

### **BASiD – Biografiedaten ausgewählter Sozialversicherungsträger in Deutschland 2007** (SUFBASiD07\_Fix und SUFBASiD07\_“*Verlaufsmerkmal*“)

Ergänzung zur Methodischen Umsetzung des SK 79:  
Erweiterung der VSKT um Merkmale aus  
den Daten der Bundesagentur für Arbeit (BA)

**Autoren:**

Janine Lange (FDZ-RV, Berlin)  
Julia Schumilow (FDZ-RV, Würzburg)  
Dr. Michael Stegmann (FDZ-RV, Würzburg)

Stand: 08.03.2012

### Inhalt

<b>0. ÜBERBLICK ÜBER DIE DOKUMENTATION ZUM SUF BASID .....</b>	<b>3</b>
<b>1. ZIELE DER DATENFUSION.....</b>	<b>4</b>
1.1 ALLGEMEINES .....	4
1.2 ERGÄNZUNG DER RV-MELDUNGEN MIT INFORMATIONEN DER BA.....	5
1.3 SCHLIEßEN ZEITLICHER LÜCKEN IN DEN RENTENVERSICHERUNGSBIOGRAFIEN.....	7
<b>2. ZUSÄTZLICHE INFORMATIONEN AUS DEN DATEN DER BA.....</b>	<b>8</b>
2.1 DATENQUELLEN DER BA.....	8
2.2 ZUSATZMERKMALE .....	10
<b>3. ZUSAMMENFÜHRUNG DER BA- UND DER RV-DATEN.....</b>	<b>11</b>
3.1 PRINZIP DER DATENZUSAMMENFÜHRUNG .....	11
3.1.1 <i>Datenzwillinge</i> .....	11
3.1.2 <i>Zusatzinformationen</i> .....	15
3.2 ABLAUF DER ZWILLINGSSUCHE .....	16
3.3 ABLAUF DER ZUSATZSUCHE.....	17
3.4 ABLAUSCHEMA DER DATENGENERIERUNG .....	19
<b>4. ANONYMISIERUNG UND DATENSCHUTZ .....</b>	<b>20</b>

### **ANLAGE**

Systematik des Datenabgleichs zwischen VSKT und Daten des IAB

# Benutzerhinweise Methodische Umsetzung

## BASiD – Biografiedaten ausgewählter Sozial- versicherungsträger in Deutschland 2007

---

### 0. Überblick über die Dokumentation zum SUF BASiD

Die Dokumentation zum SUF BASiD umfasst im Einzelnen folgende Dokumente:

- (1) Methodische Umsetzung des SK 79 in einen anonymisierten Datensatz fester Satzlänge: Sequentielle Biografiedaten
- (2) Ergänzung zur Methodischen Umsetzung des SK 79: Erweiterung der VSKT um Merkmale aus den Daten der Bundesagentur für Arbeit (BA)
- (3) Benutzerhinweise zu den Verlaufsmerkmalen und Merkmalen der Rentenberechnung
- (4) Codeplan

Diese und weitere Arbeitsmaterialien stehen auf der Homepage des Forschungsdatenzentrums der Rentenversicherung zum Download zur Verfügung<sup>1</sup>.

---

<sup>1</sup> Unter '[Datenangebot](#)' -> '[Wissenschaftliche Forschung](#)' -> '[Längsschnittdaten - Themenfiles](#)' sind die jeweiligen Texte zur Dokumentation des BASiD-Datensatzes hinterlegt.

## BASiD – Biografiedaten ausgewählter Sozial- versicherungsträger in Deutschland 2007

---

### 1. Ziele der Datenfusion

#### 1.1 Allgemeines

Der SUF BASiD ist Resultat des vom Bundesministerium für Bildung und Forschung (BMBF) geförderten Verbundprojektes „Biografiedaten ausgewählter Sozialversicherungsträger in Deutschland“ des Forschungsdatenzentrums der Rentenversicherung (FDZ-RV) und des Forschungsdatenzentrums der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (FDZ- BA/IAB).

**Ziel dieses Vorhabens** ist die Erstellung eines Datensatzes, der Längsschnittinformationen aus den Versicherungskonten der Rentenversicherung mit Meldungen aus den statistisch erfassten Biografien der Bundesagentur für Arbeit kombiniert. Dazu wurden ausgewählte prozessproduzierte Daten, die aus Meldungen zur Sozialversicherung und der Verwaltungspraxis gewonnen werden, aus beiden Häusern zusammengespielt. Das so neu entstandene Datenprodukt besitzt damit einen **erweiterten Merkmalskatalog**.

Die Versicherungskontenstichprobe (VSKT) 2007 ist der Basisdatensatz für das Projekt. Das **Stichprobendesign** des SUF BASiD erfolgte entsprechend dem der Versicherungskontenstichprobe (siehe dazu: „Benutzerhinweise - Methodische Umsetzung des SK79 in einen anonymisierten Datensatz fester Länge: Sequentielle Biografiedaten“, Kapitel 1), wobei Zusatzinformationen der BA an die bereits existierenden Daten herangespielt und zeitliche Lücken geschlossen wurden.

Dazu wurden Personen, die sich im Datenbestand zum Stichtag am 31.12.2007 befanden, in den Datensätzen des IAB ermittelt. Die Bestimmung der betreffenden Personen erfolgte anhand der Sozialversicherungsnummer. Diese wurde einmalig beim Rentenversicherungsträger erhoben und zusammen mit einer anonymen Zählnummer dem IAB zur Identifikation der Untersuchungspersonen übermittelt. Das FDZ-RV erhielt die anonyme Zählnummer (zu keinem Zeitpunkt die Sozialversicherungs-Nr.). Die Daten des IAB wurden an das FDZ-RV mittels der anonymen Zählnummer übertragen.

# Benutzerhinweise Methodische Umsetzung

## BASiD – Biografiedaten ausgewählter Sozial- versicherungsträger in Deutschland 2007

---

Die im Projektkontext verwendeten Datensätze wurden aus Gründen des Datenschutzes pseudonymisiert, so dass die Merkmale faktisch anonymisiert vorliegen. So wurden beispielsweise die Sozialversicherungsnummern durch Schlüsselnummern ersetzt.

Für die Datenweitergabe wurde der Datensatz nochmals neu sortiert und nummeriert. Diese neue Fallnummer wird dem Nutzer übermittelt. Nach Umsortierung und Vergabe der neuen Fallnummer wurde eine Stichprobenziehung nach dem Zufallsverfahren durchgeführt. Es wurde eine 25%-Zufalls-Substichprobe für eine ausgewählte Population gezogen. Dabei fand eine Altersbegrenzung bei den Untersuchungspersonen statt, sodass der SUF BASiD letztendlich zum Stichtag vom 31.12.2007 im Inland lebende Deutsche zwischen 30 und 67 Jahren beinhaltet.

Die vorliegenden Datensätze sind im Rechteckdatenformat aufgebaut und bilden die (Versicherungs)biografie als Sequenz ab. Dazu wird die dargestellte (Versicherungs)biografie der jeweiligen Person ab dem 1. Januar des Jahres dargestellt, in dem die Person 14 Jahre alt wird (siehe dazu: Benutzerhinweise - „Methodische Umsetzung des SK79 in einen anonymisierten Datensatz fester Länge: Sequentielle Biografiedaten“, S. 14). Monat 1 (z.B. SES\_1) zeigt also die Information für den Januar des Jahres, in dem die Person 14 Jahre alt wird usw. Sämtliche zeitveränderlichen Datumsangaben werden monatlich ausgegeben, das heißt, es wird für jeden Monat der Lebenszeit, falls vorhanden, ein gültiger Zustand erfasst, wodurch die personenbezogene (rentenrechtlich relevante) Historie nachvollzogen werden kann. Aus dem Datensatz wird also ersichtlich, dass eine Person in einem bestimmten Monat beispielsweise sozialversicherungspflichtig beschäftigt oder arbeitslos war. Mögliche, zeitlich parallel liegende Zustände, wie etwa eine geringfügige Beschäftigung oder Kindererziehungszeiten, werden in hierfür erstellten Zusatzmerkmalen dokumentiert.

### 1.2 Ergänzung der RV-Meldungen mit Informationen der BA

Die Verknüpfung der Datensätze von RV und BA, die so genannte Zwillingsuche, erfolgte im Wesentlichen durch einen Abgleich der einzelnen Zustände in den Biografien<sup>2</sup> sowie der entsprechenden Zeiträume (Beginn und Ende einer Episode) und der Entgeltangaben. Da

---

<sup>2</sup> Auf Seite der Rentenversicherung wurden hierzu die Variablen BYAT/BYATSO, welche Auskunft über Beitragsarten bzw. Beitragsbesonderheiten geben (siehe dazu: „Benutzerhinweise - Methodische Umsetzung des SK79 in einen anonymisierten Datensatz fester Länge: Sequentielle Biografiedaten“, S. 19ff), herangezogen. Auf Seite der Bundesagentur für Arbeit die Variable ERWSTAT\_GR.

## BASiD – Biografiedaten ausgewählter Sozial- versicherungsträger in Deutschland 2007

---

die betreffenden Angaben jeweils aus dem Meldeverfahren zur Sozialversicherung stammen, ist davon auszugehen, dass Phasen der Erwerbstätigkeit und Arbeitslosigkeit in den Biografien sowohl in den IAB- als auch RV-Daten zu finden und identisch (=Zwillinge) sind. Durch die Identifizierung dieser Zwillinge können die ergänzenden Biografieinformationen des IAB an die VSKT herangespielt werden.

Aus Grundsätzen des Datenschutzes und der wirtschaftlichen und sparsamen Datenhaltung werden in den Statistikdaten der beiden Institutionen allerdings nur die für die Erfüllung der jeweiligen Aufgaben relevanten Informationen erfasst und gespeichert. So verfügt beispielsweise nur die Rentenversicherung über Informationen zu Geburtstag und Anzahl der Kinder oder Kindererziehungszeiten, während die Daten der BA dagegen Auskunft über Maßnahmen, Arbeitssuche oder Betriebsinformationen geben.

Die **Herausforderung** besteht nun darin, die sich entsprechenden Informationen in den Datenquellen trotz der teilweise vorhandenen Abweichungen bzw. parallel auftretenden Zustände, zusammenspielen. So kann es beispielsweise vorkommen, dass eine Person in den Daten des IAB für einen Zeitraum arbeitslos gemeldet und in der VSKT kein deckungsgleicher Zeitraum zu finden ist. Diese Schwierigkeiten hängen mit der unterschiedlichen Datenhaltung in beiden Häusern zusammen.

Eine weitere mögliche Ursache für die auftretenden Abweichungen liegt in der **Kontenklärung**, die für die VSKT und die IAB-Forschungsdaten vorgenommen wird. In der VSKT können so beispielsweise unterlassene/vergessene Jahres- oder Abmeldungen der Arbeitgeber aufgedeckt und korrigiert werden. Was die Fehlerprüfung und Datenübermittlung auf Seiten der VSKT anbelangt, so wurden für das Erhebungsjahr 31.12.2007 in geringer Zahl auch fehlerhafte Datensätze übersandt. Soweit möglich, wurde dann in Absprache mit den zuständigen Trägern eine maschinelle Datenbereinigung eingeleitet. Nach dem erneuten Eingang der Daten erfolgte dann nochmals eine maschinelle Plausibilitätsprüfung. Hierbei ist zu beachten, dass generell je nach Sachlage auch durchaus unplausible Datensätze im Datenbestand belassen werden, wenn eine sinnvolle Auswertung trotzdem möglich erscheint. Aus der Erfahrung werden dabei nur sehr wenige Fälle von der Analyse ausgeschlossen.<sup>3</sup>

---

<sup>3</sup> Eine Beschreibung der VSKT-Stichprobe und der Datenaufbereitung findet sich im Vorwort der Publikation: Statistik der Deutschen Rentenversicherung (2007): Rentenanwartschaften am 31.12.2007. Band 181. Deutsche

## BASiD – Biografiedaten ausgewählter Sozial- versicherungsträger in Deutschland 2007

---

Die Konsolidierungsschritte im Rahmen der Erstellung der IAB-Beschäftigtendaten beschränken sich, abgesehen von der Einarbeitung unterlassener/vergessener, geänderter (nur für ehemals DEVO/DÜVO-Meldungen, heute DEÜV) oder stornierter Meldungen, im Wesentlichen auf Gebietsstandsaktualisierungen und RV-Korrekturen. Bei der Gebietsstandsanpassung werden alle regionalen Angaben, also Gemeinde sowie Geschäftsstelle des Arbeits- als auch Wohnorts, für sämtliche enthaltenen Jahre mittels einer vom IAB 1980 entwickelten und seitdem ständig fortgeschriebenen Verweistabelle auf den aktuellen Gebietsstand (31.12. des jeweils aktuellsten Meldejahres) umgeschlüsselt. Bei der RV-Korrektur werden mithilfe einer von der Deutschen Rentenversicherung in unregelmäßigen Abständen bereit gestellten Datei zusammengehörige Versichertendaten für den gesamten abgebildeten Zeitraum, die unter verschiedenen Nummern gemeldet wurden, zusammengeführt. Versicherungsnummern ohne Verweis werden stillgelegt, also das gesamte Konto mit den zugehörigen Daten gelöscht.

### 1.3 Schließen zeitlicher Lücken in den Rentenversicherungsbiografien

Zusätzliche Informationen, die nur in den Daten der BA vorliegen (seien es z.B. Angaben zum Leistungsbezug oder bezüglich der Einbindung in Maßnahmen), können dabei auch bestehende Lücken in der VSKT schließen und somit die Informationsdichte des ursprünglichen Datensatzes erhöhen. Verlaufsmerkmale, die aus den Angaben der BA-Daten stammen, tragen demnach dazu bei, Leerräume in den RV-Daten inhaltlich zu füllen.

So kann z.B. eine bestehende Lücke zwischen einer VSKT-Episode zur Arbeitslosigkeit und einer späteren Meldung zur Erwerbstätigkeit (etwa bei Nichtbezug von ALG II) durch Informationen zu Zeiten der Arbeitssuche und eventuell zeitgleich auftretenden Wiedereingliederungsmaßnahmen in den Arbeitsmarkt komplettiert werden. Des Weiteren lassen sich beispielsweise auch zwei Episoden der Beschäftigung, zwischen denen eine Meldungslücke liegt, durch zusätzliche Angaben hinsichtlich beruflicher Qualifizierungen und Weiterbildungen überbrücken.

---

Rentenversicherung Bund, Berlin (Dieser Band ist auch unter der Adresse [http://forschung.deutsche-rentenversicherung.de/FdzPortalWeb/discontent.do?id=main\\_fdz\\_forschung\\_laengs](http://forschung.deutsche-rentenversicherung.de/FdzPortalWeb/discontent.do?id=main_fdz_forschung_laengs) zu finden).  
Des Weiteren sei auf folgende Publikation verwiesen: Die Versicherungskontenstichprobe als Scientific Use File. Workshop des Forschungsdatenzentrums der Rentenversicherung (FDZ-RV) am 30. und 31. Oktober 2007 in Würzburg, Deutsche Rentenversicherung Bund (Hg.) DRV-Schriften Band 79, Juli 2008.

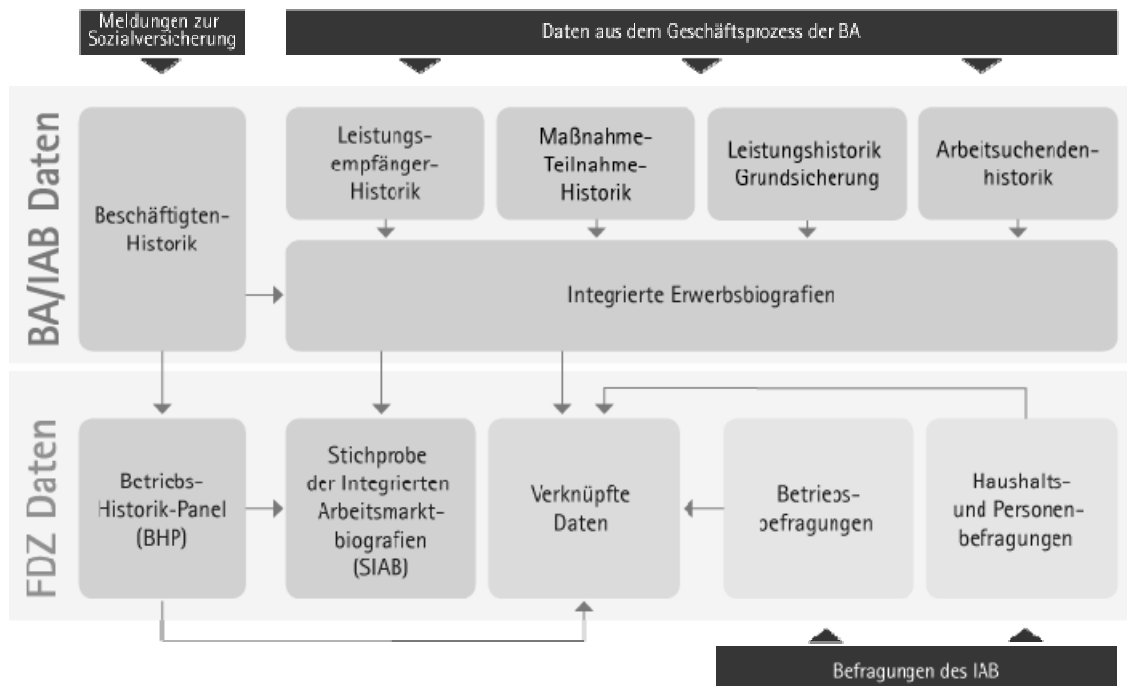
## BASiD – Biografiedaten ausgewählter Sozialversicherungsträger in Deutschland 2007

### 2. Zusätzliche Informationen aus den Daten der BA

#### 2.1 Datenquellen der BA<sup>4</sup>

Das Anliegen, einen Datensatz zu erstellen, der Erwerbsverläufe möglichst vollständig und detailliert abbildet, erfordert die Nutzung einer Vielzahl von Informationen. Um diese auch bereitstellen zu können, wurde bei der Erstellung des BASiD-Datensatzes auf unterschiedliche Datenquellen zurückgegriffen. Als Ausgangsdatsatz dient, wie schon erwähnt, die Versicherungskontenstichprobe der Rentenversicherung. An sie werden die Informationen aus den IAB-Datenquellen herangespielt. Hierbei handelt es sich um die Datenbestände **Integrierte Erwerbsbiografien (IEB)** und um das **Betriebs-Historik-Panel (BHP)**. Eine Übersicht der IAB-Datenbestände gibt die nachstehende Abbildung:

Schaubild 1: Übersicht über die BA-Daten



Quelle: FDZ-IAB Homepage ([http://fdz.iab.de/de/FDZ\\_Overview\\_of\\_Data.aspx](http://fdz.iab.de/de/FDZ_Overview_of_Data.aspx)) 08.11.2011

Aus den **Integrierten Erwerbsbiografien**<sup>5</sup> wurden umfassende Informationen über die individuelle Beschäftigtenhistorik, Leistungsbezug, die Teilnahme an Maßnahmen der

<sup>4</sup> Ausführliche und aktuelle Informationen zu den bestehenden Datenquellen des IAB befinden sich auf der Homepage des Forschungsdatenzentrums in Form von Daten- bzw. Methodenberichten oder Projektbeschreibungen (<http://fdz.iab.de/>). Des Weiteren sei hier auf den BASiD-Datenreport des IAB verwiesen. Siehe Hochfellner, Daniela; Müller, Dana; Wurdack, Anja (2011): BASiD - Biografiedaten ausgewählter Sozialversicherungsträger in Deutschland. FDZ-Datenreport, 09/2011 (de), Nürnberg.



## BASiD – Biografiedaten ausgewählter Sozialversicherungsträger in Deutschland 2007

---

aktiven Arbeitsförderung sowie auch Informationen zur Arbeitssuche aus dem Bewerberangebot in den BASiD-Biografiedatensatz übernommen. Die IEB erfassen zudem den Bezug von Arbeitslosengeld II ab dem 01.01.2005, welches im Zuge der „Hartz-Reformen“ die frühere Arbeitslosen- und Sozialhilfe zusammenfasste bzw. ablöste.

In die IEB laufen hierbei Informationen aus der **Beschäftigtenhistorik**, die Entgeltmeldungen zu sozialversicherungspflichtiger Beschäftigung (d.h. im Wesentlichen Jahres-, Ab-, Unterbrechungs- und Änderungsmeldungen) für Westdeutschland seit 1975 und seit 1992 für Ostdeutschland enthält, die wiederum aus den Meldungen der Betriebe im Rahmen des Meldeverfahrens zur Sozialversicherung stammen. Seit April 1999 können hier auch Meldungen zu geringfügiger Beschäftigung abgebildet werden. Zum anderen stammen die Informationen aus der **Leistungsempfängerhistorik**. In dieser finden sich ebenfalls ab dem Jahr 1975 alle Abmeldungen zum Bezug von Arbeitslosengeld, Arbeitslosenhilfe und Unterhaltsgeld.

In den **Maßnahme-Teilnahme-Historiken** werden Teilnahmen an Maßnahmen der aktiven Arbeitsmarktförderung abgebildet. Dazu gehören Arbeitsbeschaffungsmaßnahmen und Strukturanpassungsmaßnahmen sowie das Einstiegsgeld nach § 29 SGB II, die freie Förderung und Maßnahmen der beruflichen Weiterbildung. Bei Verwendung der IEB ist zu beachten, dass Informationen zu Maßnahmeteilnahmen sowie zu Zeiten der Arbeitssuche erst ab dem 01.01.2000 verfügbar sind, während die Informationen aus der Beschäftigtenhistorik und Leistungshistorik bereits für den Zeitraum beginnend mit dem 01.01.1975 genutzt werden können.

Informationen zu den Betrieben, in denen die Untersuchungspersonen beschäftigt waren und/oder sind, werden dem **Betriebs-Historik-Panel** entnommen, das alle Betriebe in Deutschland, die zum 30.06. eines Jahres mindestens einen sozialversicherungspflichtig Beschäftigten beziehungsweise ab dem 01.01.1999 mindestens einen geringfügig Beschäftigten zählen, enthält. Das BHP wird für Westdeutschland seit 1975, für Ostdeutschland seit 1992 aus jährlich gezogenen Querschnittsdaten gezogen und zu einem Panel verknüpft.<sup>6</sup>

---

<sup>5</sup> Einen Einblick in die IEB gewährt u.a. die Publikation von Dorner, Matthias; Heining, Jörg; Jacobebbinghaus, Peter; Seth, Stefan (2010): Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB) 1975-2008. FDZ - Datenreport, 01/2010 (de), Nürnberg.

<sup>6</sup> Für eine eingehende Variablenbeschreibung Siehe Jacobebbinghaus, Peter; Seth, Stefan (2007); Oberschachtsiek, Dirk; Scioch, Patrycja; Seysen, Christian; Heining, Jörg (2009); Spengler, Anja (2009) bzw.

## BASiD – Biografiedaten ausgewählter Sozialversicherungsträger in Deutschland 2007

### 2.2 Zusatzmerkmale

Durch die Nutzung der unterschiedlichen Quellen bietet der BASiD-Datensatz mehr Informationen als die bisher verfügbaren Versicherungsverläufe. Die folgende Abbildung gibt noch einmal die grundlegenden Inhalte des BASiD-Datensatzes wieder<sup>7</sup>:

Schaubild 2: Wesentliche Inhalte des neuen Datensatzes

<u>Querschnittsmerkmale (fixer Teil):</u>	<u>Verlaufsmerkmale (variabler Teil):</u>
➤ Geschlecht	➤ Ausbildung (BA/ IAB)
➤ Geburtsdatum	➤ Erwerbsstatus (BA/ IAB)
➤ Informationen zu Kindern (u.a. Anzahl, Geburtsangaben)	➤ Berufsstellung (BA/ IAB)
➤ Wohnort auf Bundeslandebene (letzte gespeicherte Information) (RV)	➤ Arbeitssuche (Zusatzinfo)
➤ Berufsvariablen (letzte gespeicherte Information) (RV)	➤ Maßnahmen (z.B. ABM, Weiterbildung) (Zusatzinfo)
- Berufsordnung	➤ Betriebsinformationen
- Stellung im Beruf	- Wirtschaftszweig
- Ausbildung	- Betriebsgröße
➤ Versicherungsart/-gruppe	- Betriebsnummer (anonymisiert)
➤ Werte aus der Rentenberechnung	- Lohninformationen
	- Beschäftigtenstruktur
	➤ Wohn- und Arbeitsort (BA/ IAB)
	➤ Berufsklassifikationen (Blossfeld, EGP, MPS)
	➤ Soziale Erwerbssituation (u.a. Erwerbstätigkeit, Arbeitslosigkeit, Altersteilzeit, Rente, Krankheit/ Arbeitsunfähigkeit, Pflege, Kindererziehung) (RV)
	➤ parallele Zustände (Zusatzmerkmal)

Quelle: eigene Zusammenstellung

Hethy-Maier, Tanja; Seth, Stefan (2010): Das Betriebs-Historik-Panel (BHP) 1975-2008 \* Handbuch Version 1.0.2. FDZ -Datenreport, 04/2010 (de), Nürnberg.

<sup>7</sup> Da administrative Daten auf dem Meldeverfahren der Sozialversicherung basieren, können durch den BASiD Datensatz allerdings nur Erwerbsverläufe von sozialversicherungspflichtig Beschäftigten abgebildet werden. Selbstständige finden sich lediglich in dem Fall der freiwilligen Versicherung in der RV.

### 3. Zusammenführung der BA- und der RV-Daten

#### 3.1 Prinzip der Datenzusammenführung

Die Versicherungskontenstichprobe (VSKT) stellt die Grundlage für den BASiD-Datensatz dar, da sie auf einer einheitlichen Datenquelle, dem Versicherungskonto, basiert und i.d.R. größere Biografie-Zeitfenster abbildet als die Datensätze der Bundesagentur für Arbeit (BA). Des Weiteren lässt die VSKT eine Güteprüfung der in ihr enthaltenen Informationen durch den Kontoklärungsstand zu.<sup>8</sup> Die VSKT ist als Panel organisiert, um möglichst auf geprüfte Versicherungskonten zurückgreifen zu können. Im Panel können die RV-Konten aktuell gehalten werden. Bei einer jährlichen Neuziehung müssten jeweils alle Fälle von Grund auf geprüft werden.

Bei der Datenzusammenführung werden Meldungen aus den Daten der BA an die entsprechenden Datenblöcke der VSKT herangespielt. Dabei lassen sich aus Sicht der VSKT zwei Arten von Daten der BA unterscheiden, die die Informationsbreite der VSKT erweitern können: **Datenzwillinge** und **Zusatzinformationen**.

##### 3.1.1 Datenzwillinge

Dies sind Zeiten, die sowohl an die Rentenversicherung als auch an die Bundesagentur für Arbeit im Rahmen des Meldeverfahrens der Sozialversicherung für denselben Zeitraum übermittelt werden. Im Idealfall können diese Angaben in beiden Datenquellen identifizierbar sein. Ist diese Identifikation gelungen, dann ist es möglich den jeweiligen VSKT-Datenblock durch Merkmale aus den IAB-Daten zu ergänzen (z.B. ausgeübter Beruf, Angaben zur Beschäftigung).

In diesem Kontext stellt sich die Frage: Warum ist das Zusammenspielen der Daten überhaupt nötig, wenn die Datenquelle (das Meldeverfahren zur Sozialversicherung) dieselbe ist? Die Antwort dazu ist zweigliedrig. Zum einen setzt die **RSVwV<sup>9</sup> als Rechtsverordnung** zur Ermittlung von **Statistikdaten** der Rentenversicherung enge Grenzen, sodass sich in der VSKT nicht alle Informationen wiederfinden, die bei den Rentenversicherungsträgern auch in

---

<sup>8</sup> In diesem Zusammenhang gilt es zu beachten, dass für Personen, die bereits in Rente sind, davon ausgegangen werden kann, dass das Rentenversicherungskonto vollständig ist.

<sup>9</sup> Die RSVwV stellt die allgemeine Verwaltungsvorschrift über die Statistik in der Rentenversicherung dar.

# Benutzerhinweise Methodische Umsetzung

## BASiD – Biografiedaten ausgewählter Sozialversicherungsträger in Deutschland 2007

---

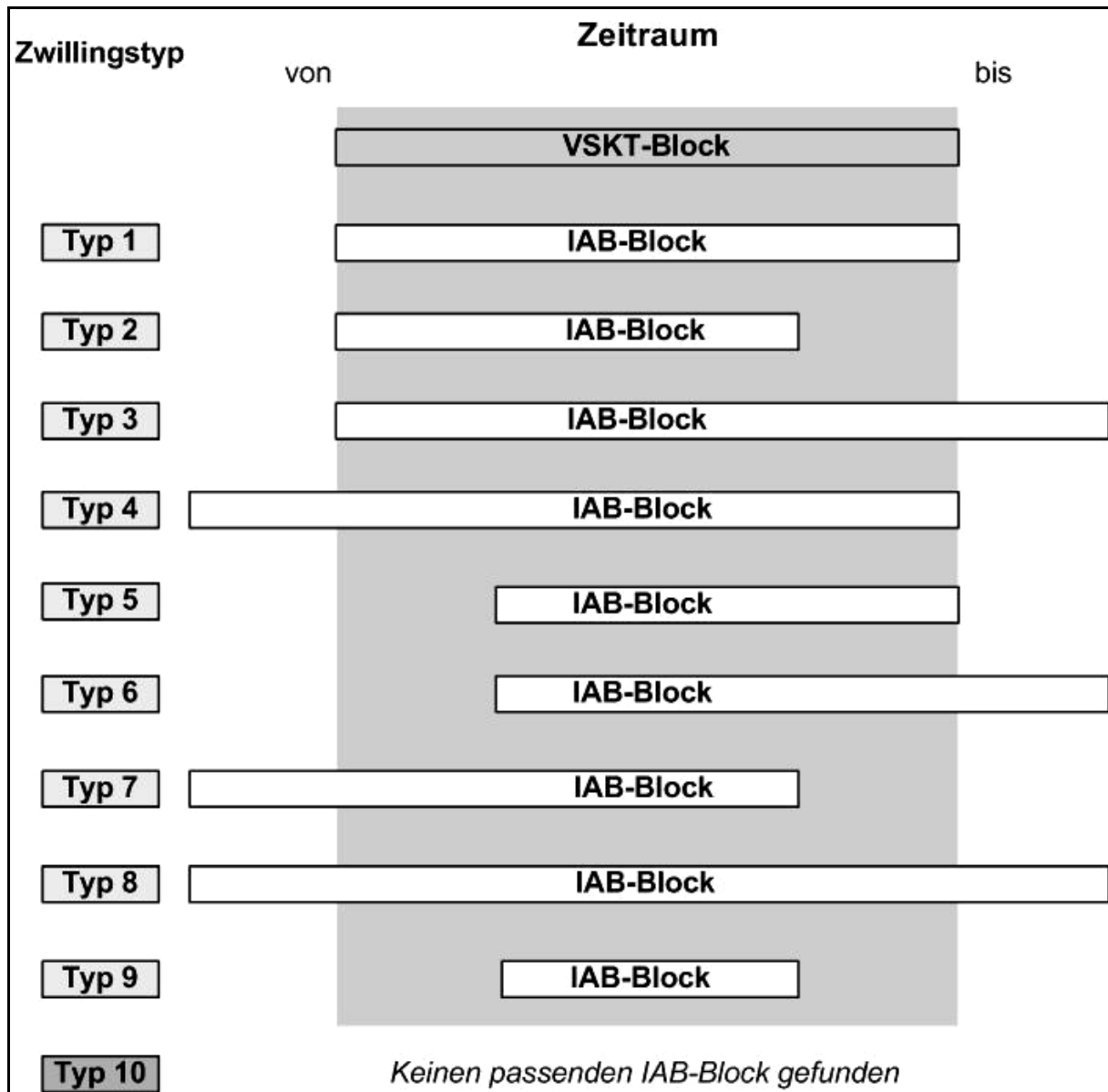
den Konten gespeichert sind. Zum anderen werden im Institut für Arbeitsmarkt- und Berufsforschung (IAB), das in der BA angesiedelt ist, mehrere prozessproduzierte Datenquellen zusammengeführt. Dadurch wird dort eine Anreicherung der Daten aus den Sozialversicherungsmeldungen erzielt (z.B. mit Betriebsdaten).

In den Daten der BA wird nach Datenzwillingen zur VSKT mit Meldungen zur **sozialversicherungspflichtigen Beschäftigung**, **Altersteilzeit**, **geringfügigen Beschäftigung** sowie **Arbeitslosigkeit** gesucht. Gleichzeitig erfolgt ein Abgleich der Tagentgeltpunkte aus der VSKT mit den Tagentgeltpunkten aus den Daten der BA für den entsprechenden Zeitraum. Jedoch können nicht immer perfekte Zwillingsblöcke, die inhaltlich und zeitlich genau übereinstimmen, gefunden werden. Für einige VSKT-Blöcke existieren zwar **inhaltlich**, aber nicht zeitlich exakt **passende IAB-Datenblöcke**. Im Schaubild 3 werden die möglichen Zwillingstypen vorgestellt.

Im Folgenden soll anhand von zwei Beispielen die Interpretation des Schaubildes 3 verdeutlicht werden. Prinzipiell bildet die VSKT die Referenzdatenblöcke, mit denen die IAB-Datenblöcke hinsichtlich ihrer zeitlichen Überschneidung mit den VSKT-Datenblöcken verglichen werden. Bei beiden Beispielen handelt es sich um inhaltlich passende Datenzwillinge. **Zwillings-Typ 1** beschreibt einen perfekten Zwilling, hier stimmt der Zeitraum des VSKT-Blocks exakt mit dem Zeitraum des IAB-Blocks überein. Beim **Typ 5** ist zwar das Ende des IAB-Zeitraums mit dem Ende des VSKT-Zeitraums identisch, jedoch beginnt der IAB-Zeitraum später als der VSKT-Zeitraum.

## BASiD – Biografiedaten ausgewählter Sozial- versicherungsträger in Deutschland 2007

Schaubild 3: Typen von Datenzwillingen



Quelle: eigene Darstellung

Darüber hinaus lassen sich **Zwillingsdatenblöcke mit verwandter Meldung** finden. Bei diesen Zwillingsdatenblöcken können ebenfalls zeitliche Differenzen vorkommen. Dabei kann es sich beispielsweise um eine Meldung für Arbeitslosigkeit handeln, die in der VSKT als Arbeitslosengeld und in den IAB-Daten als Arbeitslosenhilfe abgelegt ist. Ein weiteres Beispiel ist eine Meldung zur sozialversicherungspflichtigen Beschäftigung oder Altersteilzeit in der VSKT, für die Beschäftigungsangaben im weiter gefassten Sinne verwandte Meldungen darstellen. Dazu gehören Beschäftigungsgruppen wie Hausgewerbetreibende,

## BASiD – Biografiedaten ausgewählter Sozial- versicherungsträger in Deutschland 2007

---

behinderte Menschen in anerkannten Werkstätten oder gleichartigen Einrichtungen, kurzfristig Beschäftigte bzw. Personen in Einrichtungen der Jugendhilfe, Berufsbildungswerken oder ähnlichen Einrichtungen für behinderte Menschen. Für Episoden zur Geringfügigkeit wird nicht nach verwandten Meldungen gesucht.

Die Gründe für die inhaltlichen Differenzen zwischen den Meldungen bei der RV und BA für denselben Zeitraum können in rechtlichen Zusammenhängen liegen (rentenrechtlich ergibt sich beitragsstechnisch kein Unterschied, ob diese oder jene Meldung erfolgt) und / oder im Zustandekommen der Datenmeldung (z.B. Trennung der Meldetatbestände, etwa Arbeitslosengeld und Arbeitslosenhilfe, in der Statistikmeldung zur VSKT erst ab einem bestimmten Berichtsjahr). Daneben gibt es VSKT-Blöcke, zu denen keine Informationen in den Daten der BA ermittelt werden können. Ihr Anteil an allen an der Datenzusammenführung beteiligten VSKT-Datenblöcken liegt unter 3%. Die Gründe für diese Mismatches lassen sich nicht in allen Fällen gänzlich aufklären.

Für bestimmte VSKT-Blöcke wird überhaupt nicht nach Datenzwillingen in den Daten der BA gesucht. Dazu zählen vor allem VSKT-Datenbestände mit Zeiten in der ehemaligen DDR bis einschließlich dem Jahr 1991 sowie VSKT-West-Datenblöcke bis einschließlich dem Jahr 1974, da die Daten der BA für Personen aus den alten Bundesländern ab 1975 und für Personen aus den neuen Bundesländern erst ab 1991 (valide ab 1992) vorliegen.

Damit das Zusammenspielen, das grundsätzlich auf der Episodenebene<sup>10</sup> erfolgt, gelingt, müssen die RV- und die BA-Daten zuvor so aufbereitet werden, dass die inhaltlich zusammengehörenden Datenblöcke vergleichbare Einheiten (Jahresmeldungen) bilden und sinnvolle Datenzwillinge ergeben. Dieses Ziel wird durch die Fusion von mehreren, inhaltlich identischen Datenblöcken innerhalb eines Jahres bzw. durch das Splitting von Datenmeldungen, die über mehrere Jahre gehen, erreicht. Tabelle 1 zeigt, welches der beiden genannten Verfahren bei den jeweiligen Meldungen zur Anwendung gekommen ist.

Die Variable **QS** gibt u.a. die Qualität des Datenzwillings an. Sie unterscheidet zwischen inhaltlich exakt passenden Meldungen in den RV- / BA-Daten (Ausprägung 1 und 2) und in-

---

<sup>10</sup> Beim Datenmatching liegen sowohl die RV- als auch die IAB-Daten im Episodendatenformat vor. Dabei ist die Datenmatrix so angeordnet, dass jede Zeile eine Episode (= Zeitabschnitt, in einem bestimmten Zustand, zwischen zwei aufeinanderfolgenden Zustandswechseln) und nicht ein Individuum darstellt. Die Zahl der Episoden pro Individuum ist verschieden.

## BASiD – Biografiedaten ausgewählter Sozialversicherungsträger in Deutschland 2007

haltlich verwandten Meldungen aus beiden Datenquellen (Ausprägung 3 und 4). Mit „zeitlichen Differenzen“ bei den Ausprägungen 2 und 4 sind die Zwillingstypen 2 bis 9 gemeint, die das Schaubild 3 darstellt.

Tabelle 1: Aufbereitung der Daten für die Datenzusammenführung

VSKT-Meldung	VSKT-Datei	IAB-Datei
1. Meldungen zur sozialversicherungspflichtigen Beschäftigung	unterjährige Fusion	--
2. Meldungen Altersteilzeit	unterjährige Fusion	--
3. Meldungen Geringfügigkeit	unterjährige Fusion	unterjährige Fusion
4. Meldungen Arbeitslosigkeit	Datenblöcke werden auf Jahreslänge gesplittet, dabei werden die Zeitraummonate angepasst	Datenblöcke werden auf Jahreslänge gesplittet, dabei werden die Zeitraummonate angepasst

Quelle: eigene Zusammenstellung

### 3.1.2 Zusatzinformationen

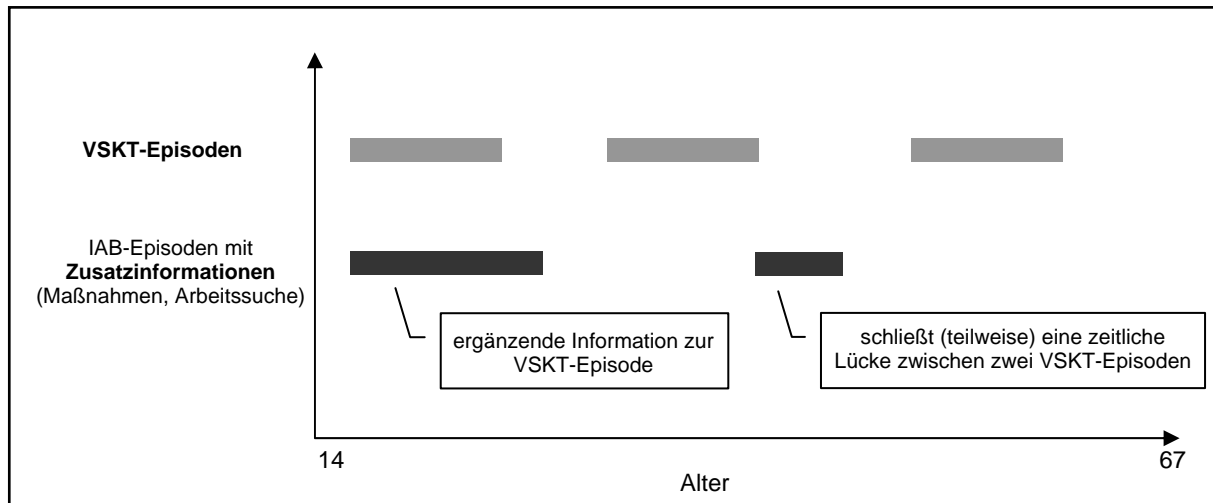
Neben den so genannten Datenzwillingen verfügen die Daten der BA auch über Informationen zur **Arbeitssuche** und zu **Maßnahmen** wie z. B. berufliche Qualifizierungen, an denen eine Person teilgenommen hat. Diese Meldungen haben keine Entsprechung in den Daten der Rentenversicherung, deshalb muss für sie kein Datenzwilling gesucht werden. Die Funktion der Zusatzinformationen wird im Schaubild 4 dargestellt. Sie können zeitlich parallelen Episoden in der VSKT als Zusatzinformation zugeordnet werden, wenn z.B. eine sozialversicherungspflichtig erwerbstätige Person gleichzeitig auf der Suche nach einer neuen Arbeitsstelle ist. Die ergänzende Zusatzinformation ist hier die Arbeitssuche. Genauso können Zusatzinformationen auch zeitliche Lücken der statistischen Daten aus dem Rentenversicherungskonto auffüllen.

Genau wie bei den Datenzwillingen gibt es auch bei Blöcken mit Zusatzinformationen solche, die zeitlich perfekt zum entsprechenden VSKT-Block passen und andere, bei denen die zeitliche Übereinstimmung nicht gegeben ist bzw. überhaupt keine Zusatzinformation

## BASiD – Biografiedaten ausgewählter Sozial- versicherungsträger in Deutschland 2007

gefunden werden konnte. Die Variable **QS\_ZUSATZ** gibt Auskunft darüber, wie gut der Ergänzungsblock aus den Daten der BA zum VSKT-Datenblock passt.

Schaubild 4: Zusatzinformationen im BASiD-Datensatz



Quelle: eigene Darstellung

### 3.2 Ablauf der Zwillingsuche

Grundmodell der Zwillingsuche<sup>11</sup> ist, dass sich für jeden Datenblock in der VSKT, der sich aus einer Meldung zur Sozialversicherung ergibt, ein gleichartiger und zeitlich identischer Datenblock in den IAB-Daten findet. Da es aber häufiger zu zeitlichen und inhaltlichen Unschärfen zwischen den Datenblöcken in den RV- und den BA-Daten kommt, wurde die Datenfusion zustandsbezogen und jeweils in mehreren Schritten durchgeführt.

Das Zusammenspielen der Datenblöcke mit Meldungen zur **sozialversicherungspflichtigen Beschäftigung** und zur **Altersteilzeit** verläuft in zwei Schritten. In einem ersten Suchlauf werden die VSKT-Datenblöcke mit **inhaltlich exakt passenden IAB-Datenblöcken** abgeglichen. Die so gefundenen Datenzwillinge können jedoch unterschiedliche zeitliche Übereinstimmungen aufweisen, wie in Schaubild 3 bereits erläutert. Sie werden in einer Ergebnisdatei abgelegt und somit mit Informationen, die ursprünglich nur am IAB-Datenblock hingen, angereichert (z.B. Betriebsmerkmale) und die Variable QS wird mit den Werten von 1 bis 2 beschickt. Die übrigen VSKT-Datenblöcke, für die im ersten Suchlauf keine inhaltlich exakten Datenzwillinge ermittelt werden konnten, gehen in einen zweiten Suchlauf ein, in

<sup>11</sup> Vgl. Anlage: Systematik des Datenabgleichs zwischen VSKT und Daten des IAB



### BASiD – Biografiedaten ausgewählter Sozialversicherungsträger in Deutschland 2007

---

dem sie mit **inhaltlich verwandten IAB-Datenblöcken** abgeglichen werden. Entsprechend erfolgt hier die Beschickung des Merkmals QS mit den Ausprägungen 3 und 4. Die im zweiten Suchlauf bestimmten Datenzwillinge werden ebenfalls in der Ergebnisdatei gespeichert. Auch bei diesen Datenzwillingen können sich, analog zu denen aus dem ersten Suchlauf, zeitliche Abweichungen ergeben.

Die Zwillingssuche von Datenblöcken mit Meldungen zur **geringfügigen Beschäftigung** erfolgt in nur einem Suchlauf, analog zum Suchlauf 1 für die Datenblöcke mit sozialversicherungspflichtiger Beschäftigung und Altersteilzeit.

Bei Datenblöcken mit Meldungen zur **Arbeitslosigkeit** findet das Zusammenspielen in drei Schritten statt. Die ersten zwei Suchläufe sind mit den beiden Suchläufen der Datenblöcke mit Meldungen zur sozialversicherungspflichtigen Beschäftigung bzw. Altersteilzeit vergleichbar. Im dritten Suchlauf werden **VSKT-Datenblöcke mit RV-spezifischen Meldungen** zur Arbeitslosigkeit (z.B. Anrechnungszeiten<sup>12</sup>) mit IAB-Datenblöcken zur Arbeitslosigkeit abgeglichen.

### 3.3 Ablauf der Zusatzsuche

Bei der Zusatzsuche<sup>13</sup> wird ein VSKT-Datenblock mit den IAB-Datenblöcken, die Informationen zu Maßnahmen und Arbeitssuche (= Zusatzinformationen) enthalten, abgeglichen. Wird ein passender IAB-Datenblock gefunden, so wird der entsprechende VSKT-Datenblock in der Ergebnisdatei um die Zusatzmerkmale ergänzt. Dabei stimmt der Datenblock mit der Zusatzinformation zeitlich nicht unbedingt genau mit dem entsprechenden VSKT-Block überein. Auch hier sind dieselben Kombinationsmöglichkeiten denkbar wie in Schaubild 3 beschrieben. Ob eine Zusatzinformation vorliegt oder nicht, lässt sich an den Dummy-Variablen ARSU (Arbeitssuche), ARSURE (Arbeitssuche registriert), ABMSAM (Arbeitsbeschaffungs-/Strukturanpassungsmaßnahmen), BQW (Berufliche Qualifizierung und Weiterbildung) bzw. SOLEI (Sonstige Leistung) erkennen. Auf dieselbe Art und Weise

---

<sup>12</sup> Anrechnungszeiten schaffen einen Ausgleich für Zeiten, in denen der Versicherte aus bestimmten persönlichen Gründen (geregelt in §§ 58 Abs. 1, 252 Abs. 1, 252a Abs. 1 SGB VI) an der Beitragsleistung gehindert war. Zu solchen Anrechnungstatbeständen gehört beispielsweise die Arbeitslosigkeit oder Ausbildungssuche mit Meldung bei einem deutschen Arbeitsamt oder Krankheit mit Arbeitsunfähigkeit.

<sup>13</sup> Vgl. Anlage: Systematik des Datenabgleichs zwischen VSKT und Daten des IAB

# Benutzerhinweise Methodische Umsetzung

## BASiD – Biografiedaten ausgewählter Sozial- versicherungsträger in Deutschland 2007

---

werden im Rahmen der Zusatzsuche auch Lücken zwischen zwei VSKT-Datenblöcken aufgefüllt. Wird kein passender Block mit Zusatzinformationen aus den IAB-Daten gefunden, wandert der VSKT-Datenblock unverändert in die Ergebnisdatei und die Variable QS\_ZUSATZ wird mit „3“ beschickt.

# Benutzerhinweise Methodische Umsetzung

## BASiD – Biografiedaten ausgewählter Sozial- versicherungsträger in Deutschland 2007

### 3.4 Ablaufschema der Datengenerierung

Die Produktion des SUF BASiD und des dabei entstehenden Endproduktes wird in Tabelle 2 veranschaulicht. Die Datengenerierung des SUF BASiD erfolgt analog zum Vorgehen des SUF VSKT (siehe dazu: „Benutzerhinweise - Methodische Umsetzung des SK79 in einen anonymisierten Datensatz fester Länge: Sequentielle Biografiedaten“, Kapitel 7) mit dem einzigen Unterschied, dass hier die VSKT-Verlaufsdaten zuvor mit den Daten der BA erweitert wurden. Wie beim SUF VSKT werden auch beim SUF BASiD die einzelnen Verlaufsmerkmale in separate Datenfiles aufgeteilt, die eine für das Matching dieser Daten an den fixen VSKT-Teil erforderliche eindeutige **Fallnummer** enthalten.

Tabelle 2: SUF BASiD – Endprodukte

<b>Fixer Datenteil</b> (Strukturmerkmale, Soziodemografie, Summenmerkmale der Rentenberechnung)	<b>Biografiebezogene                      Verlaufsmerkmale, Zusatzmerkmale                      aus den Daten des IAB,                      Betriebsmerkmale des IAB</b>
SUFBASiD07_fix	SUFBASiD07_“Merkmalsname“ (z.B. SUFBASiD07_SES)

Quelle: eigene Zusammenstellung

## BASiD – Biografiedaten ausgewählter Sozial- versicherungsträger in Deutschland 2007

---

### 4. Anonymisierung und Datenschutz

Die Aufbereitung der BASiD-Längsschnittdaten ergibt einen Datenbestand, der faktisch anonymisiert ist. Folgende Anonymisierungsschritte wurden nach dem Anonymisierungskonzept bei den fixen Merkmalen vollzogen.

Aus dem fixen Teil des SK79 werden nur Merkmale zur soziodemografischen Profilierung weitergegeben. Diese werden in ihren Ausprägungen den allgemeinen Anonymisierungsregeln des FDZ-RV folgend reduziert. Die ergänzenden Merkmale aus den Daten der Bundesagentur für Arbeit (BA) liegen ebenfalls in einer faktisch anonymisierten Form vor.

#### 1. Direkte Identifikatoren

Die Versicherungsnummer befindet sich nicht im Datensatz. Die interne Zählnummer wird nicht weitergegeben. Für die Datenweitergabe wird der Datensatz neu sortiert und nummeriert. Diese neue Fallnummer wird weitergegeben.

#### 2. Stichprobenziehung

Nach Umsortierung und Vergabe der neuen Fallnummer wird eine Stichprobenziehung nach dem Zufallsverfahren durchgeführt. Es wird eine 25%-**Substichprobe** für eine ausgewählte Population weitergegeben. Dabei findet eine Altersbegrenzung bei den Untersuchungspersonen von 30 bis 67 Jahren statt. Außerdem wird die Auswahl auf im Inland lebende Deutsche beschränkt. Bereits die Weitergabe als Substichproben und deren Umfang macht eindeutige Re-Identifikation nahezu unmöglich.

#### 3. Regionalisierung

Die Regionalangaben werden auf Ost- und West-Deutschland zurückgeführt. Bei dem RV-Merkmal zum Wohnort (WHOT\_OSTWEST) wird in Berlin zwischen Ost und West unterschieden.

#### 4. Staatsangehörigkeiten

Es werden nur deutsche Untersuchungspersonen weitergegeben.

## BASiD – Biografiedaten ausgewählter Sozial- versicherungsträger in Deutschland 2007

---

### 5. Angaben zu Entgeltpunkten

Entgeltpunktinformationen werden im variablen Datenteil auf Monatsebene weitergegeben. Im fixten Datenteil werden Entgeltpunktsummen gerundet und zusammengefasst (siehe Punkt 8).

### 6. Begrenzung von Geldbeträgen

DM- oder Euro-Beträge sind ganzzahlig gerundet nach oben begrenzt.

### 7. Datumsangaben im variablen Teil

Datumsangaben (Von/Bis-Zeiträume) werden nicht weitergegeben. Die Informationen zum Biografieverlauf werden als monatliche Angaben aufbereitet.

### 8. Werte zur Rentenberechnung im fixen Datenteil

Die Werte zur Rentenberechnung, die in AR/AV bzw. KN und Ost/West aufgeteilt sind, wurden zusammengefasst. In der Regel wird nicht nach Versicherungszweigen unterschieden. Eine Zusammenfassung von Entgeltpunktsummen Ost und West geschieht bei ausgewählten Merkmalen.

### 9. Anonymisierung der Längsschnittdaten

Neben der Anonymisierung der Querschnittsinformationen wird mit dem Ziel der faktischen Anonymisierung eine Aggregation der Verlaufsmerkmale vorgenommen.

Die detaillierten rentenrechtlichen Längsschnittinformationen, die sich aus den Variablen BYAT und BYATSO ergeben, werden reduziert auf 15 Zustände (so genannte SES). Ferner wird nur eine Monatsangabe weitergegeben und keine tagegenauen Informationen. Selbst die Kenntnis über rentenrechtliche Sachverhalte mit zeitlichem Bezug kann damit nicht mehr zur Re-Identifikation herangezogen werden.

## BASiD – Biografiedaten ausgewählter Sozial- versicherungsträger in Deutschland 2007

---

### 10. Ergänzende Merkmale aus den Daten der BA zum Erwerbsstatus

Angaben zu den ausgeübten Berufen werden auf Variablen zurückgeführt, die nur zusammenfassend über Art und Niveau der Tätigkeit berichtet (Blossfeld-Skala, EGP-Klassenschema, Magnitude-Prestigeskala). Ferner werden Angaben zum Erwerbsstatus nur monatsbezogen weitergegeben (zeitlich näher nicht differenziert).

### 11. Ergänzende Merkmale aus den Daten der BA zu den Betrieben

In den Daten der BA ist die Betriebsnummer grundsätzlich verfügbar, wird jedoch nicht an die RV weitergegeben. Folgende Maßnahmen wurden ergriffen, um eine Reidentifikation von Betrieben und deren Beschäftigten im SUF gemäß der faktischen Anonymität zu erschweren:

#### a) Angaben zum Wirtschaftszweig

Die hinteren Zahlen der Schlüsselkennziffer wurden abgeschnitten, sodass das Merkmal „Wirtschaftszweig“ als 1-Steller im SUF weitergegeben wird und damit nur stark vergrößerte Informationen enthält.

#### b) Betriebsgröße

Die Variable „Betriebsgröße“ wurde entsprechend der Anzahl der Beschäftigten eines Betriebes in fünf Kategorien aufgeteilt.

#### c) Beschäftigte

Bei den Beschäftigtenmerkmalen handelt es sich um Anteilsvariablen, die aus dem Verhältnis zur Gesamtbeschäftigtenzahl ermittelt wurden.

#### d) Bruttotagesentgelt

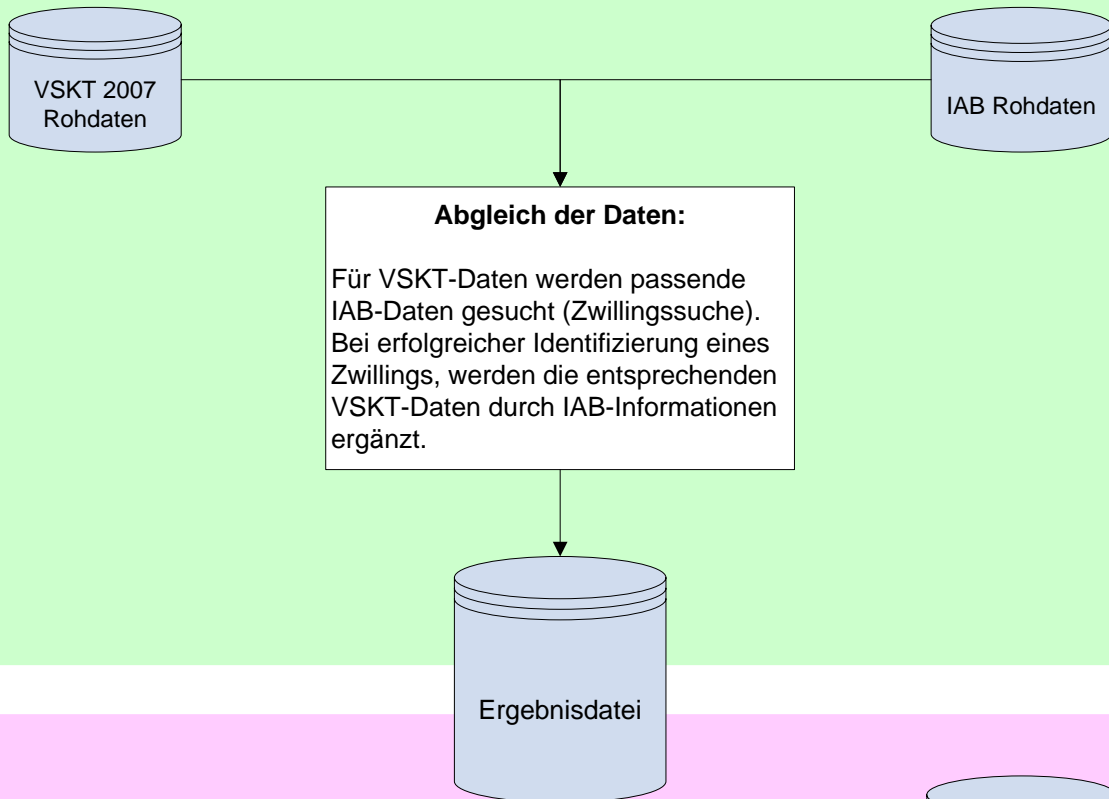
Das Bruttotagesentgelt wird anhand von drei Variablen mit den Quartilswerten 25, 50 und 75 für alle Vollzeitbeschäftigten eines Betriebs weitergegeben.

## **Inhaltsverzeichnis**

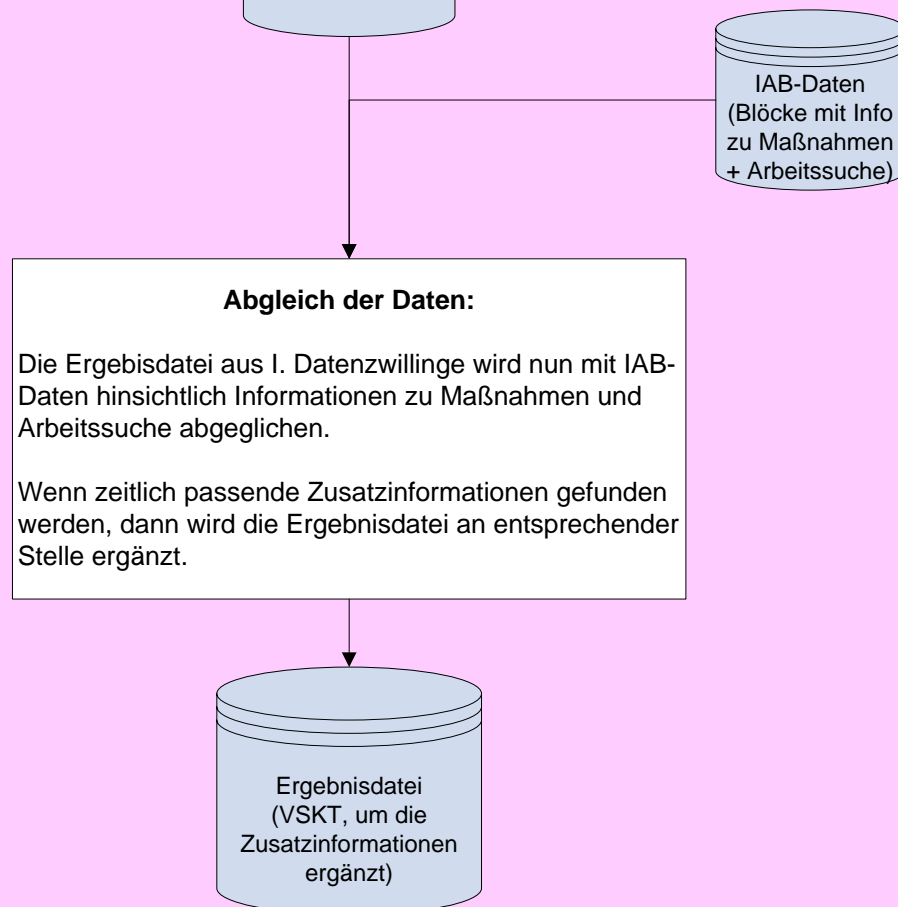
<b>Übersicht über die Systematik des Datenabgleichs</b>	<b>1</b>
<b>I. Datenzwillinge</b>	<b>2</b>
Typen von Datenzwillingen	3
Suche 1: Beschäftigte	4
Suche 2: Altersteilzeit	5
Suche 3: Geringfügigkeit	6
Suche 4: Arbeitslosigkeit	7
<b>II. Zusatzinformationen</b>	<b>8</b>
<b>Variable QS</b>	<b>9</b>
Ausprägung 6 der Variable QS	10
<b>Variable QSZUSATZ</b>	<b>11</b>

# Übersicht der Systematik des Datenabgleichs

## I. DATENZWILLINGE

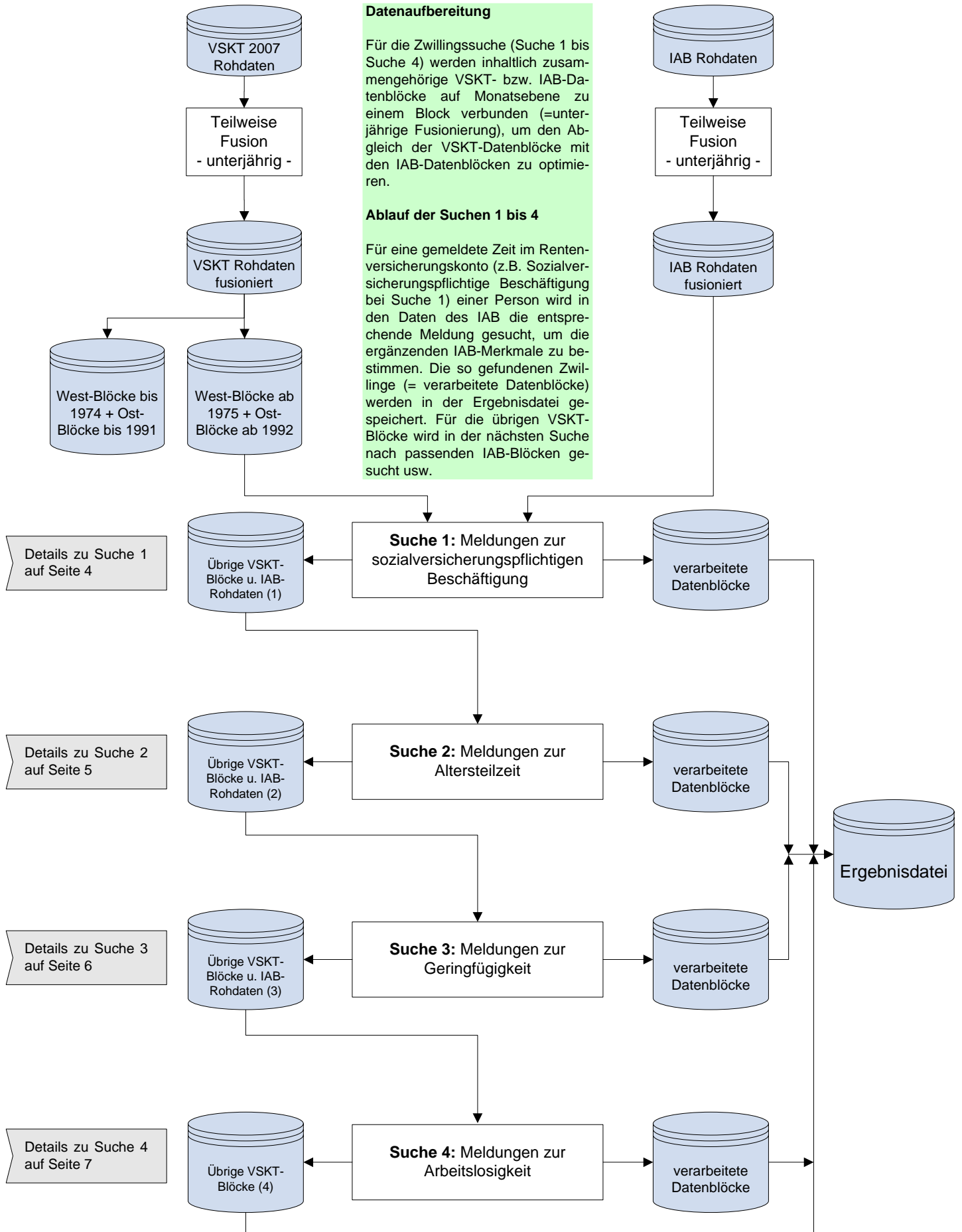


## II. ZUSATZINFORMATIONEN





# I. DATENZWILLINGE



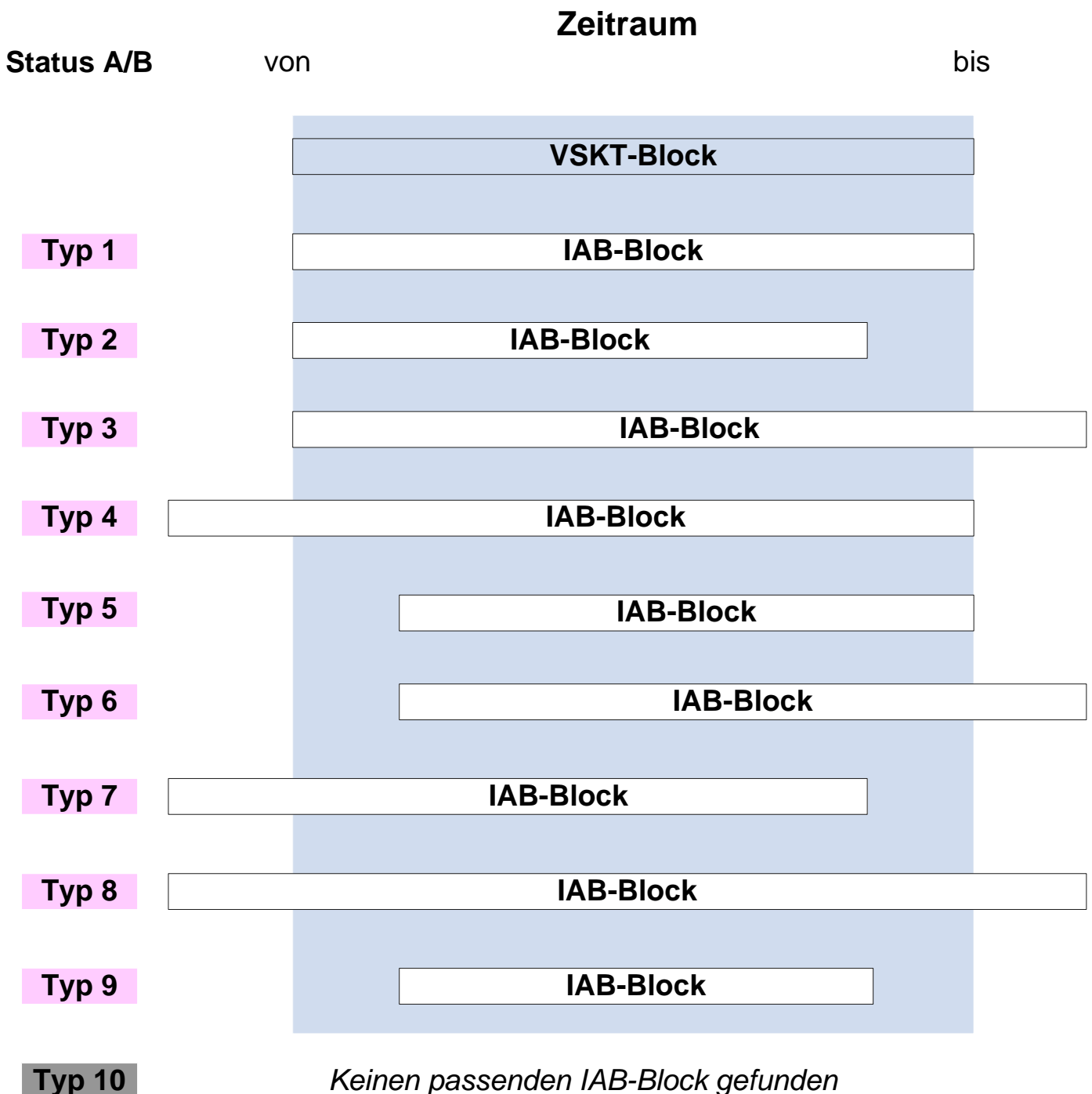
# Typen von Datenzwillingen

## Zeitliche Differenzen zwischen RV- und IAB-Blöcken

Für die RV-Blöcke existieren in den Daten des IAB nicht immer zeitlich deckungsgleiche und inhaltlich passende Blöcke. Es gibt 9 Typen von Datenzwillingen, die bei den Übersichten zu den Suchen 1 bis 4 jeweils anhand von Status A (Suchlauf 1) und Status B (Suchlauf 2) erkennbar sind.

## Beispiele für Typen von Datenzwillingen

- a) Typ 1 = perfekter Zwillig, d.h. der Zeitraum des VSKT-Blocks stimmt exakt mit dem Zeitraum des IAB-Blocks überein
- b) Typ 5 = das Ende des IAB-Zeitraums ist zwar mit dem Ende des VSKT-Zeitraumes identisch, jedoch beginnt der IAB-Zeitraum später als der VSKT-Zeitraum



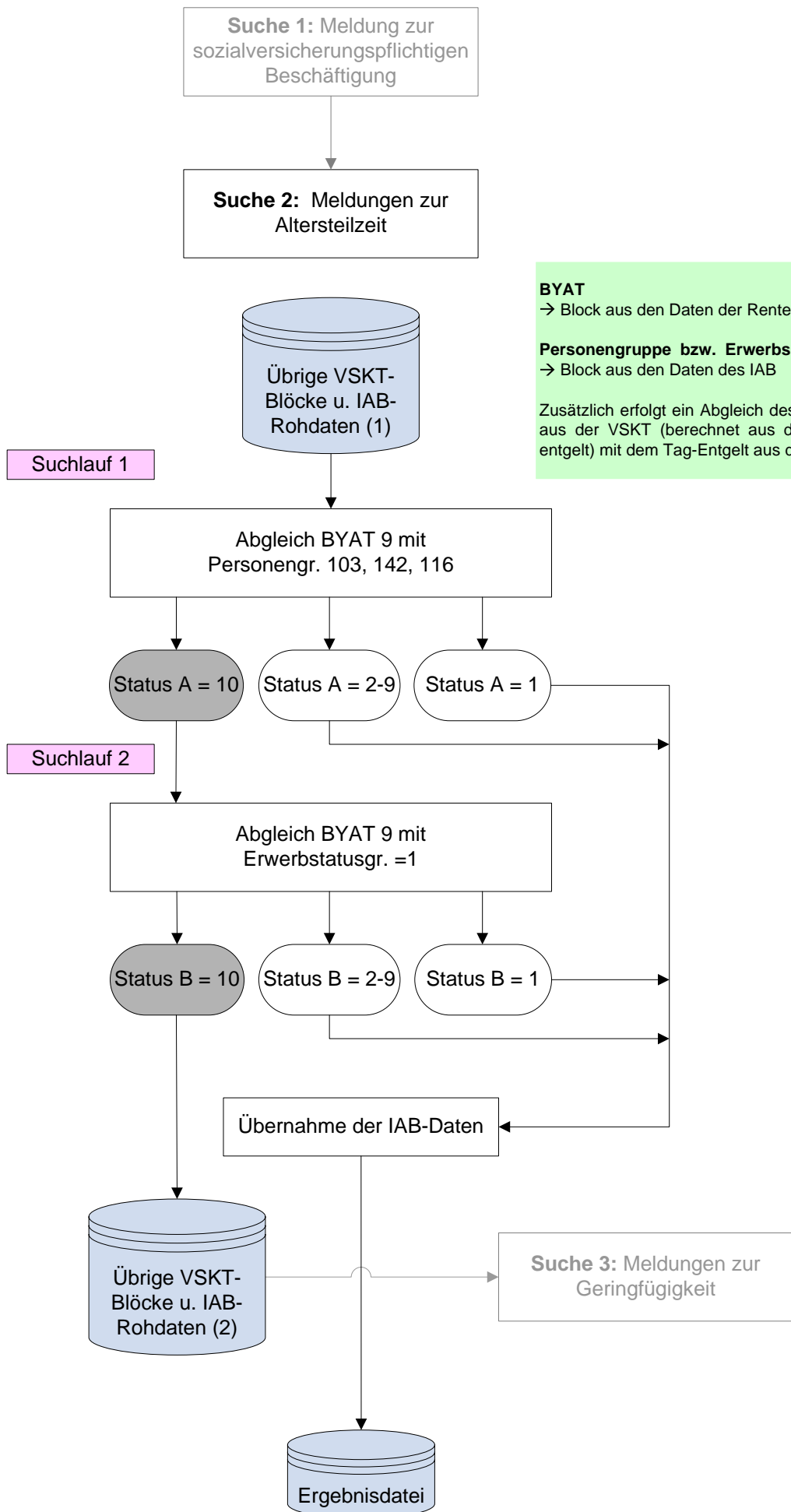
**Suche 1: Meldungen zur sozialversicherungspflichtigen Beschäftigung**



**BYAT**  
 → Block aus den Daten der Rentenversicherung

**Personengruppe bzw. Erwerbsstatusgruppe**  
 → Block aus den Daten des IAB

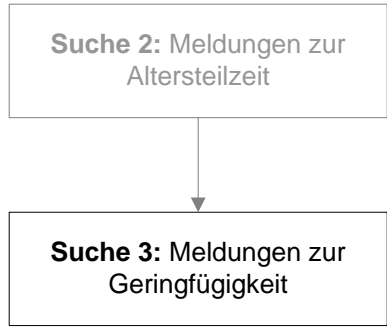
Zusätzlich erfolgt ein Abgleich des Tag-Entgelts aus der VSKT (berechnet aus dem Zeitraum-entgelt) mit dem Tag-Entgelt aus den BA-Daten.



**BYAT**  
 → Block aus den Daten der Rentenversicherung

**Personengruppe bzw. Erwerbstatusgruppe**  
 → Block aus den Daten des IAB

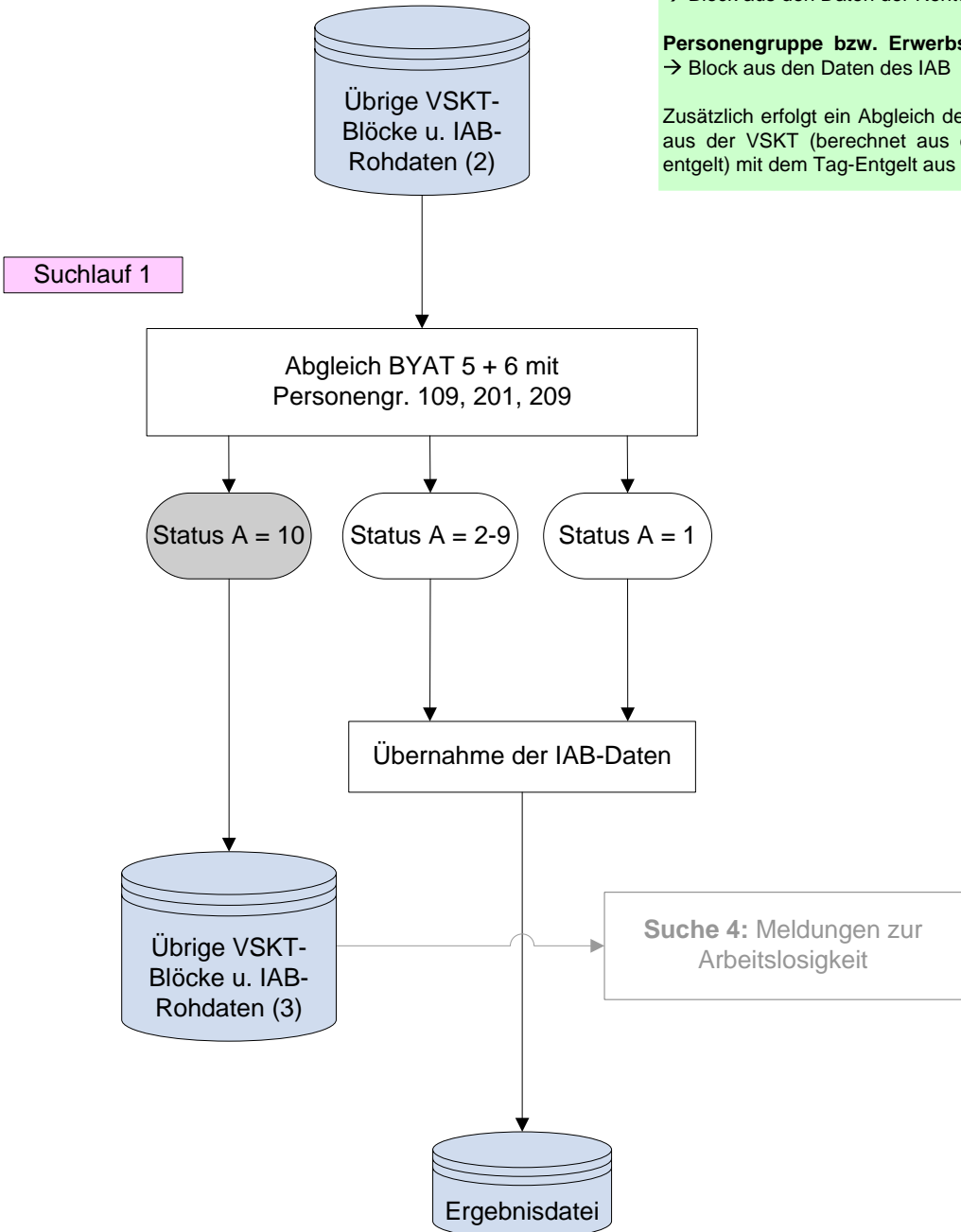
Zusätzlich erfolgt ein Abgleich des Tag-Entgelts aus der VSKT (berechnet aus dem Zeitraum-entgelt) mit dem Tag-Entgelt aus den BA-Daten.



**BYAT**  
→ Block aus den Daten der Rentenversicherung

**Personengruppe bzw. Erwerbsstatusgruppe**  
→ Block aus den Daten des IAB

Zusätzlich erfolgt ein Abgleich des Tag-Entgelts aus der VSKT (berechnet aus dem Zeitraum-entgelt) mit dem Tag-Entgelt aus den BA-Daten.



Suche 3: Meldungen zur Geringfügigkeit

Suche 4: Meldungen zur Arbeitslosigkeit

Übrige VSKT-Blöcke u. IAB-Rohdaten (3)

Suchlauf 1

Abgleich nach: BYAT = 13 / Erwerbstatusgr. = 3, 6

Abgleich nach: BYAT = 4 / Erwerbstatusgr. = 4, 5

Status A = 10

Status A = 2 - 9

Status A = 1

Status A = 1

Status A = 2 - 9

Status A = 10

Suchlauf 2

Abgleich nach: BYAT = 13 / Erwerbstatusgr. = 4, 5

Abgleich nach: BYAT 4 / Erwerbstatusgr. = 3, 6

Status B = 10

Status B = 2 - 9

Status B = 1

Status B = 1

Status B = 2 - 9

Status B = 10

Übrige VSKT-Blöcke u. IAB-Rohdaten (3)

Suchlauf 3

Abgleich nach: BYAT = 40, 41, 48 +  
BYATSO = 3 / Erwerbstatusgr. = 3, 4, 5, 6

Status A = 10

Status A = 2 - 9

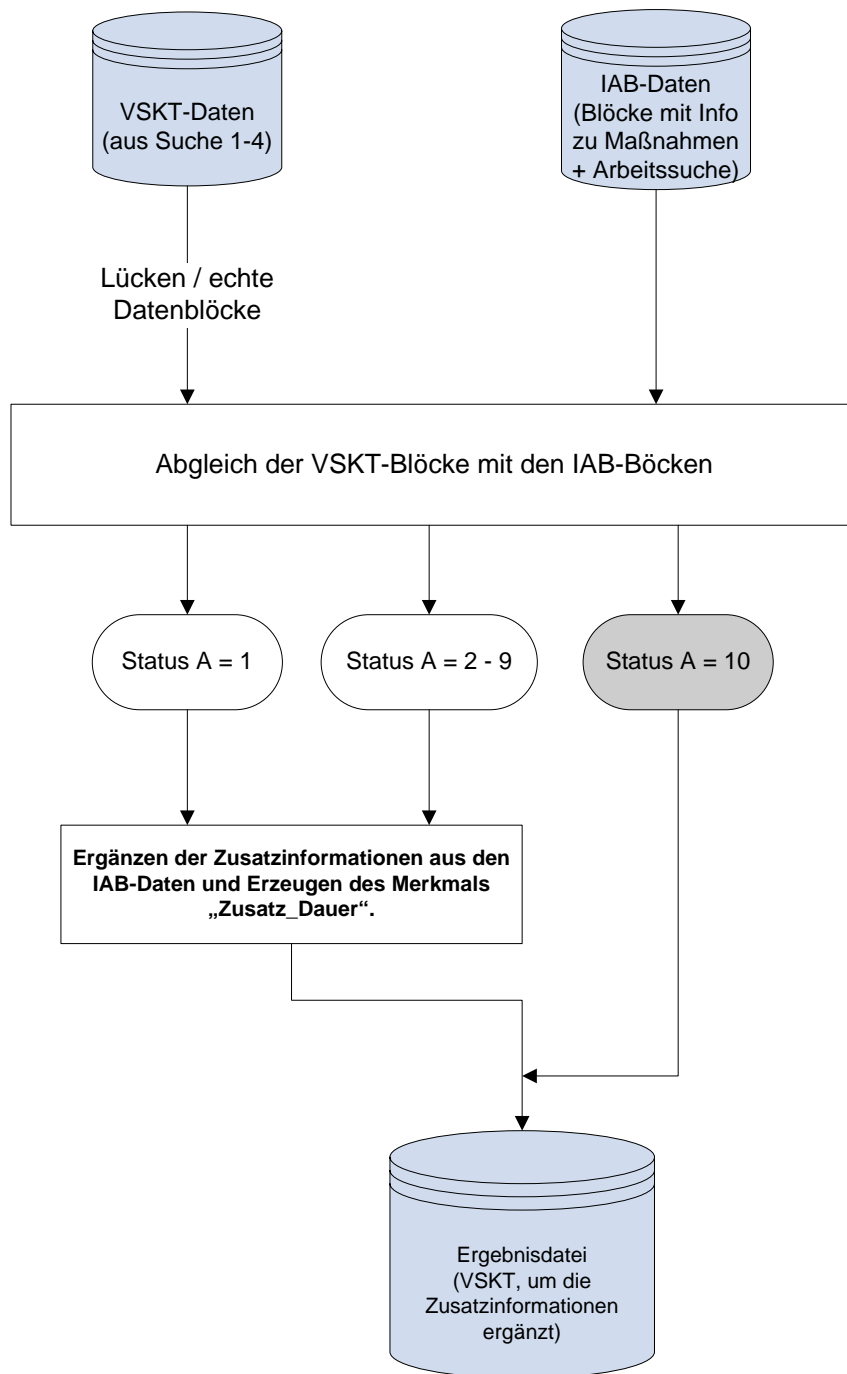
Status A = 1

Übernahme der IAB-Daten

Übrige VSKT-Blöcke (4)

Ergebnisdatei

## II. ZUSATZINFORMATIONEN



### Ablauf:

Ein VSKT-Block wird mit den IAB-Blöcken (mit Informationen zu Maßnahmen und Arbeitssuche (ARSU, ARSURE, ABMSAM, BQW, SOLEI) = Zusatzinformationen) gemäß Status A abgeglichen. Wird ein passender IAB-Block gefunden, so wird der entsprechende VSKT-Block in der Ergebnisdatei um die Zusatzmerkmale ergänzt. Wird kein passender IAB-Block zu dem VSKT-Block gefunden, bekommt der VSKT-Block den Status A = 10 und wandert in die Ergebnisdatei. Zusätzlich wird ein Merkmal generiert, das die Dauer der Überschneidung in Tagen angibt, d. h.

- für Status A = 1 entspricht die Überschneidungsdauer der Dauer des VSKT-Blocks,
- für Status A = 2 - 9 ist die Überschneidungsdauer kleiner oder gleich der Dauer des VSKT-Blocks und
- für Status A = 10 ist die Überschneidungsdauer gleich Null.

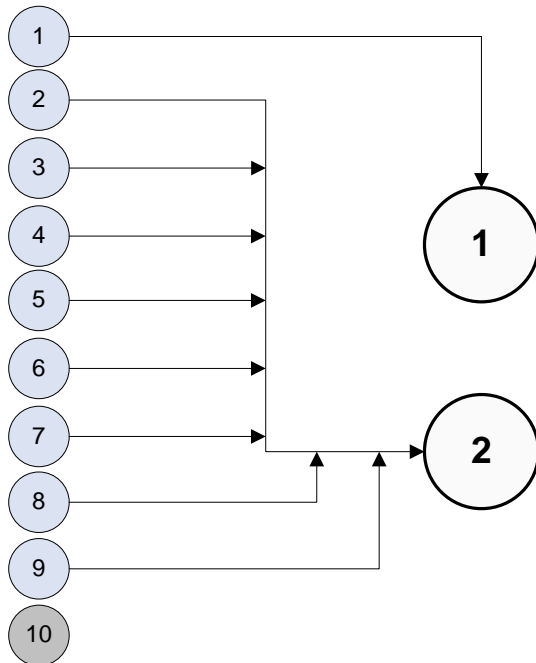
Anschließend ist der nächste VSKT-Block an der Reihe.

Alle IAB-Blöcke mit vorhandenen Zusatzinformationen werden unabhängig von Überschneidungen mit den VSKT-Blöcken in die Ergebnisdatei übernommen. Diese werden mit BYAT = 95 kenntlich gemacht.

# Variable QS (Gültig für die Suchen 1-4)

Status der Zwillinge aus Suchlauf 1

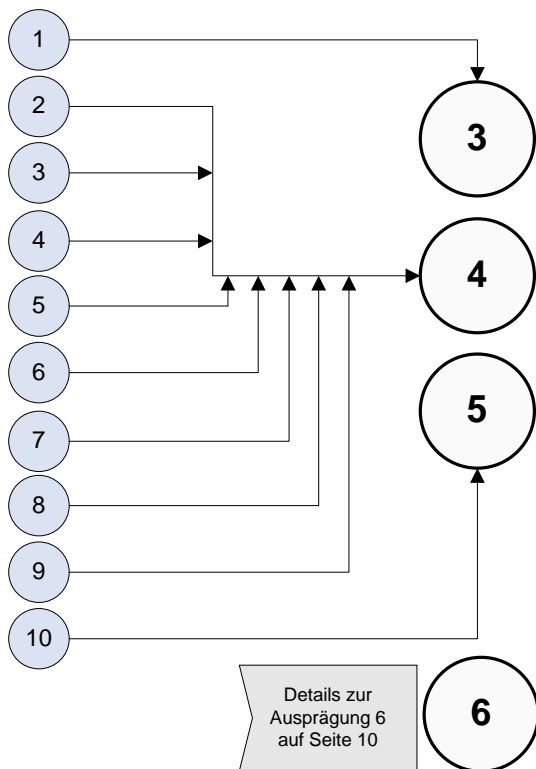
Ausprägungen der Variable QS



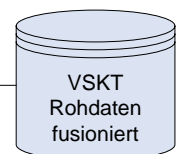
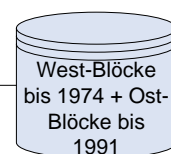
Das Merkmal „QS“ zeigt, inwieweit eine bei der GRV gespeicherte Zeit exakt in den Daten des IAB gefunden wurde bzw. mit welchen Abweichungen.

1	=	zeitidentischen Datenblock gefunden mit entsprechender Meldung
2	=	Datenblock mit entsprechender Meldung gefunden, jedoch zeitliche Differenzen
3	=	zeitidentische Datenblöcke mit verwandter Meldung gefunden
4	=	Datenblock mit verwandter Meldung gefunden, jedoch zeitliche Differenzen
5	=	keine Information gefunden
6	=	nicht nach Informationen gesucht

Status der Zwillinge aus Suchlauf 2



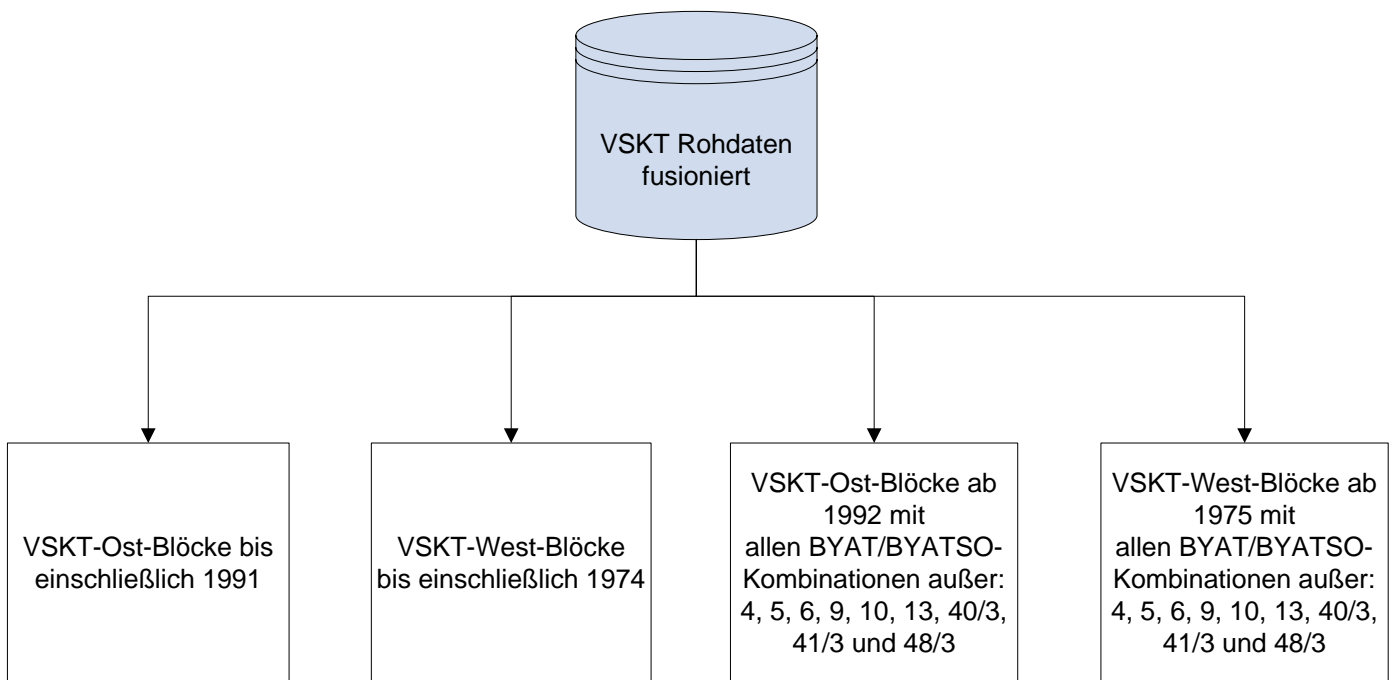
Details zur Ausprägung 6 auf Seite 10





# Ausprägung 6 der Variable QS

Ausprägung 6 der Variable QS „nicht nach Informationen gesucht“ beinhaltet alle VSKT-Blöcke für die in den Daten der BA per Definition (Meldeart und Zeitbezug) keine Speicherungen vorhanden sind. Diese VSKT-Blöcke sind nicht in die Zwillingsuche eingegangen.



# Variable QSZUSATZ (Gültig für die Zusatzsuche)

Das Merkmal „QSZUSATZ“ zeigt, inwieweit eine bei der GRV gespeicherte Zeit exakt in den Daten des IAB gefunden wurde bzw. mit welchen Abweichungen.

1	=	zeitidentischen Datenblock mit Zusatzinformation gefunden
2	=	Datenblock mit Zusatzinformation gefunden, jedoch zeitliche Differenzen
3	=	keine Zusatzinformation gefunden

