

## **Betriebe und Beschäftigte in den Linked-Employer-Employee-Daten**

**LIAB des Instituts für Arbeitsmarkt- und Berufsforschung**

*Holger Alda*



## Inhaltsverzeichnis

1. Einleitung .....	3
2. Die Betriebs- und Beschäftigtendaten .....	3
3. Verknüpfung der Betriebs- und Beschäftigtendaten: die LIAB-Datenmodelle .....	6
4. Vor- und Nachteile der Datenmodelle .....	12
5. Die technische Umsetzung der LIAB-Daten im FDZ: von Modellen zu Versionen ...	13
6. Zusammenfassung.....	16

## Datenverfügbarkeit

Der in diesem Beitrag beschriebene Datensatz ist für die Fachöffentlichkeit zugänglich. Nähere Informationen dazu auf der Internetseite: <http://fdz.iab.de/> unter der Rubrik „Integrierte Betriebs- und Personendaten“.

## Danksagung

Ohne den Service-Bereich ITM des Instituts für Arbeitsmarkt- und Berufsforschung und seine umfangreichen Vorarbeiten und Hilfestellungen wäre die Umsetzung und Bereitstellung der in den ersten drei Datenreporten abgehandelten Linked-Employer-Employee-Daten des IAB, LIAB, nicht möglich gewesen. Namentlich wird Martina Oertel, Steffen Kaimer und Wolfgang Majer gedankt. Dem Forschungsbereich „Betriebe und Beschäftigung“ des IAB unter der Leitung von PD Dr. Lutz Bellmann wird ebenfalls für die Unterstützung gedankt. Stefan Bender und Dr. Hermann Gartner (beide Kompetenzzentrum Empirische Methoden) haben das LIAB-Querschnittmodell konzipiert und die Umsetzung der LIAB-Datenmodelle durch das Forschungsdatenzentrum der Bundesagentur für Arbeit im IAB maßgeblich unterstützt. Der LIAB-Arbeitsgruppe wird für die Fülle an Anregungen im Verlauf der Erarbeitung der LIAB-Datenmodelle und der Datendokumentationen gedankt. Für die Inhalte der FDZ-Daten- und Methodenreporte sind die jeweiligen Autoren verantwortlich.

---

## 1. Einleitung

Linked-Employer-Employee-Daten im **IAB** (LIAB) sind aufeinander bezogene Arbeitgeber-Arbeitnehmer Daten. Sie bestehen aus zwei Einzelteilen, den verschiedenen Wellen des IAB-Betriebspanels und ein bis mehrerer Personendatenbanken. Durch die Verknüpfung beider Datenquellen entstehen Linked-Employer-Employee-Daten. Die Personendatenbanken entsprechen in ihrem Abdeckungsgrad, dem Aufbau und Inhalt einzelner Variablen gängigen IAB-Personendaten wie beispielsweise der IAB-Beschäftigtenstichprobe, verfügen aber über einen differenzierteren Merkmalskanon. Der Report beschreibt im zweiten Abschnitt die generellen Verknüpfungsmöglichkeiten der IAB-Personendaten mit dem IAB-Betriebspanel. Dabei wird auch auf technische Restriktionen und die Bereitstellungsmöglichkeiten im Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung eingegangen, die sich aus den Datenschutzbestimmungen ergeben. Dies mündet in so genannte Datenmodelle und Versionen<sup>1</sup>. Der Beitrag schließt mit einer Zusammenfassung und einem Ausblick auf zukünftige Aktivitäten.

## 2. Die Betriebs- und Beschäftigtendaten

Abowd/Kramarz (1999)<sup>2</sup> geben einen Überblick über die grundsätzlichen Verknüpfungsmöglichkeiten von Betriebs- und Beschäftigtendaten. Sie unterscheiden u.a. zwischen Verknüpfungen ausschließlich administrativer Daten, ausschließlich von Befragungsdaten und diversen Mischformen. Für die hier verfügbaren LIAB-Daten geben Alda/Bender/Gartner<sup>3</sup> einen Überblick über mögliche Anwendungsgebiete und die bisher mit LIAB-Daten erstellten Untersuchungen. Die Verknüpfungsmöglichkeiten der LIAB-Daten im Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung ergeben sich aus den beiden Elementen der LIAB-Daten, den Betriebs- und Personendaten.

---

<sup>1</sup> Für einzelne LIAB-Versionen (vgl. Abschnitt 5) erscheinen je eigene Datenreporte, in denen die allgemeinen Charakteristika der einzelnen Personendatenbanken beschrieben werden.

<sup>2</sup> Abowd, J./Kramarz, F. (1999): The analysis of labor markets using matched employer-employee data. In: Ashenfelter, O./Card, D. (Hrsg.): Handbook of Labor Economics, Vol. 3B, Elsevier, Amsterdam, chapter 40, pp. 2567 - 2627

<sup>3</sup> Alda, H./Bender, St./Gartner, H. (2004): Der Linked-Employer-Employee-Datensatz aus IAB-Betriebspanel und den prozessproduzierten Daten des IAB (LIAB). IAB-Discussion-Paper No.6/2005, Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg.

---

Die Personendaten stammen aus der Beschäftigten-Leistungsempfänger-Historik-Datei (BLH) des IAB. Die *Beschäftigtenhistorik* als ein Teil dieser Datei deckt dabei die sozialversicherungspflichtige *Beschäftigung* ab. Für einzelne Arbeitnehmer geben Arbeitgeber mindestens einmal im Jahr eine Meldung ab. Diese Meldungen enthalten u.a. für Beschäftigte

- ihr Alter, Geschlecht und Nationalität
- das Beginn- und Endedatum jeder Beschäftigtenmeldung
- den ausgeübten Beruf auf der Dreisteller-Ebene
- den tagesgenauen Lohn zensiert an der oberen und unteren Beitragsbemessungsgrenze
- das Bildungsniveau in Verbindung mit dem beruflichen Qualifikationsniveau
- den Wirtschaftszweig
- Kennziffern zum Arbeitsort
- einen eindeutigen Betriebsidentifikator

Indirekt lassen sich auch Beschäftigte in Freistellungszeiten identifizieren. Diese haben in der Variable Tagesentgelt einen Wert von Null eingetragen. Die Beschäftigtenhistorik-Datei wird im IAB mit Leistungsempfangsdaten, die aus Geschäftsstatistiken der Bundesagentur für Arbeit erstellt wird, zur *Beschäftigten-Leistungsempfänger-Historik-Datei* (BLH) verknüpft. Zeitliche Lücken der Beschäftigtenmeldungen werden durch Meldungen aus der *Leistungsempfänger-Historik-Datei* (LEH) des IAB aufgefüllt. Dann noch verbleibende zeitliche Lücken sind mit den vorliegenden Daten nicht beobachtbar („nicht gemeldete Zeiträume“, NGZ).

In der LEH werden Arbeitslosigkeitszustände nur erfasst, wenn Personen Leistungen der Bundesagentur für Arbeit in Anspruch nehmen. Gemeldete Arbeitslosigkeit ohne Anspruch auf Bezüge (z.B. nach einem Studium, bei Hilfe zum Lebensunterhalt oder nach Selbstständigkeit) ist in der LEH derzeit nicht beobachtbar. Die Leistungsempfänger-Historik-Datei enthält u.a. folgende Informationen:

- Alter, Geschlecht, Nationalität, Familienstand und Kinder (ja/nein) der Person
  - Kreiskennziffer des zahlenden Arbeitsamtes
  - Art und Höhe der Leistung
  - Beginn und Ende jeder Leistungsmeldung
  - Bewilligungsart
-

- Abgabegrund der Meldung
- Beendigungsgrund
- den Beruf auf der Viersteller-Ebene

Leistungsmeldungen erfolgen nur bei leistungsrechtlich relevanten Veränderungen. Eine einzelne Leistungsmeldung kann also im Gegensatz zu den Beschäftigtenmeldungen einen längeren Zeitraum als ein Jahr beschreiben.

Auf der Betriebsseite steht als Datenbasis das IAB-Betriebspanel zur Verfügung. Das IAB-Betriebspanel ist eine im Jahr 1993 erstmals und seitdem jährlich durchgeführte Arbeitgeberbefragung, zunächst in etwa 4000 westdeutschen, seit 1996 auch in etwa 4300 ostdeutschen Betrieben. Ab dem Jahr 2000 sind auf nahezu jeder Bundeslandebene repräsentative Auswertungen möglich, denn seitdem finanzieren auch westdeutsche Bundesländer entsprechende Aufstockungsstichproben. Im Jahr 2001, dem letzten Beobachtungsjahr der derzeit vorliegenden Personendaten, haben mehr als 15.000 Betriebe an der Befragung teilgenommen.

Das IAB-Betriebspanel enthält u.a. die folgenden Informationen:

- Beschäftigtenzahl (auch für verschiedene Qualifikationsgruppen)
- Anzahl befristet Beschäftigter und Leiharbeiter im Betrieb
- Wochenarbeitszeit für Vollzeitbeschäftigte und Überstunden
- Tarifbindung des Betriebs, Existenz eines Betriebsrat
- Umsatz, Vorleistungen und Exportanteil
- Investitionssumme
- Gesamtlohnsumme im Juni des Befragungsjahres
- technischer Stand der Betriebsanlagen
- Betriebsalter, Rechtsform und Unternehmensstellung
- Beurteilung der betrieblich-wirtschaftlichen Gesamtsituation
- Reorganisationsmaßnahmen und betriebliche Weiterbildungsaktivitäten (in mehrjährigen Abständen)
- Wirtschaftszweizugehörigkeit

Den Ansatz und Aufbau des IAB-Betriebspanels beschreibt beispielsweise Bellmann (2002)<sup>4</sup>. Im nächsten Abschnitt werden Eigenschaften des IAB-Betriebspanels ange-

---

<sup>4</sup> Bellmann, L. (2002): Das IAB-Betriebspanel - Konzeption und Anwendungsbereiche, Allgemeines Statistisches Archiv 86, 177 - 188

sprochen, die für die Verknüpfung der Betriebs- und Personendaten eine wichtige Rolle spielen.

### 3. Verknüpfung der Betriebs- und Beschäftigendaten: die LIAB-Datenmodelle

Die Grundgesamtheit der Betriebe in Deutschland ist in der Betriebsdatei des IAB erfasst. Aus ihr wird eine bezüglich der Betriebsgröße disproportionale Stichprobe für das IAB-Betriebspanel gezogen<sup>5</sup>, die Großbetriebe überrepräsentiert. Durch Hochrechnung der Analyseergebnisse ist der Schluss auf die Grundgesamtheit möglich<sup>6</sup>.

Die Tabelle 1 zeigt beispielhaft die Ziehungswahrscheinlichkeiten nach der Betriebsgröße in der ersten Befragungswelle 1993.

**Tabelle 1: Ziehungswahrscheinlichkeiten und Rücklaufquote des IAB-Betriebspanels in der ersten Befragungswelle 1993 (nur westdeutsche Betriebe)**

Beschäftigtenzahl	gezogene Betriebe	Auswahlwahrscheinlichkeit	antwortende Betriebe	Rücklaufquote
1 - 4	1,072	.0011	625	.67
5-9	431	.0015	250	.64
10-19	466	.0030	299	.71
20-49	862	.0089	542	.70
50-99	535	.0153	350	.72
100-199	543	.0304	376	.77
200-499	923	.0862	615	.74
500-999	479	.1504	304	.71
1.000 – 4.999	1,497	.8765	924	.72
5.000 +	115	.9127	71	.73
<b>insgesamt</b>	<b>6,923</b>	<b>.0043</b>	<b>4,356</b>	<b>.71</b>

Quelle: IAB-Betriebspanel 1993, Kölling (2000), S. 294<sup>7</sup>

<sup>5</sup> Die Schichtungsmatrix berücksichtigt neben zehn Betriebsgrößenklassen 16 Wirtschaftszweige bis zum Jahr 1999 und 20 Wirtschaftszweige ab dem Jahr 2000.

<sup>6</sup> Auch deskriptive Analyseergebnisse mit den LIAB-Personendaten können mit den Hochrechnungsfaktoren des IAB-Betriebspanels gewichtet werden.

<sup>7</sup> Kölling, A. (2000): European Datawatch: The IAB-Establishment-Panel, Schmollers Jahrbuch, 120 Jg., 291 - 300

An den Ziehungswahrscheinlichkeiten und Rücklaufquoten ändert sich im Verlauf der Jahre relativ wenig<sup>8</sup>. Die Fallzahl des IAB-Betriebspanels beträgt im Jahr 2001 beispielsweise mehr als 15500 Betriebe (etwa ein Prozent aller Betriebe), in denen etwa zehn Prozent aller sozialversicherungspflichtig Beschäftigten in Deutschland arbeiten. Entsprechende Personendatenbanken erreichen also im Zeitverlauf beachtliche Größen. Zwei Datenmodelle werden im Folgenden näher beschrieben.

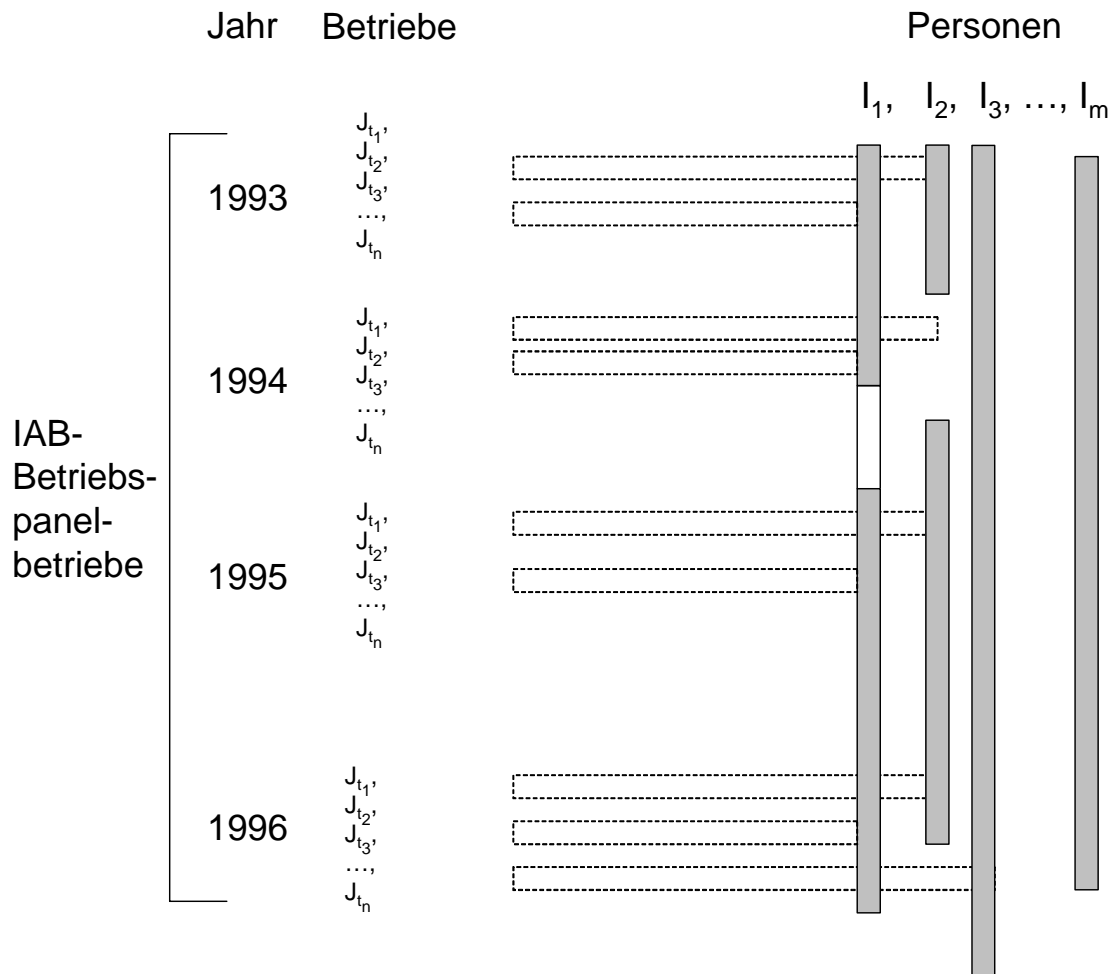
Zum einen können alle Beschäftigungs- und Leistungsmeldungen für einen bestimmten Zeitraum in Personendatenbanken abgelegt werden. Sie umfassen derzeit maximal das Jahr 1990 bis zum aktuellen Rand. Für jedes Individuum gibt es – je nach gemeldeten Vorgängen – eine beachtliche Anzahl an einzelnen Meldungen oder Spells. Zum anderen besteht die Möglichkeit, den jeweils zu einem bestimmten Zeitpunkt eines jeden Jahres gültigen Beschäftigungsspell aus der BLH zu extrahieren. Meistens handelt es sich um eine Beschäftigung am 30. Juni in einem der IAB-Betriebspanelbetriebe, weil dieses Datum der dem Referenzzeitpunkt der meisten Fragen im Fragebogen ist. Variablen in der Personendatenbank, die Leistungsbezug durch die Bundesagentur für Arbeit abbilden, werden bei der letztgenannten Datenstruktur obsolet (vgl. die Ausführungen zu einzelnen geplanten Versionen in Abschnitt 5). Mit diesen so genannten *Jahresscheiben* wird die Anzahl der einzelnen (Personen-)Beobachtungen reduziert. In der Abbildung 1 werden die beiden angesprochenen Modelle graphisch verdeutlicht.

---

<sup>8</sup> Ausnahme sind die größten Betriebe. Panelmortalität kann in den beiden obersten Größenklassen kaum ausgeglichen werden. Die Anzahl der verknüpfbaren Personendaten bzw. –konten nimmt daher bei einer steigenden Anzahl von Betrieben aufgrund großbetrieblicher Panelmortalität ab bzw. bleibt trotz diverser Aufstockungsstichproben auf der Betriebsseite nahezu unverändert.

---

Abbildung 1: Betriebe und Beschäftigte in LIAB-Daten



Die horizontalen Linien stellen befragte IAB-Betriebspanelbetriebe in dem jeweiligen Jahr dar, die vertikalen Personen. Grau hinterlegte Balkenabschnitte kennzeichnen Zeiten der Beschäftigung und hell hinterlegte gemeldete Arbeitslosigkeit. Ein unterbrochener Balken bedeutet, dass die Person weder ein sozialversicherungspflichtiges Beschäftigungsverhältnis ausübt noch Leistungen der BA in Anspruch nimmt.

Eine Jahresscheibe ist der Kreuzungspunkt der horizontalen und vertikalen Balken, d.h. die dazu gehörige Beschäftigtenmeldung zum 30. Juni. Für das Jahr 1996 stellt der längere untere horizontale Balken einen ostdeutschen Betrieb dar, der in diesem Jahr erstmalig befragt wird. Individuum I<sub>1</sub> lässt sich somit in jeder Jahresscheibe 1993 bis 1996 beobachten, eventuell in verschiedenen Betrieben. Der Leistungsbezug ist auf der Datenebene nicht identifizierbar. Person I<sub>2</sub> hat eine unterbrochene sozialversicherungspflichtige Erwerbsbiographie und kann im Jahr 1994 nicht beobachtet werden. Individuum I<sub>3</sub> ist in einem ostdeutschen Betrieb beschäftigt und daher erstmalig im Jahr



1996 beobachtbar. Der Balken für  $I_m$  soll graphisch verdeutlichen, dass es durchaus auch Beschäftigte gibt, die gar nicht in einem der IAB-Betriebspanelbetriebe auftauchen. Leistungsmeldungen bzw. die entsprechenden Variablen werden in Personendatenbanken beim Jahresscheibenmodell obsolet, denn Leistungsbezug ist durch die Bedingung, am 30. Juni in einem IAB-Betriebspanelbetrieb beschäftigt zu sein, unmöglich.

Der Unterschied zu einer Personendatenbank mit allen Beschäftigten- und Leistungsmeldungen einer Person besteht darin, dass auch Nicht-Kreuzungspunkte und die entsprechenden Erwerbszustände bestimmt werden können, also beispielsweise gemeldete Arbeitslosigkeit mit Leistungsbezug. So ist für Person  $I_1$  die Arbeitslosigkeitsphase identifizierbar und für Person  $I_2$  bekannt, dass sie im Jahr 1994 eine Erwerbsunterbrechung hat. Das ist mit Jahresscheiben nicht beobachtbar. Der Kreuzungspunkte einer Person und eines Betriebes sind, wenn alle Beschäftigten- und Leistungsmeldungen in der Personendatenbank vorliegen, Auswahlkriterien für die Aufnahme der entsprechenden Personenkonten. Es handelt sich nicht zwingend um den 30. Juni eines jeden Befragungsjahres, sondern oft ist die Bedingung für die Aufnahme einer Person und ihrer zugehörigen Meldungen eine mindestens eintägige Beschäftigung in einem der IAB-Betriebspanelbetriebe in einem zu spezifizierenden Zeitraum.

Wenn die Arbeit mit Personendatenbanken nicht möglich oder gewünscht ist, besteht eine weitere Analysemöglichkeit. Die Jahresscheibeninformationen der Beschäftigten können auf die Betriebsebene aggregiert werden, beispielsweise als Summen oder Anteile. Auf diese Weise werden Informationen auf der Betriebsebene dann auch ohne einen weiteren Rückgriff auf die Personendatenbanken verfügbar. Sie eignen sich zur Ergänzung der Daten des IAB-Betriebspanels. Im IAB-Betriebspanel gibt es beispielsweise in den meisten Jahren keine Information über den Anteil der Hochschulabsolventen an der Gesamtbelegschaft. Auch über die Alterszusammensetzung der Belegschaft ist bis zum Jahr 2002 keine genaue Information verfügbar. Analysen auf der Betriebsebene, die solche zeitvarianten Betriebscharakteristika nutzen, werden mit Aggregation der interessierenden Merkmale auf die Betriebsebene und der Hinzuspielung der IAB-Betriebspaneldaten möglich. Auf die Aggregation von Personendaten auf die Betriebsebene wird im weiteren Verlauf des Beitrags nicht mehr eingegangen.

---

Die anderen beiden Gruppen von Daten (Jahresscheiben und Verlaufsdaten von Personen) werden im Folgenden *LIAB-Querschnittsmodelle* und *LIAB-Längsschnittmodelle* genannt, wobei sich die Bezeichnungen auf die zugehörigen Personendatenbanken beziehen. LIAB-Querschnittmodelle sind also immer Personendatenbanken mit Beschäftigtenmeldungen zum 30. Juni eines jeden Befragungsjahres (in einem IAB-Betriebspanelbetrieb). LIAB-Längsschnittmodelle sind Personendatenbanken, in denen alle Beschäftigten- und Leistungsmeldungen für Personen in einem bestimmten Zeitraum vorliegen.

In den LIAB-Quer- und Längsschnittmodellen findet die Verknüpfung der Betriebs- und Personendaten in einem weiteren Arbeitsschritt statt, wo die (aufbereiteten) Betriebsinformationen zu den Personendaten dazugespielt werden. Die Tabelle 2 stellt allgemeine Eigenschaften beider Gruppen von Personendatenbanken zusammen. Die Aufstellung in Tabelle 2 wurde mit Rücksicht auf verfügbare Betriebsinformationen aus dem IAB-Betriebspanel (z.B. Alter des Betriebs) und der Informationen zu Personen aus der Beschäftigten-Leistungsempfänger-Historik-Datei vorgenommen und benennt die Quellen häufig verwendeter Betriebs- und Beschäftigtenmerkmale.

**Tabelle 2: Charakteristika potenzieller Linked-Employer-Employee-Datensätze als Verknüpfung von IAB-Personendaten mit Informationen aus dem IAB-Betriebspanel**

Nr.	Charakteristika	LIAB-Querschnittmodell	LIAB-Längsschnittmodell
1	Betriebs-ID im Personendatensatz	ja	
2	Beschäftigte und Betriebe sind in einem Datensatz vorhanden	nein	
3	Häufigkeit des Links	z.B. jährlich (30. 06.)	tagesgenaue Spells
4	Datenquelle	Beschäftigtenhistorik/ <i>Fragebogen</i>	
5	Betriebs-ID kann überprüft werden	ja	
6	Beobachtungsebene	Betrieb/ Beschäftigte gebunden an Betrieb*	Betrieb/ Person
7	Betriebsgröße muss gebildet werden bzw. wird berichtet	Bildung durch Aggregation der Beschäftigtendaten auf die Betriebsebene und Fragebogenangaben	
8	Quelle Betriebsgröße	Beschäftigtenhistorik/ <i>Fragebogen</i>	

(Fortsetzung nächste Seite)

Nr.	Charakteristika	LIAB-Querschnittmodell	LIAB-Längsschnittmodell
9	Alter des Betriebs bekannt	<i>≅ 90 Prozent</i>	
10	Frequenz der Betriebsbeobachtungen	<i>jährlich</i>	
11	Prozent der beobachtbaren Betriebe	<i>(ungewichtet) etwa 1 Prozent</i>	
12	Quelle Geburtsdatum	Sozialversicherungsnummer	
13	Quelle Mann/ Frau	Sozialversicherungsnummer	
14	Quelle schulische Bildung	Meldung der Arbeitgeber an Sozialversicherung	
15	Quelle Beruf	Meldung der Arbeitgeber an Sozialversicherung	
16	Einheit der Beschäftigtenhistorie	Person gebunden an Betrieb*	Person
17	Beginn des Beschäftigungsverhältnisses immer bekannt	nein	ja (ab 1990)
18	Stellung des Betriebs im Unternehmensverbund bekannt	<i>ja</i>	
19	Frequenz der Personenbeobachtungen	<i>jährlich</i>	tagesgenaue Spells
20	Prozent des beobachtbaren Beschäftigtenuniversums (gleich sozialversicherungspflichtige Beschäftigung)	gewichtet bis zu 90 Prozent, ungewichtet bis zu 10 Prozent	

Erläuterungen: \* mit gültigem IAB-Betriebspanel-Interview, kursive Schrift kennzeichnet Informationen, die im IAB-Betriebspanel verfügbar sind

Die ersten beiden Zeilen der Tabelle 2 beziehen sich auf die systemfreien Betriebsidentifikatoren in den Daten<sup>9</sup>. Die Daten des IAB-Betriebspanels können mit dem systemfreien Betriebsidentifikator eindeutig den Personendaten zugeordnet werden. Zeile drei verdeutlicht, dass die Verknüpfung der Personen- und Betriebsdaten eine eigenständig vom Anwender zu erbringende Leistung ist. Die Zeilen 3 bis 8 sowie 16, 17 und 20 fassen das zu den Datenmodellen Gesagte zusammen. Die Zeilen 10 bis 12 sowie 18 und 19 beziehen sich auf Merkmale, die das IAB-Betriebspanel bereitstellt bzw. auf Auswirkungen bei der Verknüpfung mit den Personendaten. Den disproportionalen Stichprobenansatz des IAB-Betriebspanels verdeutlicht Zeile 20.

<sup>9</sup> Systemfreie Identifikatoren sind fortlaufende und *eindeutige* Personen- und Betriebsidentifikationsnummern, die die originalen Sozialversicherungs- und Betriebsnummern ersetzen.

Die LIAB-Quer- und Längsschnittmodelle haben je eigene charakteristische Vor- und Nachteile. Die Tabelle 2 zeigt die grundsätzlichen Eigenschaften solcher Modelle bzw. das maximal Mögliche auf. Es sind auch Mischmodelle denkbar<sup>10</sup>. Im nächsten Abschnitt werden einige der grundlegenden Eigenschaften der beiden LIAB-Datenmodelle angesprochen.

#### **4. Vor- und Nachteile der Datenmodelle**

In diesem Abschnitt wird kurz auf grundlegende Implikationen und technische Restriktionen bei dem LIAB-Quer- und Längsschnittmodell eingegangen.

##### **LIAB-Querschnittmodell: Jahresscheiben (zum 30. Juni, derzeit von 1993 bis 2001)**

In Personendatenbanken sind die Daten der Beschäftigten in Betrieben abgelegt, die im jeweiligen Jahr an der Befragung des IAB-Betriebspanels teilgenommen haben. Es handelt sich dabei um die aktuelle Meldung der beschäftigten Personen am 30. Juni eines jeden Befragungsjahres. Darum enthalten die Personendatenbanken nur Daten aus Beschäftigungsmeldungen.

Da in den einzelnen Querschnitten ein eindeutiger Identifikator für Personen und Betriebe vorhanden ist, können die einzelnen Querschnitte miteinander verbunden werden. Somit ist es möglich, Verläufe von Personen in Panelbetrieben zu verfolgen. Ein Betriebswechsel der Person in einen nicht durch das Betriebspanel befragten Betrieb als auch die Nichtteilnahme des Betriebes an der Befragung im Folgejahr führen dazu, dass die Person nicht mehr in dieser Version des LIAB enthalten ist. Extensionen der LIAB-Querschnittsdaten entstehen durch Varianten der Jahresscheiben, beispielsweise durch die Bereitstellung von Vor- bzw. Nachgeschichte auf der Personenebene (vgl. Abschnitt 5).

Die Jahresscheiben sind ein Mittelweg in der Balance zwischen verfügbaren und generierbaren Informationen einerseits und der Größe des Datensatzes und der damit benötigten Rechenzeiten andererseits. Mit den Jahresscheiben ist nahezu jeder beliebig mögliche Aufbau von Paneldatensätzen unter den geschilderten Rahmenbedingungen möglich. Die Meldungen in den Jahresscheiben werden durch

---

<sup>10</sup> Beispielsweise können aus Längsschnitt-Personendaten Informationen für Jahresscheibendaten gewonnen werden. Das wird die dritte Version des LIAB-Querschnittmodells sein.

weitere Kennziffern der Beschäftigten (z.B. erstmaliges Eintrittsdatum in den Beschäftigungsbetrieb) ergänzt.

### **LIAB-Längsschnittmodell: auf der Personenebene vollständige Beschäftigten-Leistungsempfänger-Historikdatei-Daten**

LIAB-Längsschnittmodelle ermöglichen umfassende Analysen der Betriebs- und Beschäftigtenseite. Auf der Betriebsseite sind u.a. auch unterjährige betriebliche Personalanpassungsprozesse (theoretisch täglich) analysierbar. Auf der Beschäftigtenseite gibt es beispielsweise Informationen darüber, ob und wie lange eine Person arbeitslos oder nicht in den Arbeitsmarkt integriert war bzw. ist. Damit eine Verknüpfung mit den Betriebsdaten hergestellt werden kann, muss die Person wenigstens einen Tag in einem IAB-Betriebspanelbetrieb beschäftigt sein.

Die Personendatenbanken erreichen und überschreiten die Kapazitätsgrenze einiger gängiger Statistikprogramme – selbst bei der Bildung von Subsamples<sup>11</sup>. Die größten theoretisch denkbaren Datenbanken können nur mit wenigen Auswertungsprogrammen bearbeitet werden (SAS, TDA). Im Unterschied zum LIAB-Querschnittmodell sind in der Datenbank alle Beschäftigten- und Leistungsmeldungen einer Person in einem bestimmten Zeitraum abgelegt. Solche LIAB-Daten erlauben detaillierte Analysen mit individuellen Erwerbsverlaufsinformationen über einen längeren Zeitraum hinweg. Die Variablenaufbereitung kann anspruchsvoll sein und ist oft mit langen Rechen- und Datenaufbereitungszeiten verbunden.

## **5. Die technische Umsetzung der LIAB-Daten im FDZ: von Modellen zu Versionen**

In den beiden vorherigen Abschnitten wurde die Beschäftigten- und Betriebsseite der LIAB-Daten und ihre Eigenschaften beschrieben. Die Datenmodelle sind keine real existierenden Datensätze, sondern beschreiben die Bereitstellungsmöglichkeiten und grundsätzlichen Eigenschaften von Personendatenbanken. Auch aufgrund technischer Restriktionen (z.B. der Dateigröße) ist eine maximale Umsetzung der Datenmodelle insbesondere beim LIAB-Längsschnittmodell oft nicht möglich. Außerdem ist die Verknüpfungsqualität der Betriebs- und Personendaten diskussionswürdig. Nicht immer können IAB-Betriebspanelbetrieben auch Personendaten zugeordnet werden. Im FDZ-

---

Methodenreport Nr. 1 wird das zusammen mit anderen Verknüpfungsproblemen ausführlich diskutiert.

Der vorliegende Report ist eine Informationsgrundlage für die bestehenden und in Planung befindlichen Versionen. *Versionen sind Personendatenbanken*, die unterschiedlichen Auswahlkriterien bezüglich des abgedeckten Zeitraums, des Variablenkanons und der ausgewählten Betriebe unterliegen. Die Sozialversicherungs- und Betriebsnummer wird in jedem Datensatz durch fortlaufende und *eindeutige* Personen- und Betriebsidentifikationsnummern ersetzt (systemfreie Identifikatoren). Weitere Einschränkungen der Variablen und ihrer Inhalte gibt es nicht. Mit der Einrichtung der Gastwissenschaftlerarbeitsplätze im FDZ werden die ersten Versionen der LIAB-Datenmodelle zur Verfügung stehen. Die einzelnen Variablen der Personendatenbanken und Informationen zum Abdeckungsgrad der jeweiligen Version in Relation zum maximalen Abdeckungsgrad über Betriebe und Beschäftigte (vgl. Tabelle 2) werden in eigenen Datenreporten beschrieben. Für Gastwissenschaftler wird bei ihrem Aufenthalt im FDZ ein LIAB-Datenhandbuch zur Verfügung gestellt, das den Umgang mit den Gegebenheiten vor Ort erleichtert.

Zu Beginn des Jahres 2005 stehen im FDZ die drei folgenden Versionen zur Verfügung:

**LIAB-Querschnittmodell, Version 1:** Bedingung für die Aufnahme der Beschäftigtenmeldung in die Personendatenbank ist die sozialversicherungspflichtige Arbeitstätigkeit in einem IAB-Betriebspanelbetrieb zum 30. Juni. Die Personendatenbank enthält daher ausschließlich Variablen aus der Beschäftigtenhistorikdatei sowie Informationen zum ersten Eintrittsdatum der Beschäftigten in den Betrieb und in das Erwerbsleben. Austritte aus dem Betrieb führen auf der Personenebene und Austritte aus dem IAB-Betriebspanel auf der Betriebsebene zur Unbeobachtbarkeit der Betriebe und/oder Beschäftigten. Der Dateiumfang einzelner Jahresscheiben beträgt zwischen 140 und 205 MB<sup>12</sup>. Inhalt und Aufbau der Daten in den Personendatenbanken werden im FDZ-Datenreport Nr. 2 beschrieben.

---

<sup>11</sup> Für die Bildung von Subsamples kann sowohl auf der Betriebs- wie Personenebene angesetzt werden. Auf der Betriebsseite kann die Beschränkung auf eine bestimmte Anzahl erfolgen, bei Personendaten kann der Zeitraum 1991 bis 2001 verkürzt werden.

<sup>12</sup> Die Größenangabe bezieht sich wie alle nachfolgenden auf das Softwareanalysetool STATA 8.2. SE.

---

**LIAB-Längsschnittmodell, Version 1:** Für je etwa 2100 ost- und westdeutsche Betriebe liegen alle Beschäftigungs- und Leistungsmeldungen von 1990 bis 2001 vor. Die Bedingung für die Aufnahme des Personenkontos ist eine mindestens eintägige Beschäftigung zwischen 1996 und 2001 in einem Betrieb, der von 1999 bis 2001 in jedem Jahr an der Befragung des IAB-Betriebspanels teilgenommen hat. Der Dateiumfang beträgt für den westdeutschen Personendatensatz 1520 MB, für den ostdeutschen 1150 MB. Inhalt und Aufbau der Daten werden im FDZ-Datenreport Nr. 3 beschrieben.

**LIAB-Längsschnittmodell, Version 2:** Die Bedingung für die Aufnahme des Personenkontos ist eine mindestens eintägige Beschäftigung zwischen 1997 und 2002 in einem Betrieb, der entweder zur Version 1 des LIAB-Längsschnittmodells gehört und/oder im Zeitraum 2000 bis 2002 kontinuierlich an der Befragung des IAB-Betriebspanels teilgenommen hat. Die Personenkonten liegen für einen Zeitraum von 1993 bis 2002 vollständig vor. Insgesamt handelt es sich um die Beschäftigten in 5293 westdeutschen und 4360 ostdeutschen Betrieben. Die Personendatenbanken liegen für jeden Landesteil getrennt vor. Dank eines neuen Datenorganisationssystems und aufgrund der effizienten Speicherung der Variablen lassen sich diese Personendatenbanken an der Kapazitätsgrenze der IAB-Rechner verarbeiten. Die Personendatenbanken der Version 2 verlangen von den Anwendern eine gewisse Disziplin bei der Datenaufbereitung, denn es können nur wenige Variablen zu diesen Rohdatensätzen hinzugefügt werden. Es handelt sich um mehr als 28 Mio. westdeutsche und etwa 16 Mio. ostdeutsche Meldungen.

Im Verlauf der ersten Projektphase des Forschungsdatenzentrums werden *voraussichtlich* außerdem folgende Versionen bereitgestellt:

**LIAB-Querschnittmodell, Version 2:** Sie ist wie die Version 1 aufgebaut. Eine einzelne Jahresscheibe enthält alle Beschäftigtenmeldungen des jeweiligen Befragungsjahres und die Beschäftigten- oder Leistungsmeldung zum 30. Juni des Vor- und Nachjahres der entsprechenden Personen.

**LIAB-Querschnittmodell, Version 3:** Aufgrund der Diskussionen und Anregungen auf dem LIAB-Workshop im November 2004 in Nürnberg wird eine Querschnittsversion erstellt, die Erwerbsverlaufsinformationen der am 30. Juni beschäftigten Personen für

---



diesen Zeitpunkt generiert. Genaueres zu Art und Inhalt dieser synthetischen Zusatzvariablen wird rechtzeitig im Internet bekannt gegeben.

**LIAB-Längsschnittmodell 2, Version 3:** Die Datenbank enthält alle Personendaten für die Längsschnittbetriebe des IAB-Betriebspanels von 1993 bis zum Jahr 2001, also nur westdeutsche Betriebe. Bedingung für die Aufnahme des Personenkontos von 1990 bis 2001 ist eine mindestens eintägige Beschäftigung in einem dieser Längsschnittbetriebe im letztgenannten Zeitraum.

Die Datenreporte der ersten drei Versionen geben nach einer allgemeinen Einführung Auskunft über den Abdeckungsgrad der jeweiligen Version über Betriebe und Beschäftigte und das Personendatenmaterial (die verfügbaren Variablen) werden erläutert. Ebenso befinden sich in den Reporten Auflistungen und Erläuterungen der Variablen in den Personendatenbanken.

Der FDZ-Methodenreport Nr. 1 beschreibt die Datenverknüpfungsqualität der Betriebs- und Personendaten. Die Lektüre dieses Reports wird bei der Absicht einer konkreten Untersuchung empfohlen. Einige Versionsbeschreibungen, insbesondere die Konstruktion der Personendatenbanken, nehmen explizit Bezug auf diesen Methodenreport Nr. 1.

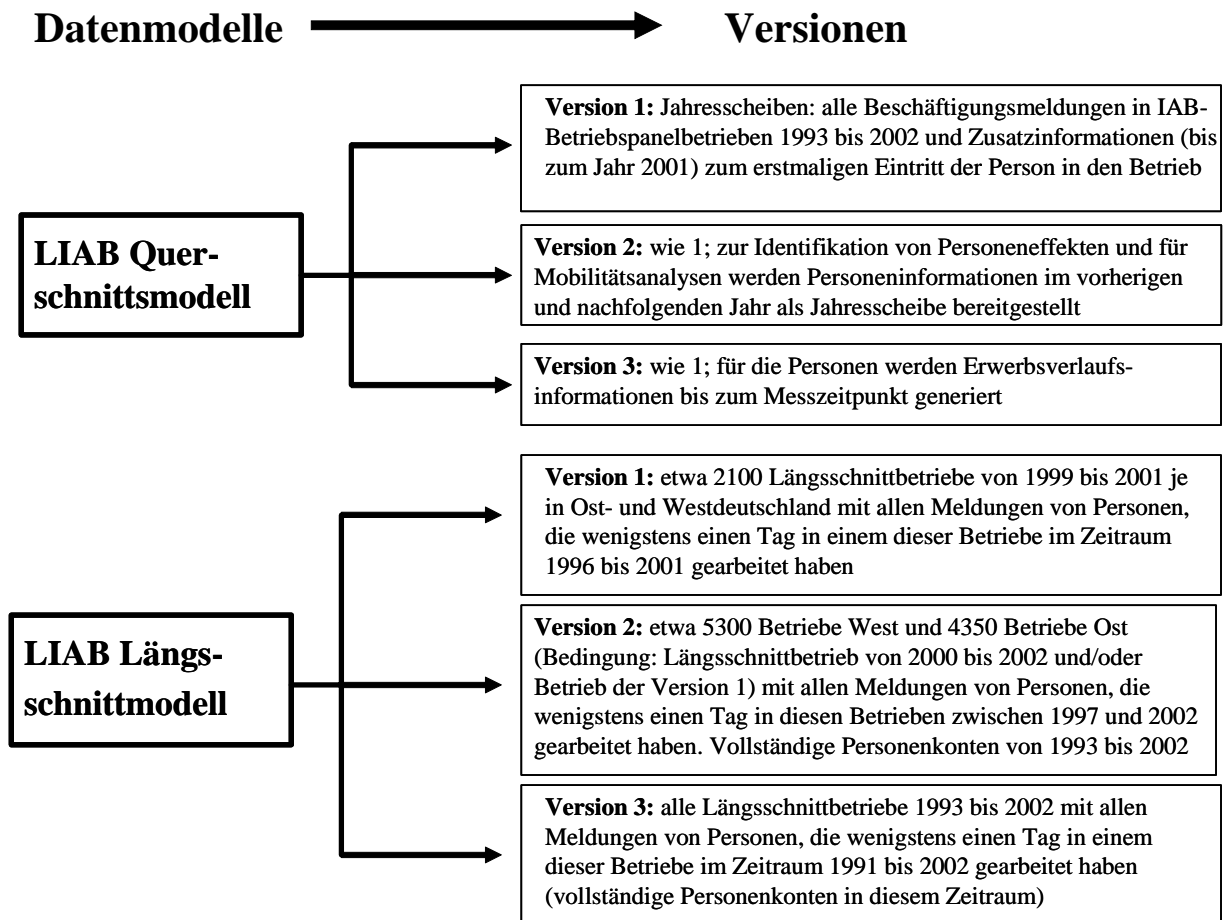
## **6. Zusammenfassung**

Der Report beschreibt Eigenschaften der LIAB-Daten und stellt die externen Wissenschaftlern bereits zugänglichen und die geplanten Versionen im FDZ der BA im IAB vor. Abbildung 1 gibt einen Überblick über die LIAB-Datenmodelle und Versionen.

---



**Abbildung 1: Genese der Linked-Employer-Employee-Daten im FDZ der BA im IAB**



Die Eigenheiten der Arbeit mit diesen Versionen sind in erster Linie in den verfügbaren Informationen über die in der Datenbank enthaltenen Personen, speziell ihrer maximal beobachtbaren erwerbsbiographischen Ereignisse, sowie in den Datenaufbereitungs- und Rechenzeiten zu sehen. Der vorliegende Beitrag ist die Basis für nachfolgende Datenreporte, in denen u.a. die einzelnen LIAB-Versionen, also die zu den entsprechenden IAB-Betriebspanelbetrieben zugeordneten Personendatenbanken, genau beschrieben werden. Die nachfolgenden Datenreporte konzentrieren sich auf die speziellen Aspekte der konkreten Version, ohne den Prozess der Datengenese oder technische Restriktionen vom Grundsatz her erneut zu beschreiben. Der vorliegende FDZ-Datenreport Nr. 1 ist auch der Versuch, relevante Vorgänge bei der Bereitstellung der LIAB-Daten im FDZ der BA im IAB transparent zu gestalten. Detaillierte Variablenbeschreibungen und Erläuterungen zum Datenzugang finden Sie auf der Web-Page des FDZ (<http://fdz.iab.de>). Dort sind auch die einzelnen Datenreporte erhältlich.

## Impressum

### FDZ *Datenreport*

Nr. 1/2005

#### Herausgeber

Forschungsdatenzentrum (FDZ)  
der Bundesagentur für Arbeit  
Regensburger Str. 104  
D-90478 Nürnberg

#### Redaktion

Dagmar Herrlinger, Annette Kohlmann

#### Technische Herstellung

Dagmar Herrlinger

#### Rechte

Nachdruck – auch auszugsweise – nur mit  
Genehmigung des FDZ gestattet

#### Bezugsmöglichkeit

Volltext-Download dieses Datenreports unter  
[http://doku.iab.de/fdz/reporte/2005/DR\\_1.pdf](http://doku.iab.de/fdz/reporte/2005/DR_1.pdf)

#### FDZ im Internet

<http://fdz.iab.de/>

#### Rückfragen zum Inhalt an

Holger Alda, Forschungsdatenzentrum (FDZ) der  
Bundesagentur für Arbeit (BA) im Institut für Arbeits-  
markt- und Berufsforschung (IAB), Regensburger  
Strasse 104, 90478 Nürnberg,  
E-Mail: [holger.alda@iab.de](mailto:holger.alda@iab.de)