sozialdaten im FDZ erfolgt kostenfrei. Unterkunft- und Verpflegungskosten tragen die Gastforscher selbst. Zwischen der Antragstellung und der Genehmigung des Datenzugangs liegen ca. drei Wochen. Der Zeitpunkt der Realisierung des Datenzugangs ist u.a. abhängig von der Verfügbarkeit freier Arbeitsplätze für Gastwissenschaftler und dem gewünschten Termin.

4. AUSBLICK

Gegenstand unseres Beitrags war die Darstellung der durch das FDZ der BA im IAB für die wissenschaftliche Forschung bereitgestellten Daten und die Beschreibung der verschiedenen Zugangswege zu diesen Daten und ihres rechtlichen Hintergrunds.

Bei der Verfolgung seines zentralen Ziels, der Gewährung und der Erleichterung des Zugangs zu den amtlichen Daten über den Arbeitsmarkt konzentriert sich das FDZ bislang auf die Bereitstellung *integrierter* Datensätze. Dies bedeutet, dass - wie auch schon bei den im Vorfeld des FDZ erstellten Scientific Use Files der IABS und des BA-Beschäftigtenpanels - die Verknüpfung von Datenmengen mit dem Ziel möglichst großer Analysepotenziale dominiert. Neue integrierte Daten liegen z.B. mit den LIAB-Modellen, die seit 2005 für Gastforscher im FDZ nutzbar sind, vor und werden gegenwärtig mit Arbeiten zu Stichproben der Integrierten Erwerbsbiographien fortgesetzt.

Darüber hinaus können jedoch auch - wie in Oertel, Passenberger und Janser (2004) beschrieben - die Quellen der geplanten integrierten Erwerbsbiographien für die Forschung bereitgestellt werden. Hierzu zählen vor allem Daten zu offenen Stellen, Arbeitsuchenden und Maßnahmen sowie Maßnahmeteilnahmen der aktiven Arbeitsförderung. Eine Priorisierung der Da-

⁴⁰ Informationen zu den Testdaten der LIAB-Modelle sind erhältlich auf den Internet-Seiten des FDZ unter <http://fdz.iab.de/pageText.asp?PageID=75>.

⁴¹ Die Richtlinien zum Gastaufenthalt sind auf den Internet-Seiten des FDZ unter http://doku.iab.de/fdz/Richtlinien_Gastaufenthalte.pdf zu finden.

tenbereitstellung erfolgt 2005 im Nachgang einer Befragung (potenzieller) Nutzer und bedarf der datenschutzrechtlichen Genehmigung.

Bei der Datenbereitstellung für die Wissenschaft steht das FDZ in mehrfacher Hinsicht vor wichtigen Herausforderungen. Einerseits bewegt sich das FDZ in einem Spannungsfeld zwischen datentechnischen Möglichkeiten (und damit auch Erfordernissen und Wünschen der Wissenschaft insgesamt) und den rechtlichen Grundlagen zur Wahrung des Sozialgeheimnisses. Beim versucht das FDZ gerecht zu werden. Hierbei sind die Anbindung des FDZ an das IAB und seinen wissenschaftlichen Beirat sowie Anregungen von Nutzerseite u.a. auf vom FDZ durchgeführten Nutzerkonferenzen und Workshops⁴² von großer Hilfe. Gleichmaßen wird das FDZ durch die für Datenschutz zuständigen Stellen in der BA und am IAB rechtlich beraten.

Eine zweite Herausforderung besteht darin, im Hinblick auf eine schnelle Bereitstellung von Daten oder der Bereitstellung von im Sinne der Datenqualität „sicheren“ Daten Prioritäten zu setzen. Um Unsicherheiten und grundlegende Fehler bei der Bearbeitung der Daten zu vermeiden, werden die für die Bereitstellung geplanten Daten IAB-intern vor der Bereitstellung geprüft⁴³. Das FDZ stellt darüber hinaus im Rahmen der von ihm herausgegebenen Online-Publikationen „FDZ-Datenreporte“ und „FDZ-Methodenreporte“⁴⁴ Informationen zur Validität der Daten bereit. Von großer Bedeutung sind im Hinblick auf die Qualitätssicherung der Daten aber auch Anregungen von Nutzerseite, die bei Veranstaltungen des FDZ, bei Gastaufenthalten oder bei der individuellen Beratung gerne aufgenommen werden. Gleichmaßen bedeutet diese Vorgehensweise, dass sich die Datenbereitstellung verlangsamt und nur sukzessive möglich ist.

Die dritte Herausforderung, vor der das FDZ steht, bezieht sich auf die Priorisierung der Bereitstellung der Daten selbst oder der Dokumentationen über die Daten. Da die vorhandenen Dokumentationen zu den Daten der BA und des IAB unsystematisch sind und nur unvollständig vorliegen, besteht eine der zentralen Aufgaben des FDZ in ihrer Prüfung, Systematisierung und Ergänzung. Aufgrund der Komplexität der Daten und ihrer Entstehung, die sich in Abbildung 1 (vgl. Abschnitt 2) widerspiegelt, sind undokumentierte Daten für die Fachöffentlichkeit nahezu unanalysierbar. Das FDZ geht zur Lösung dieses Problems einen Mittelweg: Auf dem Weg zu ausführlichen Datensatzbeschreibungen in den „FDZ-Datenreporten“ werden bereits vorliegende Dokumente im Rahmen von datensatzspezifischen „Nutzer-Guides“ im Internet-Angebot des FDZ zur Verfügung gestellt. Damit wird es potenziellen Interessenten der Daten ermöglicht, entscheiden zu

⁴² Informationen zu vom FDZ durchgeführten Veranstaltungen finden sich im Internet unter <http://fdz.iab.de/pageText.asp?PageID=30>.

⁴³ Eine interne Prüfung und Diskussion erfolgte z.B. auch bei der Erstellung der LIAB-Modelle durch die Diskussion im IAB-Arbeitskreis „Linked-Employer-Employee-Daten des IAB“.

⁴⁴ Die vom FDZ herausgegebenen Reihen sind im Internet-Angebot des FDZ unter <http://fdz.iab.de/pageText.asp?PageID=44> verfügbar.

können, welcher Datensatz für ihre Forschungsfrage relevant ist und welche Analysen sie mit ihm durchführen können.

Eine vierte und letzte Herausforderung des FDZ besteht in dem Schutz der Forscher selbst. Dieser Schutz bezieht sich einerseits auf persönliche Informationen, die auf dem Wege der Beantragung des Datenzugangs beim FDZ gesammelt werden als auch auf den Einblick in die Forschungsaktivitäten der Wissenschaftler, den die Mitarbeiter des FDZ zwangsläufig erhalten. Das FDZ gewährt den externen Forschern hier einen besonderen Vertrauensschutz: Die Mitarbeiter des FDZ sind verpflichtet, ihren Einblick in Forschungsfragen, Methoden und Analysen der Datennutzer nur zum Zwecke der Beratung, der Verbesserung des Service des FDZ sowie zur Gewährleistung der Einhaltung des Datenschutzes zu nutzen. Mitarbeiter des IAB und der BA ebenso wie andere Nutzer der Daten erhalten keinen Einblick in die Tätigkeiten von externen Wissenschaftlern. Kooperationsprojekte zwischen Mitarbeitern des IAB und des FDZ einerseits und externen Forschern sind hiervon selbstverständlich ausgenommen.

LITERATUR

- ALDA, H. (2005a). Betriebe und Beschäftigte in den Linked-Employer-Employee-Daten des Instituts für Arbeitsmarkt- und Berufsforschung (LIAB). *FDZ-Datenreport No. 1*.
- ALDA, H. (2005b). Die Verknüpfungsqualität der LIAB-Daten. *FDZ-Methodenreport No. 1*.
- ALDA, H. (2005c). Datenbeschreibung der Version 1 des LIAB-Querschnittmodells. *FDZ-Datenreport No. 2*.
- ALDA, H. (2005d). Datenbeschreibung der Version 1 des LIAB-Längsschnittmodells. *FDZ-Datenreport No. 3*.
- 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). *IAB Discussion Paper 3/2005*.
- BELLMANN, L. (2002). Das IAB-Betriebspanel. Konzeption und Anwendungsbeispiele. *Allgemeines Statistisches Archiv* **86** 177-188.
- BELLMANN, L., BENDER, S., KÖLLING, A. (2002). Der Linked-Employer-Employee-Datensatz aus IAB-Betriebspanel und Beschäftigtenstatistik der Bundesanstalt für Arbeit (LIAB). *IAB-Kompodium Arbeitsmarkt- und Berufsforschung, Beiträge zur Arbeitsmarkt- und Berufsforschung* **250** 21-29.
- BELLMANN, L., KOHAUT, S., LAHNER, M. (2002). Das IAB-Betriebspanel – Ansatz und Analysepotenziale. *IAB-Kompodium Arbeitsmarkt- und Berufsforschung, Beiträge zur Arbeitsmarkt- und Berufsforschung* **250** 13-20.
- BENDER, S., HAAS, A. (2002). Die IAB-Beschäftigtenstichprobe. *IAB - Kompodium Arbeitsmarkt- und Berufsforschung, Beiträge zur Arbeitsmarkt- und Berufsforschung* **250** 3-12.
- BENDER, S., HAAS, A., KLOSE, C. (1999). Die IAB-Beschäftigtenstichprobe 1975-1995. *ZA-Information* **45** 104-115.

- BENDER, S., HAAS, A., KLOSE, C. (2000a). The IAB Employment Subsample 1975-1995. *Schmollers Jahrbuch* **120** 649-662.
- BENDER, S., HAAS, A., KLOSE, C. (2000b). IAB Employment Subsample 1975-1995. Opportunities for analyses provided by the anonymised subsample. *IZA Discussion Paper* **117** .
- BRAND, R., BENDER, S., KOHAUT, S. (1999). Möglichkeiten der Erstellung eines Scientific Use Files aus dem IAB-Betriebspanel. *Möglichkeiten einer wissenschaftlichen Nutzung von Unternehmensdaten aus der amtlichen Statistik (Statistisches Bundesamt, Hrsg.), Schriftenreihe Spektrum Bundesstatistik* **14** 148-167.
- BUNDESGESETZBLATT (BGB1) (1972). Teil I. (Der Bundesminister der Justiz, Hrsg.), Bundesanzeiger-Verlag, Bonn.
- BUNDESGESETZBLATT (BGB1) (1999). Teil I. (Der Bundesminister der Justiz, Hrsg.), Bundesanzeiger-Verlag, Bonn.
- BÜTEFISCH, T. (2004). Datenwege und praktischer Datenzugang. *Das Forschungsdatenzentrum der gesetzlichen Rentenversicherung (FDZ-RV) im Aufbau. DRV-Schriften* **55** 20-23.
- HAMANN, S., KRUG, G., KÖHLER, M., LUDWIG-MAYERHOFER, W., HACKET, A. (2004). Die IAB-Regionalstichprobe 1975-2001: IABS-R01. *ZA-Information* **55** 34-60.
- KOCH, I., MEINKEN, H. (2004). The Employment Panel of the German Federal Employment Agency. *Schmollers Jahrbuch* **124** 315-325.
- KOMMISSION ZUR VERBESSERUNG DER INFORMATIONELLEN INFRASTRUKTUR ZWISCHEN WISSENSCHAFT UND STATISTIK (KVI) (2001). Wege zu einer besseren informationellen Infrastruktur. Nomos, Baden-Baden.
- LÜTTINGER, P., SCHIMPL-NEIMANNS, B., WIRTH, H., PAPASTEFANOU, G. (2004). The German Microdata Lab at ZUMA: Services Provided to the Scientific Community. *Schmollers Jahrbuch* **124** 455-467.
- MÜLLER, W., BLIEN, U., KNOCH, P., WIRTH, H. (1991). Die faktische Anonymität von Mikrodaten. *Schriftenreihe Forum der Bundesstatistik* (Statistisches Bundesamt, Hrsg.), Metzler-Poeschel, Stuttgart.
- OERTEL, M., PASSENBERGER, J., JANSER, M. (2004). Datenservice: Forschungsdatenzentrum der BA geht an den Start. *IAB-Kurzbericht* **09**.
- REHFELD, U. G. (2004). Datenangebot und Informationsbedarf im Bereich der Alterssicherung. Forschungsdatenzentrum der Rentenversicherung im Aufbau. *Deutsche Rentenversicherung* **1-2** 63-75.
- ZÜHLKE, S., ZWICK, M., SCHARNHORST, S., WENDE, T. (2003). Die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder. *Wirtschaft und Statistik* **10** 906-911.

Prof. Jutta Allmendinger, Ph.D.
 Institut für Arbeitsmarkt- und Berufsforschung
 Regensburger Straße 104
 90478 Nürnberg
 Deutschland
 jutta.allmendinger@iab.de

Dr. Annette Kohlmann
 Institut für Arbeitsmarkt- und Berufsforschung
 Regensburger Straße 104
 90478 Nürnberg
 Deutschland
 annette.kohlmann@iab.de