

Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB)

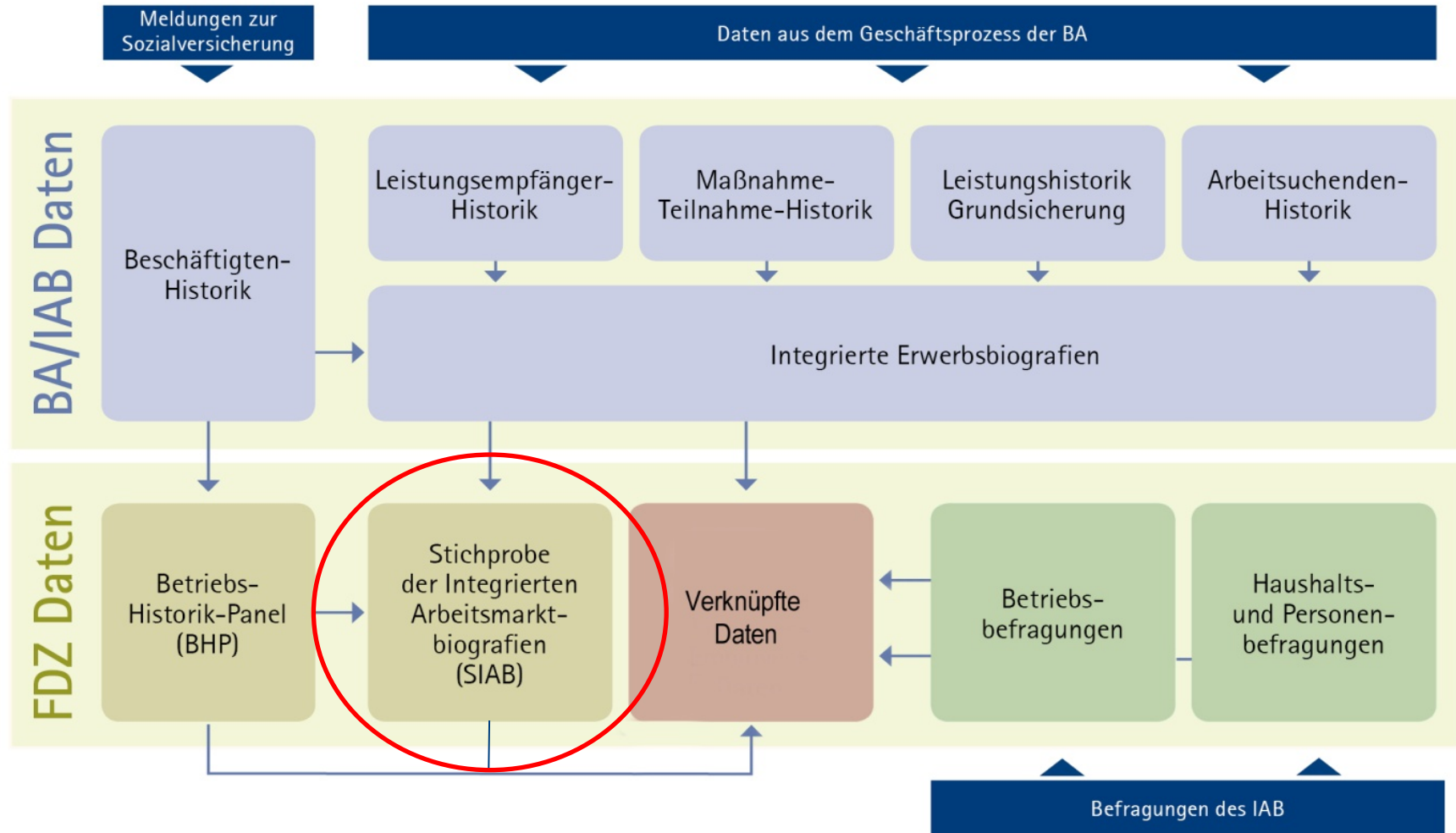
IAB Workshop

Daten des Nationalen Bildungspanels,
der amtlichen Statistik und des IAB
10.-11.06.2013

Simon Trenkle

Übersicht

- Kurzeinführung
- Datenherkunft
- Datenquellen
- Datensatzerstellung
- Variablenübersicht
- Datenqualität
- Analysepotenzial
- SIAB-Versionen
- Weiterführende Informationen



SIAB

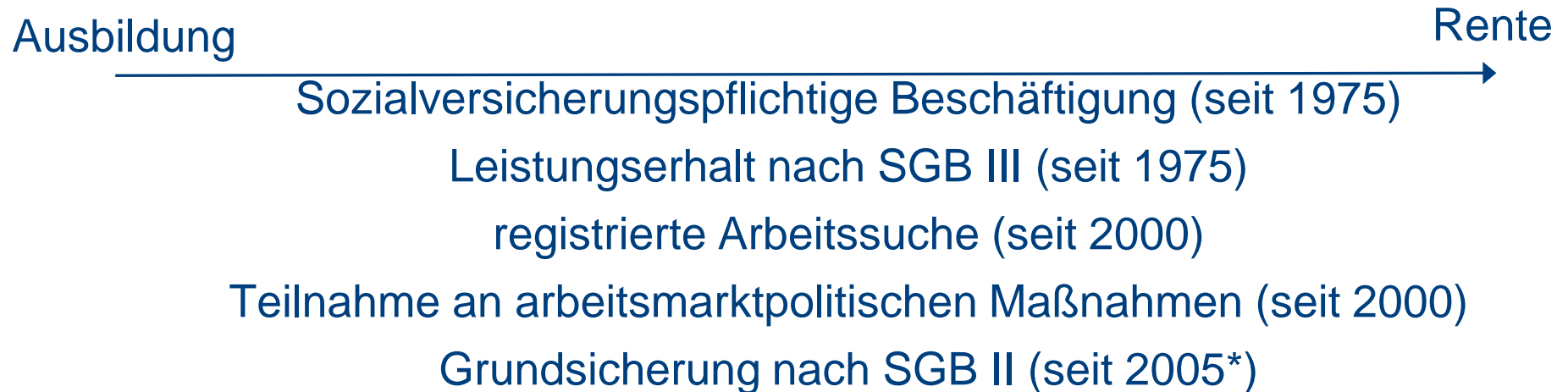
- enthält aktuell Arbeitsmarktbiografien von 1975 - 2010 (SIAB-7510)
- 2% Personenstichprobe sozialversicherungspflichtiger Personen und Personen aus den Geschäftsprozessen der BA
 - ➔ mehr als 1,6 Mio. Personen und 45 Mio. Einzelmeldungen
- Mikro-Datensatz
- biografischer Datensatz („Spell“-Format)
- tagesgenaue Informationen

Administrative Daten



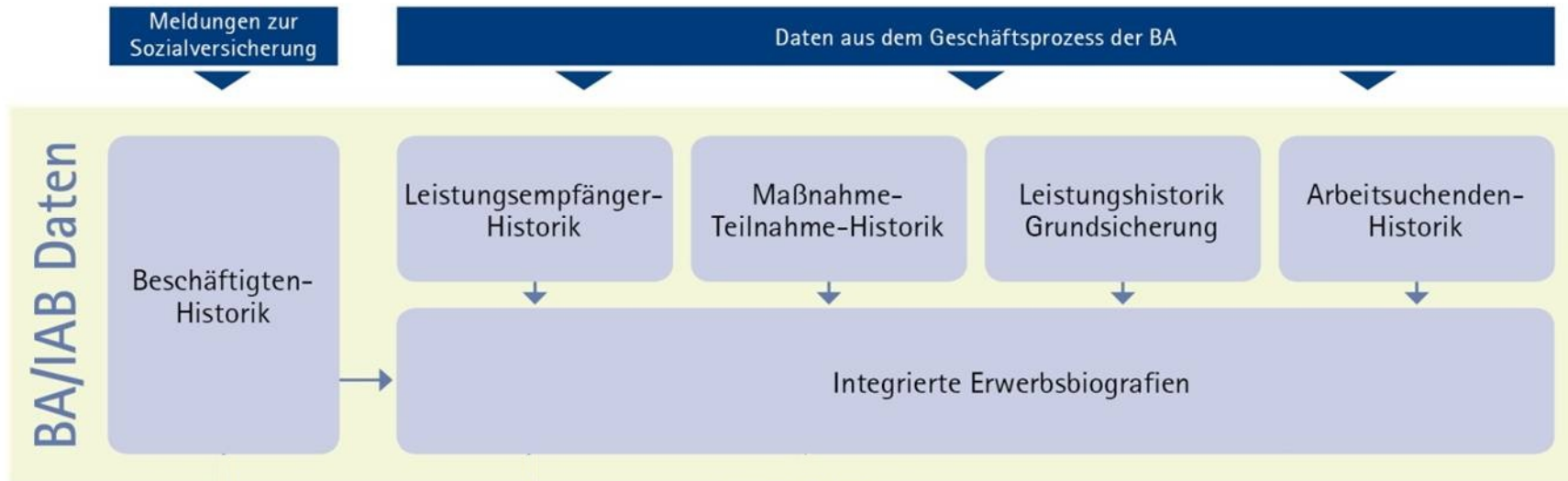
- Zahlung /Bezug von Sozialleistungen an Arbeitslose
 - Teilnahme an arbeitsmarktpolitischen Maßnahmen
 - Registrierte Arbeitssuche
 - Start- und Enddatum
- ➔ computergestützte Prozesse

- BA leitet die Daten an das IAB weiter
- IAB verknüpft Meldedaten mit den Daten aus den Geschäftsprozessen der BA
→ (vollständige) Erwerbsbiografien



- keine Informationen zu Beamten und Selbstständigen

* in den Datensätzen des FDZ erst seit 2007



- Beschäftigten- Historik (BeH): Sozialversicherungsdaten
- Leistungsempfänger- Historik (LeH): Leistungsempfänger nach SGB III
- Leistungshistorik Grundsicherung ((X)LHG): Daten über ALG-II Empfänger
- Arbeitsuchenden- Historik ((X)ASU): Daten über arbeitsuchend gemeldete Personen
- Maßnahme-Teilnahme-Historik (MTH): Daten über Teilnehmer an BA-Programmen (NICHT in der SIAB enthalten)

Beschäftigten- Historik

- enthält alle Sozialversicherungsmeldungen (entspricht ca. 80% aller Beschäftigten)
- Nicht enthalten sind: Beamte, Selbständige, Freiberufler, vor 1999: geringfügig Beschäftigte
- für Westdeutschland seit 1975, für Ostdeutschland seit 1992 gefüllt
- Quantitativ wichtigste SIAB-Quelle (ca. 65% aller Meldungen)
- Angaben zu:
 - Bruttotagesentgelt (bis Beitragsbemessungsgrenze)
 - Geburtsdatum
 - Geschlecht
 - Schulabschluss
 - Beschäftigungsdauer (tagesgenau)
 - Betriebsangaben (siehe BHP)
 - ...

Leistungsempfänger- Historik

- enthält BA-Daten über Leistungsempfänger nach SGB III (ALG, ALHi, UHG, seit 2005 ALG-I) Empfänger
- für Westdeutschland seit 1975, für Ostdeutschland seit 1992 gefüllt
- macht ca. 12% aller SIAB-Meldungen aus
- Angaben zu:
 - Zeitraum des Leistungsbezugs
 - Familienstand (verheiratet/nicht verheiratet)
 - Kinder (ja/nein)
 - Leistungssatz
 - Grund für Beendigung des Leistungsbezugs
 - ...

Leistungshistorik Grundsicherung

- enthält BA-Daten (LHG) und Daten kommunaler Träger (XLHG) über Arbeitslosengeld-II Empfänger (SGB II)
- Daten seit 2005 enthalten
- macht ca. 5% aller SIAB-Meldungen aus
- keine (reinen) Personen- sondern Haushaltsinformationen
- geringe Füllgrade in den Jahren 2005 bis 2007
- Angaben zu:
 - Zeitraum des ALG-Bezugs
 - Kinder (Anzahl)
 - Grund für Beendigung des Leistungsbezugs
 - ...

Arbeitsuchenden- Historik

- enthält BA-Daten (ASU) und Daten kommunaler Träger (XASU) über arbeitsuchende Personen
- Daten seit 2000 enthalten
- macht ca. 18% aller SIAB-Meldungen aus
- deutliche Qualitätsunterschiede zwischen Trägern
- Angaben zu:
 - (Schul)bildung
 - Schwerbehindertenstatus
 - Gesundheitliche Einschränkungen
 - ...

Quellen und Zeiträume

	BeH/LeH	Daten für Ostdeutschland	(X)ASU+ Geringfügig beschäftigte	(X)LHG (ALG-II)
1975-1991	X			
1992-1999	X	X		
2000-2004	X	X	X	
2005-2010	X	X	X	X

- Daten sind nicht immer für ganz Deutschland und für alle Quellen verfügbar
- vier „Blöcke“ unterscheidbar
- seit 2005 Daten für alle Quellen und ganz Deutschland verfügbar

Aufbereitung der Datenquellen

- Für alle Quellen werden folgende Bereinigungen durchgeführt:
 - Löschung von Sätzen, bei denen das Alter kleiner 13 o. größer 75
 - Löschung von Sätzen, bei denen Enddatum kleiner Beginndatum
 - Widersprüchliche Angaben zu Geschlecht oder Geburtsdatum zu einer Person werden bereinigt (Prioritätsregeln)
 - Falls kein Geburtsdatum vorhanden, erfolgt Löschung des Satzes
 - Ersetzung des Geburtsdatums durch das Geburtsjahr
- Zusätzlich werden verschiedene quellenspezifische Bereinigungen durchgeführt (z.B. werden nicht alle Informationen übernommen)

Zusammenführung und Stichprobenziehung

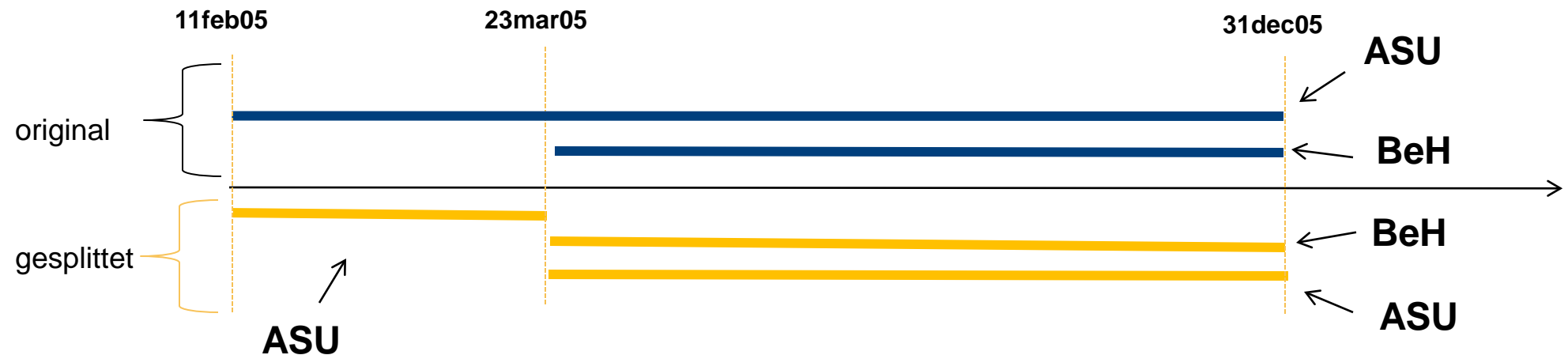
- Zusammenführung (IEB):
 - Die verschiedenen Datenquellen werden zusammengeführt (Personenidentifikator)
 - ➔ Es entstehen die Integrierten Erwerbsbiografien (IEB)
- Stichprobenziehung (SIAB):
 - Vektor aus allen Personenidentifikatoren wird erstellt
 - zufällige Ziehung der Personenidentifikatoren
 - für jede Nummer selbe Wahrscheinlichkeit (2%) gezogen zu werden
 - SIAB sowohl im Quer- als auch im Längsschnitt für alle verwendete Quellen repräsentativ

Strukturierung der SIAB

- zunächst Sortierung nach Personenummer und Beginndatum der einzelnen Dateninformation
- Aber: zeitgleiche sowie sich überschneidende Ereignisse entstehen zwischen den Datenquellen
- Bei gleichlangen Sätzen wird Priorisierungsregel angewandt
- BeH>LeH>ASU>LHG>XASU>XLHG
- Überschneidende Sätze bei einigen Analysen (z.B. Ereignisanalyse) unerwünscht
- Splittung der Daten, so dass sich komplett überdeckende, sowie komplett überdeckungsfreie Einträge entstehen
- Anzahl der Datenzeilen erhöht sich

Spell-Format und Spell-Splitting

persnr	begorig	endorig	spegepi	tengepi	tegepi	questell	tentgelt	quelle
883	11feb05	31dec05	11feb05	23mar05	23mar05	7 (ASU)	.n	7 (ASU)
883	24mar05	31dec05	24mar05	31dec05	31dec05	1 (BeH)	71.4	1 (BeH)
883	11feb05	31dec05	24mar05	31dec05	31dec05	3	.n	7 (ASU)



Enthaltene Variablengruppen

- Identifikatoren
 - Personennummer
 - Betriebsnummer
 - ➔ Betriebs-Historik- Panel (BHP) kann zugespield werden!
- Gültigkeitszeiträume
 - Beginn- und Endpunkte von Originalsätzen
 - Beginn- und Endpunkte von gesplitteten Episoden
- Personenangaben
 - Geschlecht
 - Geburtsjahr
 - Staatsangehörigkeit
 - Familienstand
 - ...

- Informationen zu Beschäftigung, Leistungsbezug, Arbeitssuche
 - Abgabe-/Beendigungs-/Einstellungsgrund
 - Tagesentgelt
 - Beruf
 - Stellung im Beruf
 - ...
- Betriebsmerkmale
 - Wirtschaftszweigangaben
 - Anzahl Beschäftigter gesamt, Vollzeit, Teilzeit
 - ...
- Ortsangaben
 - Wohnort Kreis, Arbeitsort Kreis
 - ...

Wie steht's mit der Qualität der Daten?

- Generell:
 - besonders bei Sozialversicherungsdaten hohe Zuverlässigkeit
 - Ziehung der Stichprobe aus Grundgesamtheit (keine Verzerrung)
 - langer Beobachtungszeitraum, großes N, tagesgenaue Informationen
- Aber auch:
 - teilweise schlechte Qualität bei XASU und XLHG Informationen
 - teilweise Inkonsistenzen z.B. bei Berufsangaben, Bildungsvariablen

→ Hilfe bei Bildungsvariable:

Fitzenberger, Bernd; Osikominu, Aderonke; Völter, Robert (2006): Imputation rules to improve the education variable in the IAB employment subsample. In: Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften, Jg. 126, H. 3, S. 405-436.

Eingeschränktes Analysepotenzial und Auswege

- Einschränkung auf sozialversicherungspflichtig Beschäftigte
- Tagesentgelt nur bis zur Beitragsbemessungsgrenze

➔ Hilfe:

Gartner, Hermann (2005): The imputation of wages above the contribution limit with the German IAB employment sample. In: FDZ-Methodenreport, No. 2/2005, 8 S.

Themenauswahl

- **Arbeitslosigkeit**
 - Baltagi, Badi H.; Blien, Uwe (1998): The German Wage Curve: Evidence from the IAB Employment Sample. In: Economics Letters, No. 61/2, S. 135-142.
- **Lohnungleichheit**
 - Möller, Joachim; Hutter, Christian (2011): The effects of age, skill and sector composition on the wage inequality in Germany. In: B. Genser, H. J. Ramser & M. Stadler (Hrsg.), Umverteilung und soziale Gerechtigkeit, (Wirtschaftswissenschaftliches Seminar Ottobeuren, 40), Tübingen: Mohr Siebeck, S. 9-32.
- **Humankapital**
 - Gathmann Christina; Schönberg Uta (2010): How General Is Human Capital? A Task-Based Approach. In: Journal of Labor Economics, No. 28/1, S. 1-49.
 - Osikuminu, Aderonke (2013): Quick Job Entry or Long-Term Human Capital Development? The Dynamic Effects of Alternative Training Schemes. In: Review of Economic Studies, No. 80, S. 313-342.

Die SIAB als Scientific Use File (SUF)

- SIAB nur schwach anonymisiert
- potenzielles Reidentifikationsrisiko, daher nur Zugang über Gastaufenthalt (anschließende Datenfernverarbeitung möglich)
- Für die Verwendung der Daten als SUF stärkere Anonymisierung (faktische Anonymisierung) notwendig
- Folgende anonymisierungsschritte werden durchgeführt:
 - Bestimmte sensible Merkmale grundsätzlich nicht in der SIAB-SUF enthalten (Gesundheitsstatus, Schwerbehindertenstatus)
 - Weitere Merkmale nur vergrößert erhältlich (Nationalität, Quellenangaben)
 - Gebietsangaben vergrößert, so dass mind. 100.000 EW pro Raumeinheit vorliegen

SIAB vs. SIAB-SUF und weitere Formate

	SIAB	SIAB-SUF
Aktuellste Version geht bis	2010	2008
Anzahl Sätze	45.793.010	34.862.777
Anonymisierungsgrad	Schwach anonym	Faktisch Anonym
Sensible Merkmale erhältlich	Auf Antrag	Nein/Vergrößert
Erweiterungen	BHP-Module	Nein

- Außerdem:
 - Campus-Files
 - Testdaten

Interesse geweckt?

- Weiterführende Infos auf unserer Homepage:
→ <http://fdz.iab.de/>
- Arbeitshilfen in deutsch und englisch (auch auf unserer Homepage):
 - Berge, Philipp vom; König, Marion; Seth, Stefan (2013): Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB) 1975-2010. In: FDZ-Datenreport, 01/2013, 65 S.
 - Eberle, Johanna; Schmucker, Alexandra; Seth, Stefan (2013): Programmierbeispiele zur Datenaufbereitung der Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB) in Stata. In: FDZ-Methodenreport 04/2013, 16 S.
- Ansprechpartner im FDZ
 - Philipp vom Berge
 - Simon Trenkle
 - Kontakt über: iab.fdz@iab.de

Vielen Dank für Ihre Aufmerksamkeit

Simon Trenkle

simon.trenkle@iab.de

www.iab.de

<http://fdz.iab.de>

- Block I: 1975-91
 - Daten für Westdeutschland über BeH sowie Leistungsbezug nach SGB III
- Block II: 1992-99
 - Zusätzlich Daten für Ostdeutschland
- Block III: 2000-04
 - Informationen über Arbeitssuchend gemeldete Personen (ASU) sowie geringfügig beschäftigte Personen kommen dazu
- Block IV: 2005-10
 - Daten zu Grundsicherung nach SGB II (ALG II) nach Harzreform enthalten

- Noch zu tun:
 - Meldeverfahren
 - Administrativ vs. Befragungsdaten
- Checken:
 - Datum bei Quellen
 - Ziehverfahren (Welche Nummer?)

Wie ist der Datenzugang möglich?

- Gastaufenthalt
 - Zugang zu schwach anonymisierten SIAB-Daten
 - Ansprechpartner bei Fragen vor Ort (in Nürnberg)
 - anschließende Datenfernverarbeitung möglich
- Scientific Use Files
 - Zugang zu SIAB-Regionalfile (faktisch anonymisierte Daten)
 - Aktuell mit Daten bis 2008 (SIAB_R_7508)
 - Version bis 2010 in Planung
- Außerdem erhältlich:
 - Campus-Files
 - Testdaten