

Betriebe und Beschäftigte in den Linked-Employer-Employee-Daten des IAB, LIAB

**zur Analyse im Forschungsdatenzentrum der Bundesagentur für Arbeit
im Institut für Arbeitsmarkt- und Berufsforschung, FDZ**

**Präsentation auf dem LIAB-Workshop am 26. November 2004
im Verwaltungszentrum der Bundesagentur für Arbeit**

**Holger Alda, Forschungsdatenzentrum
E-Mail: Holger.Alda@iab.de**

Die Themen:

- **Anforderungen an Linked-Employer-Employee-Daten**
- **Betriebe und Personen in den IAB-Daten**
- **LIAB-Datenmodelle**
- **Konsistenz der Betriebs- und Personendaten**
- **LIAB-Versionen**
- **erste Version des LIAB-Querschnittmodells**
- **erste Version des LIAB-Längsschnittmodells**
- **Testdaten**
- **Zusammenfassung und Ausblick**

Anforderungen auf der Datenebene

Linked-Employer-Employee-Daten sollten die grundlegende Eigenschaften haben, dass die Beobachtungsebene „Betrieb“ möglichst vollständig durch Personendaten wiedergegeben wird. Dazu gehört beispielsweise :

- **Identifikation der Betriebsbelegschaft über einen Identifikator (Betriebsnummer)**
- **Identifikation des Ein- und Austrittes von Personen in und aus Betrieben**
- **Informationen zur betrieblichen Personalstruktur (Beschäftigtenmix)**
- **Informationen über betriebliche Merkmale**
- **Informationen über Personenmerkmale**
- **Daten auf der Betriebs- und Personenebene sollten sich auf die gleiche Einheit beziehen**
- **für bestimmte Fragestellungen: Erwerbsverläufe von Personen sollten möglichst auch analysierbar sein, wenn sie in mehr als einem Betrieb arbeiten**

Betriebe und Personen in den LIAB-Daten

Betriebe: IAB-Betriebspanel

Personen: Beschäftigten-Leistungsempfänger-Historik-
Datei des IAB, BLH

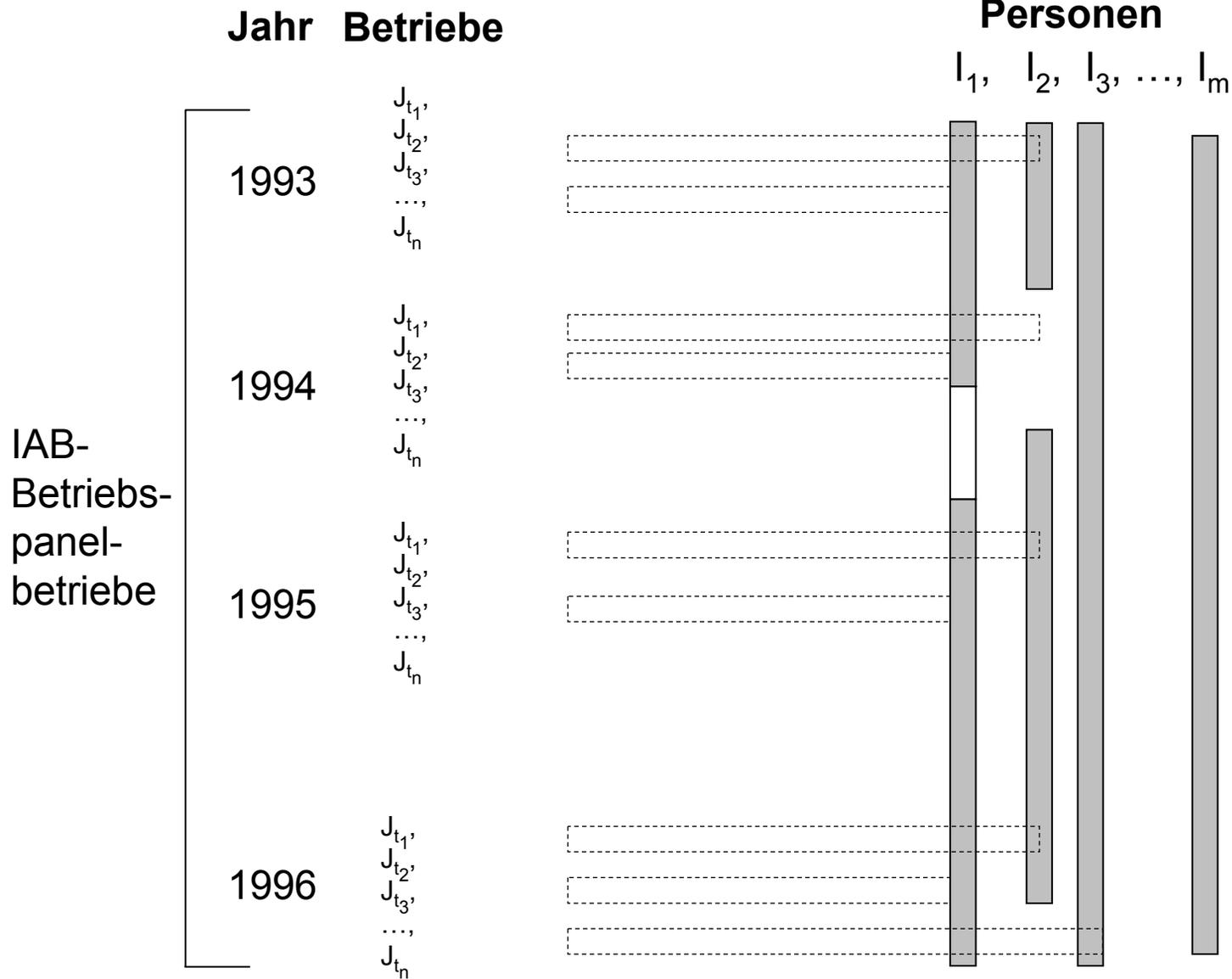
- **Das IAB-Betriebspanel erfüllt die meisten der allgemeinen Anforderungen an Linked-Employer-Employee-Daten**
- **Die BLH stellt die zur Zeit bestmögliche Ergänzung der Betriebsdaten auf der Personenebene dar**

Die LIAB-Datenmodelle

Betriebe (des IAB-Betriebspanels) sind Ausgangspunkt

- nach ihnen richtet sich immer die Ziehung von Personendaten
- der disproportionale Stichprobenansatz des IAB-Betriebspanels mit mehr Großbetrieben lässt eher große Personendatenbanken erwarten

➔ Lösung: LIAB-Datenmodelle



Es gibt zwei verschiedene Möglichkeiten, Personendaten abzulegen := LIAB-Datenmodelle

1.) LIAB-Querschnittsmodell

- enthält die Beschäftigtenmeldungen zum 30. Juni eines jeden Befragungsjahres
- spart Speicherplatz und Rechenzeiten ein
- Querschnitte können zu einem Panel verbunden werden
- wenn Personen den Betrieb verlassen, sind keine weiteren Informationen für diese Personen verfügbar
- die Beobachtungsebene Betrieb wird für die Zeitdauer der Teilnahme am IAB-Betriebspanel konstant gehalten

2.) LIAB-Längsschnittmodell

- enthält alle Beschäftigungs- und Leistungsmeldungen von Personen in einem bestimmten Zeitraum
- lange Datenaufbereitungs- und Rechenzeiten
- Querschnitte von Betrieben können zu einem Panel verbunden werden
- wenn Personen den Betrieb verlassen, sind weitere Erwerbsverlaufinformationen vorhanden
- die Beobachtungsebenen Betrieb *und* Personen werden für die Teilnahme am Panel konstant gehalten
- darüber hinaus ist die Beobachtungsebene Betrieb auf der Personenebene eventuell länger verfügbar
- die Entwicklung der im Betrieb Verbleibenden kann über einen längeren Zeitraum verfolgt werden (z.B. zur Bestimmung der Betriebszugehörigkeitsdauer)

Ausfälle und Inkonsistenzen der Betriebs- und Personendaten entstehen, wenn

a) die Anzahl der sozialversicherungspflichtig Beschäftigten nach den Angaben der Betriebe oder laut Angabe aus der Betriebsdatei des IAB (Ziehungsgrundlage des IAB-Betriebspanels) nicht mit der Anzahl der Personenmeldungen übereinstimmt oder

b) keine Personendaten zu einem befragten Betrieb gefunden werden

a) : Angaben können durch kurzfristige betriebliche Personalfluktuatation oder Erinnerungslücken ungenau sein → bei einem Vergleich sollten Toleranzen zugelassen werden, hier :

1 bis 5 Beschäftigte : 40 Prozent Abweichung

5 bis 19 Beschäftigte: 30 Prozent Abweichung

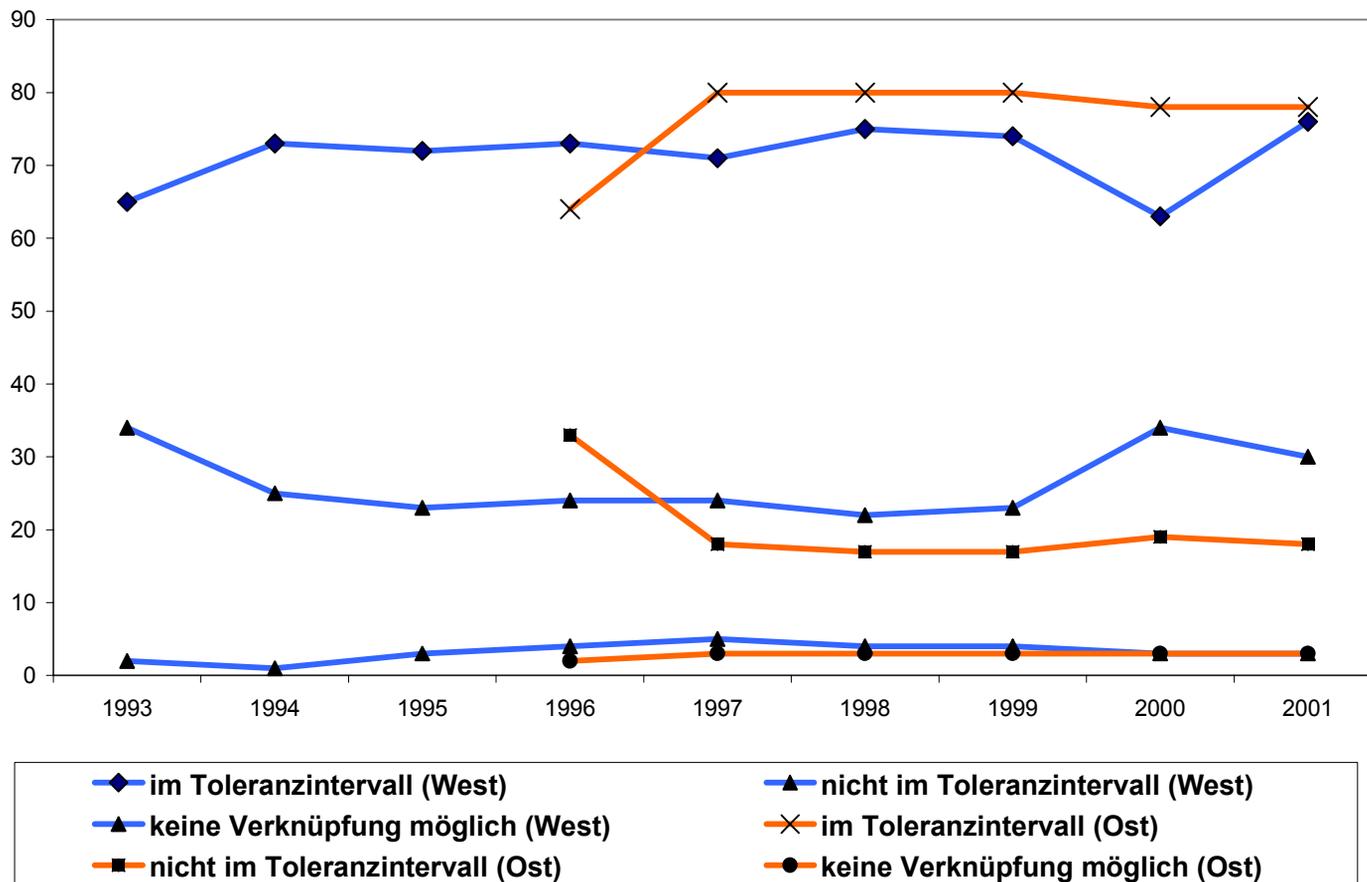
20 bis 100 Beschäftigte : 20 Prozent Abweichung

100 und mehr Beschäftigte: 10 Prozent Abweichung

b) kann direkt untersucht werden

Ausfälle und Inkonsistenzen der Betriebs- und Personendaten

Ergebnis für die derzeit zur Verfügung stehenden LIAB-Daten (Angabe in Prozent der Betriebe; Stichtag 30. Juni 1993 bis 2001;)



Die erste Version des LIAB-Querschnittmodells

Kenngröße/ Jahr	1993	1994	1995	1996	1997
Beobachtungen	2652776	2235644	1916065	2416359	2126793
davon Meldungen					
geringfügige	0	0	0	0	0
Teilzeit < 18 h	22859	20080	18751	18725	17139
Teilzeit > 18 h	217274	189929	169861	238675	220740
Frauen	863894	736168	641492	970092	836764
Westkonten	2652776	2235644	1916065	1640493	1472966
weitere Informationen					
zensierte Entgelte	253795	207101	182826	152600	122334
<i>Personen</i> mit mehr als einem Job	69007	12628	32351	6630	6504
Anzahl Betriebsnummern	4189	4038	3933	8056	8374
Datensatzgröße in MB*	194	163	140	176	155
Anzahl Variablen	27	27	27	27	27
Kenngröße/ Jahr	1998	1999	2000	2001	2002
Beobachtungen	2108118	2026658	2488589	2532641	
davon Meldungen					Die Daten
geringfügige	0	56046	78229	93396	
Teilzeit < 18 h	21337	77913	111106	134887	für das
Teilzeit > 18 h	241481	243986	275241	315197	
Frauen	830405	822021	887386	1007955	Jahr 2002
Westkonten	1441127	1384491	1702241	1946816	
weitere Informationen					liegen
zensierte Entgelte	120250	111572	139428	141491	
<i>Personen</i> mit mehr als einem Job	27272	20082	21532	24618	noch nicht
Anzahl Betriebsnummern	8752	9353	13361	14878	
Datensatzgröße in MB*	153	164	182	205	vor
Anzahl Variablen	27	27	27	27	

* in STATA 8.2. SE

Exkurs: Gruppierung der verfügbaren Merkmale in der LIAB-Personendatenbank der ersten Version des LIAB-Querschnittmodells unverzichtbar: Betriebs- und Personenidentifikatoren

„Anwendermerkmale“	oft entbehrliche Merkmale	„sensible“ Merkmale
Spellzähler quellenbez.	Spellzähler personenbezogen	AA-Bezirk Arbeitsort
Mann/Frau	Beginn,-Endedatum	AA-Bezirk Wohnort
Berufskennziffer	Abgabegrund d. Meldung	Kreiskennz. Arbeitsort
Tagesentgelt	Dauer d. Meldung in Tagen	Kreiskennz. Wohnort
Personengruppe	Kennz. Datumsbereinigung	WZ 1973
Stellung im Beruf	Währungskennzeichen	WZ 1993
Ausbildung der Person	Ost/West-Kennzeichen	
Alter der Person		
Staatsangehörigkeit		
Eintrittsdatum Betrieb		
Eintrittsdatum Erwerbsleben		

Die erste Version des LIAB-Längsschnittmodells

Kennzahl	Datensatz	
	West	Ost
Anzahl Betriebsnummern der Version 1 (Grundlage)	2131	2116
Anzahl Personen im Datensatz	1.078.735	729.440
Meldungen insgesamt	14.473.545	10.955.976
Meldungen von Leistungsbezug	1.375.873	2.964.702
Meldungen von Personen, die in beiden Datensätzen enthalten sind	42.105	42.105
Meldungen aus dem jeweils anderen Landesteil	346.207	950.676
Anzahl Personen, die mindestens einen Tag zeitgleich mehr als ein Beschäftigungsverhältnis haben	236.737	145.604
Gesamtanzahl der Betriebsnummern	562.548	370.990
Anzahl weiterer IAB-Betriebspanelnummern**	29.025	19.925
Datensatzgröße in MB*	1.520	1.150
Anzahl Variablen	31	31

* Größe der entsprechenden STATA© 8.2. SE-Datensätze;

** zum Teil in den Jahren 2002 und 2003 erstmals befragt

Arbeitshilfe Testdaten

Die Testdaten dienen der Erstellung fehlerfrei laufender Auswertungsprogramme bereits vor dem Gastaufenthalt und der Analyse mit den Originaldaten.

1) Betriebsseite:

- etwa 100 (Querschnittmodell) oder 55 (Längsschnittmodell) Betriebe. Die Datenstruktur entspricht vollständig den *neuen* IAB-Betriebspaneltestdaten (Panelmortalität wird berücksichtigt, die Daten enthalten einen eindeutigen Betriebsidentifikator)
- Größenstruktur: 1 Betrieb mit 1000 Beschäftigten
2 Betriebe mit 500 Beschäftigten
2 Betriebe mit 250 Beschäftigten usw. (vgl. Merkblatt)
- Testdaten bilden über die Variable, die die Anzahl der sozialversicherungspflichtig Beschäftigten beinhaltet, ansatzweise die angesprochenen Konsistenzprobleme ab

Arbeitshilfe Testdaten

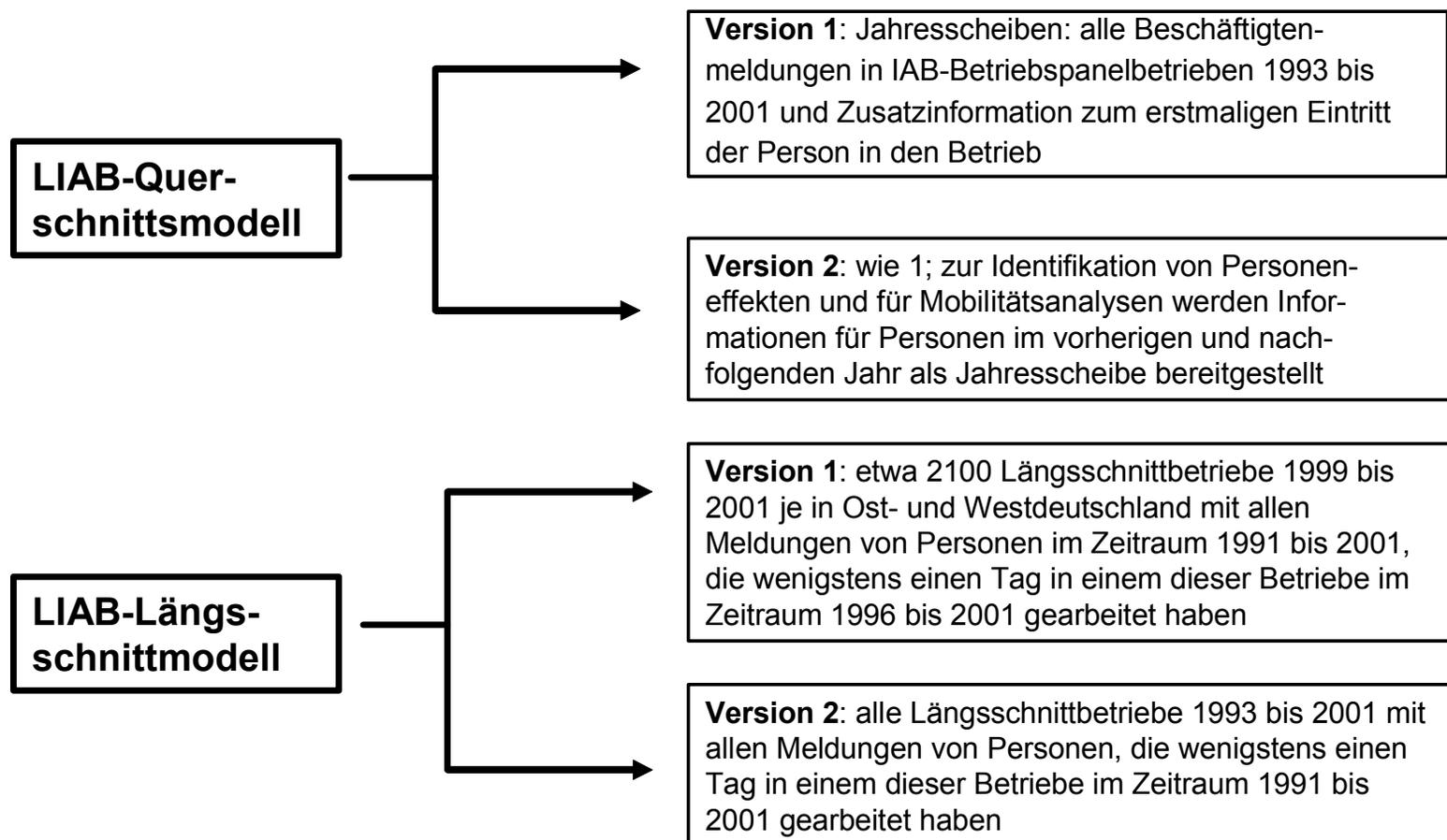
2) Personendaten:

- für die erste Version des LIAB-Querschnittmodells stehen jährlich etwa 4800 Personen (Beobachtungen) zur Verfügung
- jedes Jahr scheiden etwa zehn Prozent der Beschäftigten aus den Längsschnittbetrieben aus und etwa zehn Prozent kommen neu hinzu
- für die erste Version des LIAB-Längsschnittmodells stehen etwa 71 Tsd. Beobachtungen (das sind 6462 Personen) zur Verfügung
- Die Verlaufsdaten werden durch konsistente Beginn- und Enddatums-Meldungen abgebildet; sonstige Merkmalsinhalte sind – mit Ausnahme des Betriebsidentifikators – mehr oder minder zufällig gefüllt

In den ersten Versionen beider Modelle sind die Personentestdaten vollständig belabelt und geben den Wertebereich der Originaldaten inklusive der Vercodung wider

Zusammenfassung und Ausblick I

Datenmodelle → Versionen



Zusammenfassung und Ausblick II

- **eine der wenigen Möglichkeiten, in Deutschland Linked-Employer-Employee-Daten zu erzeugen, besteht in der Verbindung des IAB-Betriebspanels und der IAB-Personendaten (BLH)**
- **die Datenstruktur ist sehr komplex und es gibt technische Restriktionen**
- **die Lösung besteht in LIAB-Datenmodellen, die für unterschiedliche Untersuchungsgegenstände das passende Personendatenbankmodell vorselektieren**
- **konkrete Personendatenbanken werden über Versionen umgesetzt, die Auszüge aus dem theoretisch maximal möglichen Betriebs- und Beschäftigtenuniversum extrahieren und der Wissenschaft zur Verfügung stellen**
- **Datenmodelle und Versionen sind offen für zukünftige wissenschaftliche Entwicklungen und Bedürfnisse**