

FDZ *Datenreport*

Dokumentation zu Arbeitsmarktdaten

Nr. 01/2007

Dokumentation des

„Scientific Use Files
der Integrierten Erwerbsbiographien“
(IEBS-SUF V1)

Version 1.0

Juli 2007

Ralf Zimmermann, Steffen Kaimer, Dirk Oberschachtsiek



Für Fragen zum IEBS-SUF V1 stehen Ihnen die oben genannten Autoren zur Verfügung.
Fragen zur IEBS und zum IEBS-SUF V1 werden gerne von Dirk Oberschachtsiek,
IAB (0911 179 1615; Dirk.Oberschachtsiek@iab.de beantwortet)

Inhaltsverzeichnis

1. Einleitung	5
2. Aufbau des Scientific Use Files der Integrierten Erwerbsbiografien (IEB)	7
2.1. <i>Inhalt und Umfang der IEB</i>	7
2.1.1. Verknüpfung und Aufbau.....	8
2.1.2. Datenqualität.....	9
2.1.3. Besonderheiten zur Qualität der MTH- und BewA-Meldungen.....	10
2.1.4. Veränderung von Merkmalsausprägungen (Zeitkonsistenz der Merkmale).....	14
2.2. <i>Genese des IEBS-SUF V1</i>	15
2.2.1. Variablenumfang des Scientific Use Files der IEB.....	17
2.2.2. Mengengerüst des IEBS-SUF V1.....	20
3. Datenschutzrechtliche Anforderungen an das Scientific Use File der IEB	21
4. Faktische Anonymisierung der IEB	23
4.1. <i>Besonderheiten bei Beschäftigungs-, Leistungsempfangs- und Arbeitsuchezeiten</i>	24
4.2. <i>Besonderheiten Maßnahmeteilnahmezeiten</i>	24
4.3. <i>Umfang und Art der Datenveränderung</i>	25
5. Analysepotential des Scientific Use Files der IEB	28
6. Beschreibung der Variablen des Scientific Use Files der IEB	29
6.1. <i>Identifikationsnummern</i>	29
6.1.1. Personennummer („persnr“).....	29
6.1.2. Betriebsnummernzähler („betnrz“).....	29
6.2. <i>Spellanfang und –ende</i>	30
6.2.1. Beginn- und Enddatum des Originalspells („begorig“ und „endorig“).....	30
6.2.2. Beginn- und Enddatum einer Episode („begepi“ und „endeipi“).....	30
6.3. <i>Generierte technische Merkmale</i>	31
6.3.1. Quelle der Information („quelle“).....	31
6.3.2. Kombination der Quellen („kom_quelle“).....	31
6.3.3. Spellzähler und Anzahl der Spells pro Konto („spell“ und „nspell“).....	31
6.3.4. Spellzähler und Anzahl der Spells pro Episode („level2“ und „nlevel“).....	32
6.3.5. Spellzähler und Anzahl der Spells pro Episode und Quelle („level1“ und „nlevel1“).....	32
6.3.6. Bereinigungskennzeichen („berknz“).....	32
6.3.7. Status des Endedatums („stendat“).....	33
6.3.8. Kennzeichen des BewA-Spells („ken_asu“).....	33
6.4. <i>Personenstatus vor, während und nach dem aktuellen Spell</i>	34
6.4.1. Erwerbsstatus vor Arbeitsuche („estatvor“).....	34
6.4.2. Fortsetzung der Arbeitslosigkeit nach Arbeitsunfähigkeit („krankvor“).....	34
6.4.3. Erwerbsstatus („erwstat“).....	34
6.4.4. Grund des Spellendes („grund“).....	37
6.5. <i>Persönliche Merkmale</i>	41
6.5.1. Geburtsjahr („gebjahr“).....	41
6.5.2. Geschlecht („sex“).....	42
6.5.3. Staatsangehörigkeit („nation“).....	42
6.5.4. Schwerbehindertenstatus („schweb“).....	43
6.5.5. Schulabschluss („schbild“).....	43
6.5.6. Ausbildung („bild“).....	43

6.5.7.	Familienstand („famstand“)	44
6.5.8.	Geburtsjahr, jüngstes Kind („gebjahr_k“)	45
6.5.9.	Kinderzahl im Haushalt („kind“)	46
6.6.	<i>Angaben zu Beschäftigungsverhältnis und Arbeitsuche</i>	46
6.6.1.	Stellung im Beruf und Arbeitszeit („stib“)	46
6.6.2.	Berufsgruppe („beruf“)	47
6.6.3.	Tagesentgelt / täglicher Leistungssatz („tentgelt“)	47
6.6.4.	Wirtschaftszweig 93-Klassifizierung („w93“)	48
6.6.5.	Dauer der Arbeitslosigkeit („dualo“)	49
6.6.6.	Geplantes Endedatum („endplan“)	50
6.6.7.	Berufserfahrung („erfahrung“)	51
6.6.8.	Qualifikationsstufe („quali“)	51
6.6.9.	Mobilität („mobil“)	52
6.6.10.	Gesundheitliche Einschränkungen („gesund_ein“)	53
6.6.11.	Gesundheitliche Einschränkungen – Auswirkung auf Vermittlung („auswirkung“)	53
6.6.12.	Arbeitszeit des Stellengesuchs („arbeit“)	54
6.6.13.	Restanspruchsdauer Arbeitslosengeld („restanspruch“)	54
6.7.	<i>Ortsangaben</i>	55
6.7.1.	Wohnort – Regionaltyp der Arbeitsagentur („wo_aatyp“)	55
6.7.2.	Wohnort – Ost-West („wo_ost_west“)	56

1. Einleitung

Die Integrierten Erwerbsbiografien (IEB) sind ein Datenangebot des IAB (Institut für Arbeitsmarkt- und Berufsforschung) und werden seit 2004 erzeugt. Sie umfassen vier verschiedene Quellen und sind damit die umfangreichste Datenmenge zu erwerbshistorischen Meldungen auf Personenebene, die vom Forschungsdatenzentrum der BA am IAB (FDZ) angeboten werden. Ausgangspunkt der IEB sind Meldungen, die die Rentenversicherungsträger zu Beschäftigtenzeiten erfassen und an die Bundesagentur für Arbeit (BA) weiterleiten sowie Informationen, die im Rahmen der Arbeitsmarktförderung und der Arbeitslosenversicherung durch die Bundesagentur für Arbeit erfasst werden. Alle Meldungen liegen in Kontenform vor und sind tagesgenaue Angaben zu Erwerbszeiten. Ergänzt werden diese Informationen mit Personenmerkmalen als auch mit Leistungs- und Lohnangaben, die für den Zeitraum der jeweiligen Erwerbsmeldung gültig sind.

Der vorliegende Datenreport beschreibt eine anonymisierte Stichprobe zu diesen Integrierten Erwerbsbiografien, die der Wissenschaft- und Forschung in Form eines Scientific Use Files (IEBS-SUF) zur Verfügung gestellt wird. Hierzu dokumentiert der Report die Arbeitsschritte, die im Rahmen zur Erstellung des Scientific Use Files umgesetzt worden sind und liefert eine Dokumentation der enthaltenen Variablen. Im Mittelpunkt des Papiers steht die Gegenüberstellung der Informationen aus dem IEBS-SUF V1 und aus der IEBS V1.0 (Stichprobe der IEB). Letzter Datensatz weicht von dem IEBS-SUF V1 durch ein deutlich geringeres Ausmaß an Anonymisierung ab und steht der Forschung lediglich im Rahmen von Gastaufenthalten und Datenfernverarbeitung über das FDZ der Bundesagentur für Arbeit (im IAB) zur Verfügung. Beide Datensätze basieren auf der gleichen Ausgangsversion der Integrierten Erwerbsbiografien (Version 3.01).

Im Gegensatz zur IEBS V 1.0 umfasst der IEBS-SUF V1 ein erweitertes Variablenspektrum. Damit möchte das IAB einen für die Forschung attraktiven Datensatz zur Verfügung stellen, auf dessen Grundlage sich vor allem die Untersuchungen im Rahmen der Evaluation der Maßnahmen zur Umsetzung der Vorschläge der Hartz-Kommission (Hartz-Evaluation I-III) nachvollziehen lassen.¹ Maßgebliches Kriterium für die Erstellung eines Scientific Use Files sind jedoch die durch Gesetzgeber vorgegebenen Bestimmungen zum Datenschutz. Diese stellen sicher, dass Persönlichkeitsrechte gewahrt bleiben und regeln, dass in den zur Verfügung gestellten Daten nicht oder nur sehr schwierig individuelle Fälle identifiziert werden können bzw. dass die Informationen in den Daten „missbräuchlich“ verwendet werden. Dies

¹ Der Auftrag zur Bearbeitung der Machbarkeitsstudie erfolgte im Rahmen der Hartz-Evaluation I-III vom Bundesministerium für Wirtschaft und Arbeit (BMWA) an das IAB (Institut für Arbeitsmarkt- und Berufsforschung), Abteilung Informations- und Technik Management: IZM).

führt dazu, dass das Ausgangsmaterial vor einer Veröffentlichung „faktisch anonymisiert“ werden muss.

Im Kern wird bei der vorliegenden Anonymisierung der IEB von einem bereits vorhandenen Stichprobenauszug zu den Integrierten Erwerbsbiografien – der IEBS – ausgegangen. Diese wird um bestimmte Variablen (insbesondere Regionalidentifikatoren) reduziert. Darüber hinaus werden je nach Quelle der IEB unterschiedliche Anonymisierungen vorgenommen. Das Einhalten der gesetzlichen Vorgaben zur Umsetzung der faktischen Anonymisierung wird dann auf Grundlage von Heuristiken bzgl. der repräsentierten Grundgesamtheit umgesetzt. Konkret bedeutet dies, dass für das Kriterium der faktischen Anonymität eine minimale Fallzahl univariater Häufigkeitsauszählungen verwendet wird. Liegen die Fallzahlen unterhalb bestimmter Werte, dient dies als Entscheidungskriterium für eine Anonymisierung des Datenmaterials.

Die nachfolgende Dokumentation beschreibt hierzu zunächst das Ausgangsmaterial für diese Anonymisierung und skizziert die angewandten Anonymisierungskriterien sowie das Vorgehen bei der Datenanonymisierung. Darüber hinaus wird dargestellt, in welchem Umfang Informationen durch die Anonymisierung – gegenüber der IEBS – verloren gehen.

2. Aufbau des Scientific Use Files der Integrierten Erwerbsbiografien (IEB)

Die Integrierten Erwerbsbiografien sind das Ausgangsmaterial des IEBS-SUF. Der nachfolgende Abschnitt gibt einen Überblick über die Inhalte, den Aufbau und die Datenqualität dieses Datenmaterials.

2.1. Inhalt und Umfang der IEB

Die IEB sind ein auf Prozessdaten der Bundesagentur für Arbeit und der Rentenversicherungsträger basierender Datensatz der vier verschiedene Quellen zum Erwerbsverhalten integriert. Diese umfassen Beschäftigtenmeldungen, Zeiten des Leistungsempfangs (Arbeitslosenversicherung und Arbeitslosenhilfe) sowie Zeiten erfasster Arbeitsuche und Teilnahmen an Maßnahmen der Aktiven Arbeitsmarktförderung. Das Besondere an den Integrierten Erwerbsbiografien ist, dass die Meldungen aus den verschiedenen Quellen auf Personenebene chronologisch aufbereitet sind.

Die IEB umfassen folgende Quellen:

a) Die Beschäftigtenhistorik (BEH)

In dieser sind die gesetzlich vorgeschriebenen Beschäftigtenmeldungen durch die Betriebe an die Rentenversicherungsträger zu sozialversicherungspflichtigen Beschäftigungszeiten von Arbeitnehmern enthalten. Mit der Änderung des Meldeverfahrens 1999 werden auch Meldungen zu geringfügig Beschäftigten und Praktikanten o.ä. bei der Erfassung von Beschäftigtenmeldungen berücksichtigt. Die Datenquelle umfasst Informationen zu Beschäftigungszeiten, zum Einkommen und zur Art der Beschäftigung sowie zur Qualifikation des Arbeitnehmers und wird seit 1990 in die IEB integriert. Grundlage der Integration von Beschäftigtenmeldungen ist eine Datenübermittlung durch die Rentenversicherungsträger an die Bundesagentur für Arbeit.

b) Die Leistungsempfängerhistorik (LEH)

Die Leistungsempfängerhistorik basiert auf Daten der Bundesagentur für Arbeit und weist Informationen (Dauer, Art und Höhe der Leistung) zum Empfang von Arbeitslosengeld, Arbeitslosenhilfe und Unterhaltsgeld aus. Basis dieser Informationen sind Leistungsempfangszeiten, die über das Fachverfahren CoLei (Computergestützte Leis-

tungsgewährung) verwaltet werden. In den LeH-Teil der IEB gehen Informationen über alle Leistungsempfangszeiten ein, die am oder nach dem 01.01.1990 enden.

c) Die Maßnahmeteilnehnehistorik (MTH)

Die Maßnahmeteilnehnehistorik enthält ebenfalls Informationen von der Bundesagentur für Arbeit. Ausgangspunkt sind die Meldungen zu Teilnahmen an Leistungen der aktiven Arbeitsmarktpolitik im Rahmen des SGB III, die über das Fachverfahren CoSach (Computergestützte Sachverwaltung) zur Verfügung stehen. Die Maßnahmeteilnehnehistorik umfasst individualisierte Meldungen zu erhaltenen Förderleistungen (Art der Maßnahme). Dies schließt die Maßnahmeteilnehnehedauern als auch ergänzende persönliche Merkmale ein sowie weitere Informationen zur Maßnahme. Neben der lokalen Verwaltung von Förderleistungen werden durch das Fachverfahren auch statistische Zwecke bedient, in dem die Förderleistungen auf Personenebene regelmäßig zentral zugespielt und aufbereitet werden. Diese sind Ausgangspunkt der in die IEB integrierten Maßnahmeteilnehmungen.

d) Die Arbeitsuchendehistorik (BewA)

Die Arbeitsuchendehistorik enthält ebenfalls Meldungen von Arbeitsuchenden, die bei der Bundesagentur für Arbeit erfasst werden. Grundlage sind Eingaben zum Bewerberangebot, die bei Meldungen zur Arbeitssuche von einer Person durch die Arbeitsagentur erhoben werden. Die BewA-Informationen zu Personen sind Grundlage für die Vermittlungsleistungen und den Leistungen der Arbeitslosenversicherung (Lohnersatzzahlungen). Enthalten sind hier u.a. Informationen zur Qualifikation des Bewerbers, zur Art der gewünschten Arbeitssuche als auch sonstige vermittlungsrelevante Informationen wie Einschränkungen der „Arbeitsfähigkeit“ (z.B. gesundheitliche Einschränkungen oder eingeschränkte Mobilität). Grundsätzlich kann davon ausgegangen werden, dass die BewA-Informationen zur Qualifikation und zur Person von größerer Zuverlässigkeit sind als die Informationen aus anderen Verfahren (vgl. Fitzenberger/ Osikominu/ Völter, 2005).

2.1.1. Verknüpfung und Aufbau

Die Datenquellen der IEB kommen aus unterschiedlichen und unabhängig geführten Fachverfahren / EDV-Systemen. Zur Verknüpfung der unterschiedlichen Erwerbsverlaufsinformationen werden beim Aufbau der IEB zwei zentrale Identifikatoren verwendet (die BA-Kundennummer und die Sozialversicherungsnummer). Durch diese lassen sich die separaten Informationen auf Personenebene zu Erwerbsverlaufskonten zusammenführen. Zu beachten ist, dass zwar die BeH-Meldungen meist ohne BA-Kundennummer erfasst sind, dass jedoch für die BA-Meldungen nahezu vollständig eine Versicherungsnummer im DataWareHouse geführt wird.

Da angenommen werden kann, dass sich die Versicherungsnummer bei einer Person nur selten ändert, wird diese als zentraler Personenidentifikator verwendet. Durch dieses Vorgehen wird erreicht, dass: 1. eine Zuordnung der BA-Beschäftigtenmeldungen möglich ist und, dass 2. Personen mit mehr als einer BA-Kundennummer über die Relation Kundennummer-Versicherungsnummer eindeutig einer Versicherungsnummer zugeordnet werden können. Der Füllgrad und die Qualität der Relation sind daher entscheidend für die Vollständigkeit der Meldungen auf Individualebene.

Integriert werden daher nur solche Personen, bei denen eine eindeutige Relation zwischen Kundennummern (Personenidentifikator der Bundesagentur für Arbeit) und der Sozialversicherungsnummer (Personenidentifikator der Rentenversicherungsträger) möglich ist. Validiert wird die Eindeutigkeit durch die Überprüfung des Geburtsdatums und des Geschlechts.

2.1.2. Datenqualität

Beim Umgang mit den Integrierten Erwerbsbiografien sind unterschiedliche Sachverhalte zu berücksichtigen, die mit dem Aufbau der Datenbasis aus den verschiedenen Quellen zusammenhängen. Dies betrifft zum Beispiel Probleme zur Identifikation von Personen, potentielle Inkonsistenzen von Informationen, nicht-plausible Überschneidungen von Meldungen usw. Darüber hinaus existieren jedoch auch quellenspezifische Probleme, die mit möglichen Systemwechseln in der Art der Datenerfassung zusammenhängen. Einige zentrale und bekannte Besonderheiten im Umgang mit der Qualität der Integrierten Erwerbsbiografien werden nachfolgend kurz vorgestellt.

Datenaufbereitung: Der Aufbau der IEB erfolgt in mehreren Schritten. Zunächst werden nur Meldungen mit gültigen Relationen zwischen Kundennummer und Versicherungsnummer integriert. Für diese wird anschließend ein neuer eindeutiger Personenidentifikator aufgebaut, der vorrangig über die Versicherungsnummer definiert ist. Die Versicherungsnummer wird deshalb verwendet, da sie mit höherer Sicherheit personeneindeutige Zuordnungen ermöglicht. Anschließend werden Altersbereinigungen durchgeführt, so dass Sätze gelöscht werden, bei denen zum Spellbeginn das Alter kleiner 13 oder größer 75 beträgt. Darüber hinaus bleiben Meldungen unberücksichtigt, bei denen das Beginndatum größer ist als das Enddatum. Ausgeklammert werden zudem alle Konten, die Sätze mit variierendem Geschlecht oder Geburtsdatum enthalten.

Vollständigkeit: Im Prinzip umfassen die IEB in ihren Meldungsarten (Beschäftigung, Leistungsempfang, Maßnahmeteilnahmen, Arbeitsuche) vollständige Grundgesamtheiten, so dass alle real vorkommenden Meldungen in der IEB erfasst werden. Technisch gesehen ist dieses

jedoch nur bedingt der Fall. So werden bestimmte Meldungen erst mit Verzögerungen in die IEB integriert.² Mit zunehmender Entfernung der Datenlage vom aktuellen Rand dürften die IEB daher ein zunehmend vollständigeres Abbild der Grundgesamtheit enthalten. Weitere Abweichungen können aber durch den spezifischen Aufbau der IEB entstehen, bei dem nur solche Fälle in die IEB integriert werden, die in gültigen Relationen zwischen zentralen Personenidentifikatoren vorliegen. Im Zweifel weisen die IEB damit regelmäßig geringere Fallzahlen aus als die Statistiken der Bundesagentur für Arbeit.

Konsistenz der Merkmale aus unterschiedlichen Quellen: Außerdem gilt es bei den Integrierten Erwerbsbiografien zu berücksichtigen, dass jeweils für unterschiedliche Quellen ähnliche Informationen zum selben Sachverhalt vorliegen. So sind z.B. in der MTH als auch in den BewA-Meldungen Merkmale zur Qualifikation enthalten. Diese Informationen sind nicht notwendigerweise widerspruchsfrei und müssen inhaltlich je nach Fragestellung vom Forscher selbst mit Regeln berücksichtigt werden.

Widerspruchsfreiheit von Meldungen aus unterschiedlichen Quellen: Ebenfalls sollte beim Umgang mit den Integrierten Erwerbsbiografien berücksichtigt werden, dass sich Meldungen zu einer Person aus den verschiedenen Quellen überschneiden können. Üblicherweise ist dies zum Beispiel der Fall bei Perioden des Leistungsbezugs zu denen i.d.R. parallel-liegende Spells zur Arbeitsuche existieren. Hier gilt ebenfalls, dass derartige Überschneidungen ggf. inhaltlich auf Plausibilitäten geprüft werden müssen. Zu beachten ist dabei, dass sowohl die Arbeitsuchendenzeiten als auch die MTH-Meldungen jeweils von den Sachbearbeitern per Hand in den Arbeitsagenturen eingegeben werden. Eine Übersicht zu den potentiellen Überschneidungen kann Jaenichen et al. (2005) entnommen werden (http://doku.iab.de/fdz/reporte/2005/DR_4.pdf).

2.1.3. Besonderheiten zur Qualität der MTH-und BewA-Meldungen

Bezüglich der Qualität der Meldungen einzelner Quellen gilt es beim Umgang mit den Integrierten Erwerbsbiografien einige Besonderheiten zu berücksichtigen. Dies gilt insbesondere für die Meldungen zu Maßnahmeteilnahmezeiten (MTH) und den Arbeitsuchendenmeldungen (BewA), da diese Quellen relativ jungen Ursprungs sind und erst ab 2000 über ein Fachverfahren (Co-Sach) zur Verfügung stehen, das verknüpfbare Identifikatoren auf Personenebene berücksichtigt.

² Sowohl bei der Bundesagentur für Arbeit als auch beim Meldeverfahren durch die Rentenversicherungsträger können Änderungen noch zum Teil Jahre nach der Ersterfassung erfolgen.

MTH³

Bei den MTH-Meldungen ist zunächst zu berücksichtigen, dass die Beginn- und Enddaten vom Fachpersonal eingegeben werden und dass bei der Eingabe keine Plausibilitätsprüfungen vorgenommen werden. Dies kann im verstärkten Maße zu widersprüchlichen Überschneidungen von Meldungen aus anderen Quellen führen.

Darüber hinaus gilt es bei der Qualität der integrierten MTH-Meldungen zu beachten, dass die Ausgangsdaten zu den Maßnahmeteilnahmen erst durch technische Aufbereitungsroutinen erforderlich sind. Diese begründen sich damit, dass im Ausgangsmaterial systembedingt zu jeder Teilnahme zunächst mehrere Meldungen vorliegen. Durch das Meldeverfahren werden automatisch Meldungen zum Zeitpunkt des Zugangs (Teilnahmebeginn) und Meldungen zum Zeitpunkt des Abgangs (Teilnahmeende) erzeugt (vgl. Kruppe/Oertel, 2003). Zusätzlich werden zu jeder Teilnahme noch Bestandsmeldungen geliefert, die jeweils monatlich solange geliefert werden, wie sich eine Person in einer Maßnahme befindet. Dadurch existieren im Ausgangsmaterial der MTH regelmäßig Meldungen zum Zugang, zum Bestand und zum Abgang. Hintergrund dieser Art der Datenlieferung sind die statistischen Verarbeitungszwecke der Bundesagentur für Arbeit, die regelmäßig aggregierte Zahlen für Zugänge, Bestände und Abgänge in/aus Maßnahmen ausweist.

Diese Mehrzahl an Meldungen wird so aufbereitet, dass vorrangig die Abgangsmeldungen für die Integration in die IEB berücksichtigt werden. Dieses Vorgehen hat den Vorteil, dass mit einer höheren Wahrscheinlichkeit nur solche Teilnahmen berücksichtigt werden, denen tatsächliche Teilnahmen zugrunde liegen. Darüber hinaus können mögliche zeitliche Verschiebungen in den Teilnahmen berücksichtigt werden, die nach dem angesetzten Teilnahmebeginn liegen. Weiterhin werden nur solche Maßnahmeteilnahmen aufgenommen, wenn für die Abgangssätze auch Zugangssätze existieren. Für die Anfangszeit beim Aufbau der MTH (2000) führt dies zu einer systematischen Unterschätzung der Maßnahmeteilnahmen (sowohl bei den Zugängen als auch bei den Beständen), da ggf. einigen Maßnahmeteilnahmen keine Zugangssätze zugeordnet werden können.

BewA

Die BewA-Daten wurden 1997 mit Bestandsdaten zu den Arbeitsuchendenmeldungen und den zugehörigen Merkmalen zum Bewerberangebot aufgebaut. Seit dieser Uraldung hat es mehrere Umstellungen zentraler Indentifikatoren bei der Bundesagentur für Arbeit gegeben, die Auswirkungen auf die Vollständigkeit der integrierten BewA-Meldungen haben können. Tech-

³ Die nachfolgenden Ausführungen zur MTH beziehen sich auf die Art der Datenlieferung, die bis März 2005 gilt. Mit März 2005 hat es eine flächenweite Systemumstellung gegeben, die eine andere Art der Datenversorgung beinhaltet.

nisch haben diese Systemumstellungen Auswirkungen wie und ob BewA-Meldungen zu einer Person zugeordnet werden können. Von Bedeutung ist vor allem, dass ein systemübergreifend genutzter Personenidentifikator erst seit 2000 existiert. Dieser löst die BewA-Eingabenummer als zentralen Identifikator ab.

Die Kundennummer wird allerdings erst ab 2003 flächendeckend zentral kontrolliert (Einführung einer zentralen PersonenDatenVerarbeitung). Zwischen 2000 und 2003 wurde die Vergabe von Kundennummern nur lokal gesteuert. Dies kann dazu führen, dass eine Person, die vor 2003 mit BewA-Meldungen erfasst wurde, unterschiedliche Identifikatoren aufweist. Soweit möglich berücksichtigt die IEB derartige Konstellationen, indem nur Personen integriert werden, für die eine Zuordnung zwischen Versicherungsnummer und BA-Kundennummer (bei älteren Sätzen die BewA-Eingabenummer) möglich ist. Dadurch lässt sich weitestgehend sicherstellen, dass ab 2000 eine eindeutige Personenidentifikation bei den BewA-Meldungen möglich ist.

Eindeutige Zuordnungen von Meldungen für den Zeitraum vor 2000 können nachträglich jedoch durch bestimmte Relationen in den Identifikatoren erzeugt werden. Dieses Vorgehen hat jedoch zur Folge, dass bestimmte BewA-Sätze beim Kontenaufbau einer Person unberücksichtigt bleiben.⁴ Die nachstehende Illustration veranschaulicht dieses Problem. Danach können bei zwei von vier möglichen Kontenaufbauten Konstellationen existieren, bei denen die BewA-Meldungen in den IEB-Konten nicht vollständig enthalten sind. Grundsätzlich kann davon ausgegangen werden, dass die BewA-Meldungen einer Person mit hoher Sicherheit vollständig sein dürften, wenn das Konto einer Person BewA-Meldungen aufweist, die den Zeitraum zwischen 01/2000 und 06/2003 berühren:

⁴ Zwei mögliche Fehlertypen könnten hierbei eine Rolle spielen: Zum einen werden Maßnahmeteilnehmer oder Arbeitssuchende ohne Erwerbshistorie nicht berücksichtigt, da für diese ggf. eine Zuordnung von Versichertennummer und BewA-MTH-Identifikator weniger wahrscheinlich ist, zum anderen könnten die Teilnehmer / Arbeitssuchendenmeldungen überschätzt werden, wenn zu einem Teilnehmer / Arbeitssuchenden mehrere BewA-MTH-Identifikatoren existieren.

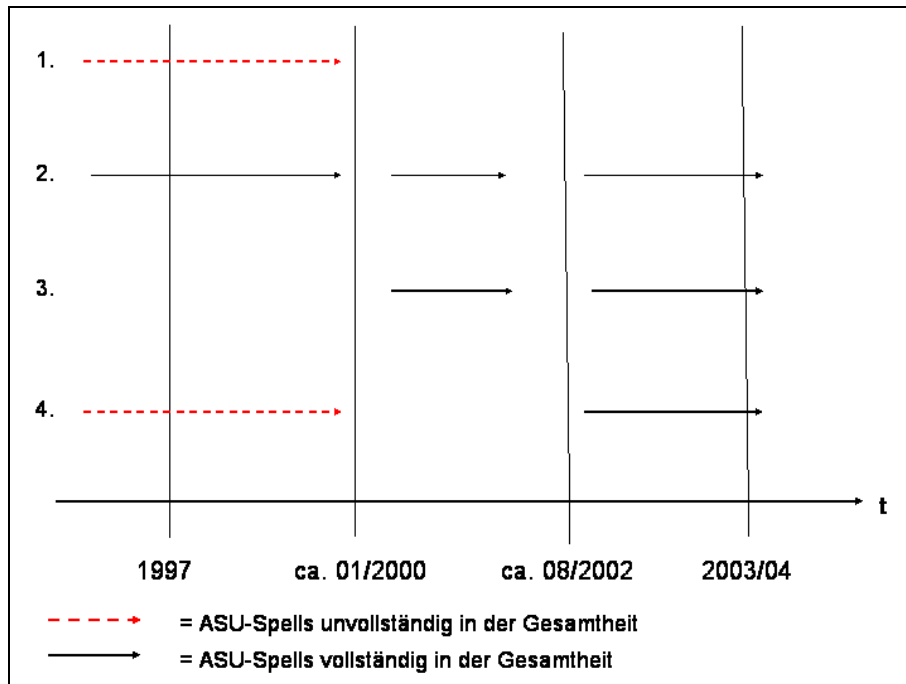


Abbildung 1: BewA-Kontenaufbauten mit unterschiedlichem Vollständigkeitsgrad

1. Konten, die ausschließlich BewA-Spells vor 2000 enthalten, sind hinsichtlich der BewA-Spells unvollständig (ca. 1% aller Konten mit BewA-Spells). Meldungen in diesem Zeitraum sind nur über die BewA-Eingabenummer erfasst.
2. Konten, die (zusätzlich zu BewA-Zeiten vor 2000) BewA-Spells beinhalten, die in den Zeitraum 01/2000 – 06/2003 hineinragen (oder darin liegen), sind vollständig hinsichtlich der BewA-Spells (ca. 1/3 aller Konten mit BewA-Spells). Für diese Meldungen existieren sowohl BewA-Eingabenummern als auch BA-Kundennummern.
3. Konten, die ausschließlich BewA-Spells beinhalten, die nach 01/2000 beginnen, sind hinsichtlich der BewA-Spells ebenso vollständig (ca. 2/3 aller Konten mit BewA-Spells). Für diese Meldungen existieren sowohl BewA-Eingabenummern als auch BA-Kundennummern.
4. Konten, die BewA-Spells vor 2000 und nach 06/2003 enthalten, jedoch keine Spells im Zeitraum 01/2000 – 06/2006, sind für den Zeitraum vor 2000 nicht vollständig (weniger als 1% aller Konten mit BewA-Spells). Für Konten, die nur BewA-Spells nach 6/2006 (nur die Kundennummer als Identifikator) enthalten, ist nicht bekannt, ob vor 2000 BewA-Spells (auf Basis von BewA-Eingabenummern erfasst) vorlagen (ca. 30% aller Konten mit BewA-Spells).

2.1.4. Veränderung von Merkmalsausprägungen (Zeitkonsistenz der Merkmale)

Für die Qualität der IEB ebenfalls von Bedeutung sind Veränderungen von Merkmalsausprägungen, die aufgrund von Systemwechsel bei der Erfassung von Merkmalen in den jeweiligen Meldeverfahren vorgenommen wurden. So ist zu berücksichtigen, dass die IEB jeweils den aktuellsten Differenzierungsgrad aufweisen. Dies kann jedoch dazu führen, dass keine Meldungen für ältere Zeiträume – die ohne die aktuelle Differenzierung erfasst wurden – ausgewiesen werden.

Veranschaulicht sei dieser Aspekt im Folgenden am Beispiel von Stichtagsauszählungen zur Personengruppe „Praktikanten“. Die nachfolgende Tabelle enthält hierzu die Mengengerüste der Jahre 1996 bis 2001. Dabei zeigt sich, dass die Fallzahlen von 1998 auf 1999 einen erheblichen Sprung machen.

	15.06. des Jahres ...						
	1996	1997	1998	1999	2000	2001	...
Praktikanten (n)	4	7	91	1069	1351	1291	...

Abbildung 2: Stichtagsauszählungen der Personengruppe Praktikanten

Ursache für diesen Sprung ist allerdings keine Änderung in dem tatsächlichen Aufkommen an Praktikantenstellen, sondern eine Änderung im Meldeverfahren. Erst ab 1999 unterscheidet das Verfahren zwischen verschiedenen Personengruppen, die gemeldet werden, bei denen unter anderem auch Praktikanten separat erfasst und ausgewiesen werden. Im Meldeverfahren vor 1999 war diese Differenzierung noch nicht enthalten, so dass u.a. Praktikanten vor 1999 nicht gesondert ausgewiesen wurden. Die gezählte Häufigkeit von bspw. 4 Praktikanten zum Stichtag 30.06.1996 lässt sich dadurch erklären, dass rückwirkende Beschäftigtenmeldungen berücksichtigt wurden und dabei den aktuelleren Kodierungen entsprechen. Im Abgleich mit der Anzahl an Praktikanten, die seit 1999 ausgewiesen wird, ist daher davon auszugehen, dass zu den Stichtagen vor 1999 regelmäßig ca. 1000-1200 Praktikanten in der Stichprobe enthalten sind, ohne als solche gekennzeichnet zu sein.

2.2. Genese des IEBS-SUF V1

Technisch betrachtet stellt der SUF eine Zufallsstichprobe aus den Integrierten Erwerbsbiografien dar (ca. 2,19% aller Personen der IEB). Im Grunde entspricht der Datensatz damit in seiner Struktur und in den Merkmalen einer anonymisierten IEBS V.1.0.

Die Genese des IEBS-SUF V1 lässt sich in fünf unterschiedliche Teilschritte zerlegen:

- 1) Stichprobenziehung
- 2) Exklusion von Meldungen vor dem 01.01.1992
- 3) Stichtagsabgrenzungen/ Auszählungen
- 4) Anonymisierungen
- 5) Episodensplitting und Neuberechnung der Alo-Dauer

Ausgangspunkt des IEBS-SUF V1 ist die IEB in der Version 3 (IEB V.3).⁵ Aus dieser wird eine Stichprobe gezogen, bei der alle Spells jener Personen selektiert werden, die an acht festgelegten Kalendertagen (Tag und Monat) Geburtstag haben. Dieses Vorgehen ist aus zwei Gesichtspunkten günstig:

- Zukünftige Aktualisierungen des Scientific Use Files der IEB
Aus Gründen der Vergleichbarkeit ist es wünschenswert, in einer zukünftigen Version des Scientific Use Files der IEB den Erwerbsverlauf jener Personen aktualisieren zu können, die auch in der Vorgängerversion enthalten waren. Über die acht festgelegten Geburtstage können diese Personen unaufwendig wieder gezogen und mit ihren aktualisierten Erwerbsbiographien in eine neue Scientific Use File-Version aufgenommen werden. Ebenso werden dabei Personen gezogen, die zwar den gleichen Geburtstag aufweisen, aber in den IEB V.3 nicht enthalten waren.
- Synchronität des Scientific Use Files der IEB mit der IEBS V.1.0
Die in der IEBS V.1.0 enthaltenen Personen wurden auch über eine Geburtstagsstichprobe aus den IEB V.3 mit den gleichen Kalendertagen wie in der Machbarkeitsstudie Scientific Use File der IEB gezogen. Eine einheitliche Festlegung von Auswahlgrundlage und Kalendertagen in IEBS V.1.0 und Scientific Use File der IEB stellt die gleichartige Struktur in den beiden Datenbasen sicher, die für die Umsetzung des 2-stufigen Analysemodells von großer Relevanz ist.

In einem zweiten Schritt werden alle Meldungen ausgeklammert, die vor dem 01.01.1992 beginnen. Dieser Ausschluss von Meldungen berücksichtigt die Tatsache, dass für den davor

⁵ Für weitere Informationen zu den IEB V.3 vgl. http://doku.iab.de/fdz/reporte/2005/DR_6.pdf

liegenden Zeitraum systematisch Meldungen zu Beschäftigtenzeiten und Leistungsbezügen aus Ostdeutschland unterrepräsentiert sind.

In einem dritten Aufbereitungsschritt werden unterschiedliche Stichtagsabgrenzungen und -auszählungen vorgenommen. Diese sind Grundlage für die nachfolgenden Anonymisierungen. Wie oben bereits beschrieben werden univariate Häufigkeiten als Grundlage für die repräsentierte Menge in der Grundgesamtheit verwendet. Das Unterschreiten bestimmter Fallzahlen wird dann als Kriterium für eine erforderliche Anonymisierung verwendet. Das genaue Vorgehen bei der Anonymisierung ist weiter unten beschrieben.

In Schritt vier (siehe weiter unten) werden dann schließlich unterschiedliche Anonymisierungsverfahren für die Daten angewendet. Abschließend (Schritt fünf) wird ein Episodensplitting umgesetzt, bei dem die Spells der angereicherten und anonymisierten Stichprobe so gesplittet werden, dass exakt parallel-liegende und überlappungsfreie Episoden entstehen.⁶ Alle technischen Merkmale werden erst nach diesem Episodensplitting erzeugt.

Die nachfolgende Abbildung illustriert nochmals das Grundgerüst des IEBS-SUF V1.

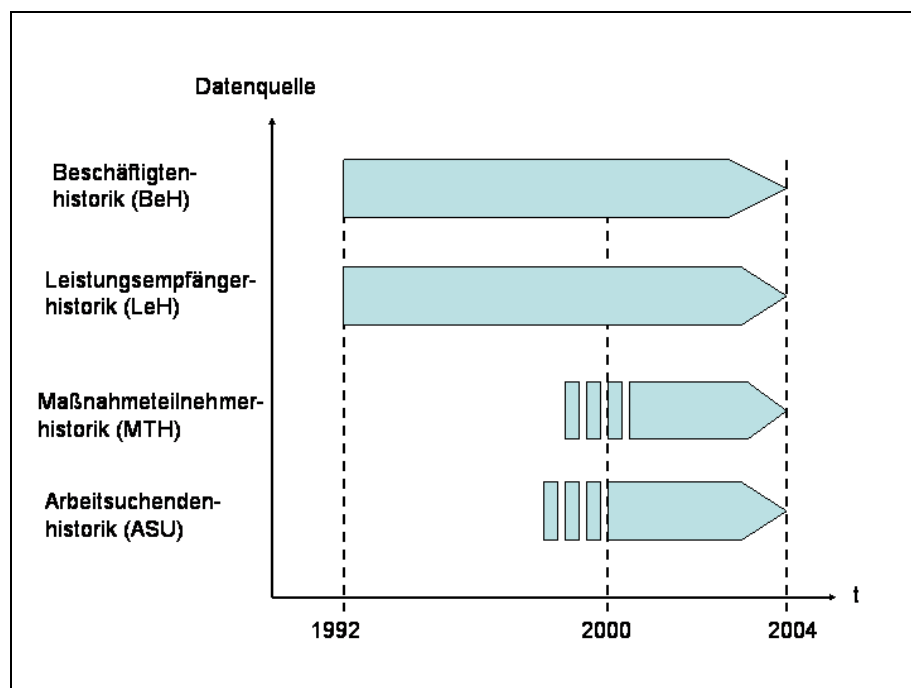


Abbildung 3: Datenquellen und abgebildete Zeiträume des Scientific Use Files der IEB

⁶ Eine genauere Beschreibung des Episodensplittings kann der IEBS-Dokumentation entnommen werden: http://doku.iab.de/fdz/reporte/2005/DR_6.pdf

2.2.1. Variablenumfang des Scientific Use Files der IEB

Das Merkmalspektrum für den Scientific Use File der IEB lehnt sich an die Merkmale an, die im Rahmen der Hartz-Evaluation I-III vom IAB an die evaluierenden Institute geliefert wurden. Hierzu wurden Ergebnisse einer Befragung verwendet, die bei den evaluierenden Instituten durchgeführt wurde.⁷ Die Ergebnisse dieser Daten-Evaluation gehen in die Erstellung des IEBS-SUF V1 ein, so dass dieser ein, aus der Sicht der evaluierenden Institute, inhaltlich validiertes Variablenspektrum enthält.

Im Vergleich zur IEBS V.1.0 ist der IEBS-SUF V1 um folgende Variablen angereichert:

- a) BewA Kennzeichen
- b) Familienstand
- c) Geburtsjahr des jüngsten Kindes
- d) Kinderanzahl im Haushalt
- e) Erfahrung
- f) Qualifikation
- g) Mobilität
- h) Gesundheitliche Einschränkungen
- i) Gesundheitliche Einschränkungen mit Vermittlungsrelevanz
- j) Gewünschte Arbeitszeit im Stellengesuch
- k) Restanspruchsdauer beim Arbeitslosengeldbezug
- l) Wohnort Ost-West

Der IEBS-SUF V1 ist allerdings gegenüber der IEBS V.1.0 auch um einige Variablen reduziert. Diese beziehen sich insbesondere auf die regionalen Identifikatoren, die aufgrund von Anonymisierungsaspekten in dem IEBS-SUF V1 nicht enthalten sind. Aus dem Datenumfang entfernt

⁷ Dabei wurden sie gebeten, im Fragebogen anzugeben, welches Merkmal zu welchem Analysezweck verwendet wurde (Matching, erklärende Variable, Ergebnisvariable, etc.).

sind aber auch weitere Variablen, die aus inhaltlichen Gründen nicht aufgenommen wurden.⁸ Tabelle 1 stellt die Merkmale aus dem IEBS-SUF V1 und aus der IEBS V.1.0 einander gegenüber. Alle Variablen, die zur IEBS V.1.0 abweichen, sind entsprechend im Variablennamen markiert. Das Kürzel „_a“ steht für eine vorgenommene Anonymisierung.

Variablenname	Variablenname in IEBS V1.0	Bezeichnung	BeH	LEH	MTH	BewA
Identifikationsnummern						
persnr	persnr	Personennummer				
	satznr	Satznummer				
	masnr	Maßnahmenummer				
betnrz	(betnr)	Betriebsnummer				
Spellanfang und -ende						
begorig	begorig	Beginndatum Originalspell				
endorig	endorig	Endedatum Originalspell				
begepi	begepi	Beginndatum der gesplitteten Episode				
endepi	endepi	Endedatum der gesplitteten Episode				
Generierte technische Merkmale						
quelle	quelle	Quelle				
kom_quel	kom_quel	Kombination der Quellen				
spell	spell	Spellzähler pro Konto				
nspell	nspell	Anzahl der Spells pro Konto				
level2	level2	Spellzähler pro Episode				
nlevel2	nlevel2	Anzahl der Spells pro Episode				
level1	level1	Spellzähler pro Episode und Quelle				
nlevel1	nlevel1	Anzahl der Spells pro Episode und Quelle				
berknz	berknz	Bereinigungskennzeichen				
stendat	stendat	Status des Ende-Datums				
knz_asu		BewA-Kennzeichen				
Personenstatus vor, während und nach dem aktuellen Spell						
estatvor	estatvor	Erwerbsstatus vor Arbeitsuche				
krankvor	krankvor	Fortsetzung der Arbeitslosigkeit nach Arbeitsunfähigkeit				
erwstat_a	erwstat	Erwerbsstatus: Personengruppe, Leistungsart, Maßnahmeart, Arbeitsuchestatus				
grund_a	grund	Grund des Spellendes				
	sna	Status nach Abgang				

⁸ Dies betrifft beispielsweise die Variable Status nach Abgang „sna“. Recherchen haben hier z.B. gezeigt, dass die Validität dieser Variablen stark eingeschränkt ist.

Variablen- name	Variablen- name in IEBS V1.0	Bezeichnung	BeH	LEH	MTH	BewA
Persönliche Merkmale						
gebjahr_a	gebjahr	Geburtsjahr				
sex	sex	Geschlecht				
nation_a	nation	Staatsangehörigkeit				
schweb_a	schweb	Schwerbehindertenstatus				
schbild	schbild	Schulabschluss				
bild	bild	Ausbildung				
famstand		Familienstand				
gebjahr_k		Geburtsjahr, jüngstes Kind				
kind		Kinderanzahl im Haushalt				
Angaben zu Beschäftigungsverhältnis und Arbeitsuche						
stib	stib	Stellung im Beruf und Arbeitszeit				
beruf_a	beruf	Beruf				
	vstyp	Rentenversicherungsträger				
	beitgr	Beitragsgruppe	ab 99			
tentgelt_a	tentgelt	Tagesentgelt / täglicher Leistungssatz				
	gleitz	Gleitzzone	ab 03			
w93_a	w93	Wirtschaftszweig 93	ab 99			
dualo	dualo	Dauer der Arbeitslosigkeit				
endplan	endplan	Geplantes Ende der Maßnahme				
erfahrung		Berufserfahrung				
quali		Qualifikationsstufe				
mobil		Mobilität				
gesund_ein		Gesundheitliche Einschränkungen				
auswirkung		Gesundh. Einschr. mit Auswirkung				
arbeitszeit		Gewünschte Arbeitszeit				
restanspruch		Restanspruchsdauer Alo-Bezug				
Ortsangaben						
	ao_bula	Arbeitsort Bundesland (Kreis, Gemeinde)				
	wo_bula	Wohnort Bundesland (Kreis, Gemeinde)	ab 99			
	ao_rd	Arbeitsort Regionaldirektion (Arbeitsagentur, Geschäftsstelle)				
	wo_rd	Wohnort Regionaldirektion (Arbeitsagentur, Geschäftsstelle)	ab 99			
wo_aatyp_a	wo_aatyp	Wohnort Regionaltyp der Arbeitsagentur	ab 99			
wo_ost_west		Wohnort Ost-West				

- In den Spells dieser Quelle vorhanden
- In den Spells dieser Quelle nicht vorhanden

Tabelle 1: Variablen des Scientific Use Files der IEB

2.2.2. Mengengerüst des IEBS-SUF V1

Die nachfolgende Tabelle beschreibt das Mengengerüst des IEBS-SUF V1 und stellt dieses den Fallzahlen in der IEBS V.1.0 gegenüber. Hierzu sind die Fallzahlen der Spells jeweils getrennt für die einzelnen Quellen der Integrierten Erwerbsbiografien ausgewiesen. Insgesamt enthält der IEBS-SUF 18,959,982 gegenüber 20,870,743 Spells in der IEBS V.1.0 (gesplittete Spells).

Quelle	Anzahl Spells in dem IEBS-SUF V1 (Anteil in %)	Anzahl Spells in der IEBS V.1.0 (Anteil in %)
BeH	12.162.676 (64.15%)	13.760.149 (65.93%)
LeH	3.153.769 (16.63%)	3.457.471 (16.57%)
MTH – ABM, SAM, usw.	155.924 (0.82%)	155.732 (0.75%)
MTH – FbW, TM, DSL	202.759 (1.07%)	202.139 (0.97%)
MTH – FF	66.289 (0.35%)	66.071 (0.32%)
MTH – ESF-BA	10.201 (0.05%)	9.670 (0.05%)
BewA	3.208.364 (16.92%)	3.219.511 (15.43%)
Total	18.959.982 (100%)	20.870.743 (100%)
Personenzahl	1.320.530	1.370.032

Tabelle 2: Anzahl und Anteil gesplitteter Spells pro IEB-Quelle

Durch das spezifische Abschneideverfahren (nur Spells ab dem 01.01.1992) ist der Datensatz des Scientific Use Files gegenüber der IEBS V.1.0 leicht reduziert. So enthält der IEBS-SUF V1 bezogen auf die Anzahl der gesplitteten Spells eine um rund 2 Mio. Spells geringe Datenmenge als die IEBS V.1.0. Darüber hinaus hat das Abschneideverfahren auch Auswirkungen auf die Anzahl der erfassten Personen. Dies betrifft vor allem solche Individuen, für die nur Beschäftigten- und Leistungsempfangsmeldungen vor 1992 vorliegen. So enthält der IEBS-SUF V1 1.320.530 und die IEBS V.1.0 1.370.032 Personen. Dies entspricht einer Abweichung des SUF gegenüber der IEBS V.1.0 von 3,6 % (=49.502 Individuen).

3. Datenschutzrechtliche Anforderungen an das Scientific Use File der IEB

Der Gesetzgeber sieht für die Bereitstellung von Daten für Forschungszwecke bestimmte datenschutzrechtliche Anforderungen vor. Dies betrifft vor allem das verfassungsmäßig festgelegte Recht auf informationelle Selbstbestimmung und Bestimmungen, die näher in §16 Abs. 6 BStatG formuliert sind. Unter anderem wird dort geregelt, dass die in einem für die Forschung zur Verfügung gestellten Datensatz enthaltenen „Einzelangaben nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft zugeordnet werden können“.

Die gesetzlichen Bestimmungen sind jedoch nicht weiter konkretisiert, was dazu geführt hat, dass die Datenschutzerfordernungen im Rahmen eines Gemeinschaftsprojektes der Universität Mannheim, des Zentrums für Umfragen, Methoden und Analysen und dem Statistischen Bundesamt genauer festgelegt wurden (vgl. Müller et al. 1991).⁹ Als ein Ergebnis des Projektes wurden Regeln zur Weitergabe von Mikrodaten aufgestellt, welche die per Gesetz geforderte sogenannte „faktische Anonymität“ der Einzelangaben sicherstellen sollen. Bei der Bereitstellung von Scientific Use Files orientiert sich das IAB an diesen Vorgaben.

Grundlage jedes Scientific Use Files ist demnach neben dem Entfernen von direkten Identifikatoren eine „nur nach dem Zufallsprinzip“ gezogene Stichprobe (Müller et al. 1991: 401). Als weitere Anonymisierungsmaßnahmen sollen

1. Regionalangaben nur vergrößert weitergegeben werden.
2. Angaben über Staatsangehörigkeit nur so aggregiert weitergegeben werden, dass eine Nationalität oder identifizierbare Gruppe von Nationalitäten in der Bundesrepublik Deutschland mindestens 50.000 Einwohner umfasst.
3. alle übrigen („sichtbaren“) Variablen so aggregiert werden, dass in den univariaten Randverteilungen jede ausgewiesene Merkmalsausprägung für die Grundgesamtheit der Bundesrepublik Deutschland mindestens 5.000 Fälle umfasst (vgl. Bender et al. 1996: 36).

Abweichend von der zweiten Mikrozensus-Regel wird vom IAB für das Merkmal „Staatsangehörigkeit“ eine Grenze von minimal 30.000 Personen festgelegt (vgl. Bender et al. 1996: 37).

Diese Anonymisierungsregeln beziehen sich jedoch in erster Linie auf Querschnittsdaten, d.h. Informationen, die zu einem Zeitpunkt für eine bestimmte Stichprobe vorliegen. Die IEB hingegen enthalten Längsschnittdaten, so dass entsprechende neue Konkretisierungen der

datenschutzrechtlichen Bestimmungen erforderlich werden. Im Kern bleibt aber auch hier die durch das Gemeinschaftsprojekt vorgeschlagene Vorgabe erhalten, dass ein faktisch anonymisierter Datensatz nur eine Mindestanzahl von Personen repräsentieren darf.

Diese Umsetzung der Vorgaben von minimalen Fallzahlen erfolgt im vorliegenden Fall, indem mehrere Stichtagsauszählungen über den gesamten abgebildeten Zeitraum für alle Merkmale betrachtet werden. Analog der zweiten und dritten Anonymisierungsregel muss jede Merkmalsausprägung in jeder Stichtagsauszählung dann die entsprechende Mindesthäufigkeit aufweisen, um als faktisch anonym zu gelten (vgl. Bender et al. 1996: 36f). Wird die notwendige Fallzahl in einer Ausprägung nicht erreicht, dann wird die Ausprägung mit einer oder mehreren anderen Ausprägungen zusammengelegt, so dass die neue synthetische Kategorie die Grenze der Mindesthäufigkeit erreicht.

Die Grenzen für Fallzahlen, die Müller et al. für das Kriterium der „faktischen Anonymität“ festsetzen, werden anhand der Grundgesamtheit bestimmt. Die Überprüfung der Grundgesamtheit ist jedoch langwierig und kostenintensiv, teilweise liegen sogar keine geeigneten Daten für die Überprüfung vor. Darum hat sich im IAB die Regel bewährt, dass auch über die Hochrechnung der Häufigkeiten der geforderten Stichprobe aus der Grundgesamtheit festgestellt werden kann, ob faktische Anonymität gegeben ist. Im Hinblick auf die Struktur der IEB-Daten ist jedoch zu berücksichtigen, dass bei den Stichtagsauszählungen Personen und nicht Spells gezählt werden.

Ein gravierender Nachteil dieses Vorgehens ist jedoch, dass wesentliche Informationen durch die Anonymisierung verloren gehen. Neben diesen einfachen Heuristiken zur Umsetzung einer faktischen Anonymisierung von Mikrodaten durch Vergrößerungsstrategien hat die Literatur daher mittlerweile zahlreiche andere Verfahren vorgeschlagen, die mit einem geringeren Informationsverlust verbunden sind. Unter anderem wird vorgeschlagen, dass Merkmalsausprägungen auf Grundlage der Korrelationsstruktur der vorhandenen Daten imputiert oder zufällige Wertetauschungen vorgenommen werden. Bei stetigen Variablen schlägt die Literatur zudem vor, dass zufällige Störgrößen zu den Werten addiert oder subtrahiert werden. Gemeinsam ist diesen Verfahren, dass sie allgemein einen geringeren Informationsverlust aufweisen dürften als ein einfaches Vergrößern der Daten. Aus methodischen Gesichtspunkten werden derartige Verfahren jedoch für die Erstellung des IEBS-SUF V1 nicht angewendet.

⁹ Grundlage der Arbeiten von Müller et al. war der Mikrozensus. Da der Mikrozensus eine Stichprobe aus der Gesamtheit der Wohnbevölkerung der Bundesrepublik darstellt, wird in den Regeln als Bezugsgröße für die minimalen Fallzahlen auch die Grundgesamtheit der Bundesrepublik Deutschland gewählt.

4. Faktische Anonymisierung der IEB

Zur Überprüfung der faktischen Anonymität des Scientific Use Files der IEB werden Stichtagsauszählungen zu Grunde gelegt. Liegen die Fallzahlen in der Grundgesamtheit unterhalb der von Müller et al. (1991) geforderten Fallzahl von 5.000 bzw. 30.000 Personen, werden Anonymisierungsverfahren angewendet. Von zentraler Bedeutung für die Anonymisierung wird damit die Bestimmung des zugrunde liegenden Auswahlssatzes.

Bezogen auf die oben beschriebenen Fallzahlen in dem IEBS-SUF V1 und dem Mengengerüst der IEB (Konten: 62.233.767; Sätze: 772.944.208) ergibt sich ein Stichprobensatz von ca. 2,1%. Damit würden die Schwellenwerte für eine Anonymisierung bei 105 bzw. 630 Personen liegen.¹⁰ Hierbei ist jedoch zu berücksichtigen, dass sich der Auswahlssatz von 2,12% erst nach Anwendung des Abschneideverfahrens (Ausklammern der Spells vor 1992, siehe oben) ergibt. Im Prinzip müsste der Auswahlssatz daher um die Art des Abschneideverfahrens und der damit zusammenhängenden zweistufigen Stichprobenziehung korrigiert werden.

Um jedoch diese zweistufige Stichprobenziehung nicht explizit für die Bestimmung des Auswahlssatzes berücksichtigen zu müssen, wird als Grundlage daher die ursprüngliche Stichprobenziehung (8 Geburtstagstage) verwendet. Bei dieser ergibt sich ein Auswahlssatz von ca. 2,2% ($8 : 365 * 100 = 2,19 \%$). Dieses Vorgehen sorgt im Zweifel für tendenziell höhere Anonymisierungskriterien. Insgesamt gelten damit folgende minimale Fallzahlen als Entscheidungsgrundlage für eine Datenanonymisierung:

- 110 Personen bzw.
- 660 Personen für „besondere Arten personenbezogener Daten“ (Staatsangehörigkeit, Schwerbehindertenstatus, gesundheitliche Einschränkungen und Auswirkung der gesundheitlichen Einschränkungen auf die Vermittlung)

Von der strengen Anwendung der Fallzahl-Kriterien wird unter bestimmten Umständen jedoch abgewichen – und zwar immer dann, wenn nachgewiesen werden kann, dass die IEB nicht der vollständigen Grundgesamtheit entspricht. Dies bedeutet, dass in Zeiträumen, in denen die IEB unvollständige Datenlagen aufweisen, geringe Fallzahlen als Anonymisierungskriterium herangezogen werden können. In solchen Fällen lässt sich unter Rückgriff auf Angaben über die Grundgesamtheit (z. B. anhand der Amtlichen Nachrichten der BA) nachweisen, dass kleinere Fallzahlen nicht gegen die datenschutzrechtlichen Vorgaben verstoßen.

¹⁰ $105 = 5.000 \times 2,1\%$; $630 = 30.000 \times 2,1\%$

4.1. Besonderheiten bei Beschäftigungs-, Leistungsempfangs- und Arbeitsuchezeiten

Im Grundsatz werden zur Überprüfung der faktischen Anonymität bei den Beschäftigungs-, Leistungsempfangs- und Arbeitsuchezeiten Stichtagsauszählungen zum 30.06. jeden Jahres (1990 – 2004) für die in den Quellen enthaltenen Merkmale durchgeführt. Werden die geforderten Fallzahlen je Ausprägung dabei nicht erreicht, wird eine Zusammenlegung der Merkmale allerdings unabhängig vom Stichtag für den gesamten abgebildeten Zeitraum durchgeführt. Dies führt zu folgendem Ergebnis: Wird bei einer Auszählung eine zu geringe Fallzahl ermittelt, wird die Meldungsart für den gesamten Beobachtungszeitraum zu einer übergeordneten Kategorie zusammengefasst.

4.2. Besonderheiten Maßnahmeteilnahmezeiten

In Bezug auf Maßnahmeteilnahmen wird beim Scientific Use File der IEB eine neuartige Anonymisierungsregel umgesetzt. Das Besondere an den Maßnahmeteilnahmemeldungen begründet sich damit, dass Maßnahmen „An- und Auslaufzeiten“ haben, was dazu führt, dass in der Phase der Einführung und der Abschaffung eines Maßnahmetypus jeweils nur wenige Personen mit einer entsprechenden Maßnahmeteilnahme erfasst sind. Dies hätte bei strikter Anwendung der oben genannten Regeln zur Folge, dass Maßnahmetypen über ihre gesamte Laufzeit zusammengefasst werden. Eine wesentliche Eigenschaft des Scientific Use Files der IEB soll jedoch sein, differenzierte Analysen von Maßnahmeteilnahmen zu ermöglichen. Diese Eigenschaft würde durch die Zusammenlegungen von Maßnahmen verloren gehen.

Um trotz notwendiger Anonymisierungsmaßnahmen einen für die Forschung interessanten Scientific Use File anbieten zu können, wurden folgende Anonymisierungsregeln für Daten zu Maßnahmeteilnahmen entwickelt:

- Maßnahmeteilnahmen werden in Jahresscheiben unterteilt ¹¹
- Stichtagsauszählungen zu Maßnahmemerkmale werden in jeder Jahresscheibe quartalsweise durchgeführt (zum 15.02./15.05./15.08./15.11.)
- Die Anonymisierungsregeln, die eine minimale Fallzahl vorschreiben, werden auf die quartalsweisen Stichtagsauszählungen übertragen
- Wird zu einem Quartalsstichtag in einer Kategorie nicht die minimale Fallzahl erreicht, wird diese Kategorie mit einer oder mehreren anderen Kategorien zusammengelegt. Dies wird allerdings nur für die entsprechende Jahresscheibe umgesetzt.

¹¹ Das genaue Vorgehen bei der Erzeugung von Jahresscheiben kann dem Anhang entnommen werden.

Folgende Abbildung (Abbildung 4) illustriert dieses Vorgehen anhand eines hypothetischen Szenarios, bei dem zu Beginn einer Maßnahme und zum Auslaufen jeweils die geforderten Fallzahlen unterschritten werden. Eine detailliertere Darstellung kann dem Anhang entnommen werden. Bei Anwendung der oben angesprochenen Anonymisierungsregeln würde dies dazu führen, dass die gesamte Maßnahmeart für den gesamten Beobachtungszeitraum vergrößert ausgewiesen wird. Bei der jahresweisen Anonymisierung, wird die Maßnahmeart dagegen nur für diejenigen Jahre in eine übergeordnete Maßnahmekategorie zusammengefasst, in denen eine von vier Quartals-Fallzahlen unterhalb der geforderten Fallzahl liegt.

	Jahresscheibe				
	2000	2001	2002	2003	2004
Hypothetische Teilnehme- zahlen zur Maßnahme XY	100	200	300	120	90
ursprünglicher Anony- misierungsumfang					
angepasster Anonymi- sierungsumfang					
		Dezidierte Ausweisung der Maßnahme möglich, ausreichende Fall- zahlen			
		Dezidierte Ausweisung der Maßnahme nicht möglich, Fallzahlen zu gering			

Abbildung 4: Prinzip der jahresweisen Anonymisierung von Maßnahmeteilnahmen

4.3. Umfang und Art der Datenveränderung

Nachfolgend sind die Ergebnisse der Überprüfung der Anonymisierungskriterien zusammenfassend dargestellt. Hierzu weist die Tabelle aus, für welche Variablen Anonymisierungen erforderlich werden.

Variablenbezeichnung	Variablenname	Anonymisierung	Besonderheiten
Identifikationsnummern			
Personennummer	persnr	systemfrei	-
Betriebsnummernzähler	betnrz	systemfrei	-
Spellanfang und -ende			
Beginndatum des gemeldeten (Original-)Zeitraums	begorig	Nein	-
Enddatum des gemeldeten (Original-)Zeitraums	endorig	Nein	-
Beginndatum der Episode nach dem Episodensplitting	begepi	Nein	-
Enddatum der Episode nach dem Episodensplitting	endepi	Nein	-
Generierte technische Merkmale			
Quelle	quelle	Nein	-
Kombination der Quellen	kom_quel	Nein	-
Spellzähler pro Konto	spell	Nein	-
Anzahl der Spells pro Konto	nspell	Nein	-
Spellzähler pro Episode	level2	Nein	-
Anzahl der Spells pro Episode	nlevel2	Nein	-
Spellzähler pro Episode und Quelle	level1	Nein	-
Anzahl der Spells pro Episode und Quelle	nlevel1	Nein	-
Bereinigungskennzeichen	berknz	Nein	-
Status des Endedatums	stendat	Nein	-
Kennzeichen des Vollständigkeitsgrads der BewA-Spells in einem Konto	knz_asu	Nein	-
Personenstatus vor, während und nach dem aktuellen Spell			
Erwerbstatus vor Arbeitsuche	estatvor	Nein	-
Fortsetzung der Arbeitslosigkeit nach Arbeitsunfähigkeit	krankvor	Nein	-
Erwerbsstatus: Personengruppe, Leistungsart, Maßnahmeart, Arbeit-suchestatus	erwstat_a	Ja	für BeH und MTH, bei MTH: Anonymisierung über Jahres-scheiben – Zusammenfassung von Kategorien
Grund des Spellendes	grund_a	Ja	Für LeH und BewA - Zusammenfassung von Kategorien

- Fortsetzung der Tabelle auf der nächsten Seite -

Persönliche Merkmale			
Geburtsjahr	gebjahr_a	Ja	Anonymisiert für Personen mit Spell unter 16 und über 65 Jahre
Geschlecht	sex	Nein	-
Staatsangehörigkeit	nation_a	Ja	Reduzierung auf 4 Ausprägungen
Schwerbehindertenstatus	schweb_a	Ja	Reduzierung auf 3 Ausprägungen
Schulabschluss	schbild	Nein	-
Ausbildung	bild	Nein	-
Familienstand	famstand	Nein	-
Geburtsjahr, jüngstes Kind	gebjahr_k	Ja	anonymisiert gegenüber IEB – Zusammenfassung von Kohorten vor 1963
Kinderzahl im Haushalt	kind	Ja	anonymisiert gegenüber IEB - Kategorisierung
Angaben zu Beschäftigungsverhältnis und Arbeitsuche			
Stellung im Beruf und Arbeitszeit	stib	Nein	-
Berufsordnung	beruf_a	Ja	Ausweis der 2seller plus Zusammenfassung
Tagesentgelt	tentgelt_a	Ja	Kategorisierung auf 14 Ausprägungen
Wirtschaftszweig 93	w93_a	Ja	Ausweis der 3steller plus Zusammenfassungen
Dauer der Arbeitslosigkeit	daualo	Nein	-
Geplantes Endedatum der Maßnahmeteilnahme	endplan	Nein	-
Berufserfahrung	erfahrung	Nein	-
Qualifikationsstufe	quali_n	Nein	-
Mobilität	mobil_n	Nein	-
Gesundheitliche Einschränkungen	gesund_ein	Nein	-
Gesundheitliche Einschränkungen - Auswirkungen auf Vermittlung	auswirkung	Nein	-
Gewünschte Arbeitszeit des Stellengesuchs	arbzeit	Ja	anonymisiert gegenüber IEB – Zusammenfassung von Kategorien
Restanspruchsduer Arbeitslosengeld	restanspruch_a	Ja	anonymisiert gegenüber IEB - Kategorisierung
Ortsangaben			
Wohnort Regionaltyp der Bundesagentur	wo_aatyp_a	Ja	Reduzierung auf 6 Ausprägungen
Wohnort Ost-West	wo_ost_west	Ja	anonymisiert gegenüber IEB - Kategorisierung

Tabelle 3: Variablen des Scientific Use Files der IEB und der IEBS V.1.0

5. Analysepotential des Scientific Use Files der IEB

Im September 2006 wurde ein Prototyp des Scientific Use Files der IEB erstellt. Zu diesem Zeitpunkt waren für den Großteil des angestrebten Variablenspektrums bereits Anonymisierungsvorschläge erarbeitet worden, die im Prototyp umgesetzt wurden.

Zur Überprüfung des Analysepotentials der anonymisierten Daten hatte sich Herr Caliendo¹² vom Deutschen Institut für Wirtschaftsforschung (DIW Berlin) bereit erklärt, den vorliegenden Prototyp des Scientific Use Files der IEB zu testen. Die Grundidee des Tests war, Evaluationsanalysen zur Wirksamkeit des Instruments „Überbrückungsgeld“ (ÜG), die im Rahmen der Hartz-Evaluation I-III durchgeführt wurden, mit Hilfe des anonymisierten Prototyps zu replizieren. An diesen war Herr Caliendo maßgeblich beteiligt.

Technisch wurden zu diesem Zweck aus dem Prototypen des Scientific Use Files der IEB-Datenauszüge zum Überbrückungsgeld analog dem Vorgehen der Datenlieferungen zur Hartz-Evaluation (Modul 1e) erstellt. Anschließend wurden die Auswertungen – soweit dies die Neukategorisierungen der Daten zur Anonymisierung ermöglichten – entsprechend des methodischen Verfahrens bei der Hartz-Evaluation (Modul 1e) umgesetzt.

Die Ergebnisse deuten darauf hin, dass eine direkte Vergleichbarkeit nur bedingt möglich ist – d.h. die Ergebnisse mit anonymisierten und nicht-anonymisierten Daten voneinander abweichen. Weitere Untersuchungen zeigen aber, dass die Verteilungen der Mehrzahl der Variablen im Prototyp und Hartz-Lieferungen ähnlich sind. Die Analysen zur Teilnahmewahrscheinlichkeit an Überbrückungsgeld dagegen, zeigen ähnliche Ergebnisse wie bei den Analysen der Hartz-Evaluation.

Die durchgeführten Untersuchungen zum Analysepotential lassen sich dahingehend interpretieren, dass eine eingeschränkte Reproduzierbarkeit der Analysen der Hartz-Evaluation möglich ist.

Weitere Untersuchungen wurden ergänzend mit dem fertigen Datensatz des Scientific Use Files der IEB durchgeführt. Bei diesen Analysen wurden univariate Verteilungsmaße zwischen der IEBS V.1.0 und dem Scientific Use File der IEB gegenübergestellt.

12 Kontakt: Dr. Marco Caliendo; DIW Berlin – Abteilung Staat; Königin-Luise-Str. 5 - 14195 Berlin ; www.caliendo.de

6. Beschreibung der Variablen des Scientific Use Files der IEB

Im nachfolgenden Abschnitt werden die Variablen des IEBS-SUF V1 kurz beschrieben. Zu den Variablen, die nicht in der IEBS V.1.0 enthalten sind oder bei denen Anonymisierungen vorgenommen wurden, ist ergänzend eine Beschreibung der Variablenausprägungen aufgeführt (ohne Fallzahlen). Um die Art der Datenanonymisierung transparent zu halten sind hierzu die Ausprägungen der IEBS V.1.0 bzw. der IEB den Ausprägungen im IEBS-SUF V1 gegenübergestellt. Alle weiteren Informationen (Detailbeschreibungen der Variablen, Besonderheiten, Ausprägungen und Fallzahlen) zu den im Scientific Use File vorhandenen Variablen, die auch in der IEBS V.1.0 enthalten sind, können der Dokumentation der Stichprobe zu den Integrierten Erwerbsbiografien (IEBS V.1.0) entnommen werden (http://doku.iab.de/fdz/reporte/2005/DR_6.pdf).

6.1. Identifikationsnummern

6.1.1. Personennummer („persnr“)

Gefüllt für	Alle Quellen
Art der Datenveränderung	systemfrei
Beschreibung	Die Personennummer ist ein systemfreier Identifikator für eine Person. (Der ursprüngliche Personenidentifikator wird über die Relation Sozialversicherungsnummer und BA-Kundennummer generiert.)

6.1.2. Betriebsnummernzähler („betnrz“)

Gefüllt für	BeH
Art der Datenveränderung	Neue Variable – angelehnt an Betriebsnummer
Beschreibung	Der Betriebsnummernzähler weist für jede Kombination Personennummer-Betriebsnummer einen eindeutigen Zähler zu. Damit lassen sich Betriebswechsel auf Personenebene identifizieren. Gleiche Betriebe im Erwerbsverlauf einer Person erhalten den gleichen Zähler, wodurch Wiedereinstellungen im gleichen Betrieb erkennbar werden. Dieses Attribut wurde anstelle der Betriebsnummer in den Datensatz aufgenommen, um zusätzliche betriebspezifische Datenanonymisierungen zu vermeiden. Ein erläuterndes Beispiel ist in der untenstehenden Tabelle aufgeführt (die dort aufgeführten Werte sind fiktiv).

Persnr	Betriebsnummer (Beispiel)	betrnz_n
1	33	1
1	21	2
1	33	1
2	70	1

6.2. Spellanfang und –ende

6.2.1. Beginn- und Enddatum des Originalspells („begorig“ und „endorig“)

Gefüllt für	Alle Quellen
Art der Datenveränderung	keine
Beschreibung	Beide Merkmale markieren das tagesgenaue Beginn- bzw. Enddatum eines Originalspells

6.2.2. Beginn- und Enddatum einer Episode („begepi“ und „endepi“)

Gefüllt für	Alle Quellen
Art der Datenveränderung	keine
Beschreibung	Beide Variablen markieren das tagesgenaue Beginn- bzw. Enddatum einer gesplitteten Episode. Die Beginn- und Enddaten einer Episode sind dabei so gesplittet, dass nur überschneidungsfreie oder vollständig parallele Episoden erzeugt werden. Ohne Überschneidungen der Originalspells sind Beginndatum eines Originalspells und Beginndatum eines gesplitteten Spells identisch (begorig = begepi). Entsprechendes gilt für das Enddatum eines Originalspells und das Enddatum einer Episode.

6.3. Generierte technische Merkmale

6.3.1. Quelle der Information („quelle“)

Gefüllt für	Alle Quellen
Art der Datenveränderung	Technisches Merkmal
Beschreibung	
Dieses Merkmal beschreibt, aus welcher Datenquelle (MTH, BeH, LeH oder BewA) die Information des Satzes stammt. Bei den MTH-Sätzen wird zusätzlich noch nach den einzelnen Teilverfahren in CoSach unterschieden.	

6.3.2. Kombination der Quellen („kom_quelle“)

Gefüllt für	Alle Quellen
Art der Datenveränderung	Technisches Merkmal
Beschreibung	
Das Merkmal wird nach dem Episodensplitting erzeugt und zeigt an, ob und welche Überschneidungen zwischen verschiedenen Quellen vorliegen. Überschneidungen innerhalb einer Quelle werden nicht berücksichtigt.	

6.3.3. Spellzähler und Anzahl der Spells pro Konto („spell“ und „nspell“)

Gefüllt für	Alle Quellen
Art der Datenveränderung	Technisches Merkmal
Beschreibung	
Beide Variablen zeigen an, um den wievielten Originalspell es sich bei einem Konto (=Person) handelt. Hierzu enthält die Variable „spell“ einen laufenden Zähler der Originalspells eines Kontos. Die Variable „nspell“ markiert dagegen wie viele Originalspells je Konto existieren.	

6.3.4. Spellzähler und Anzahl der Spells pro Episode („level2“ und „nlevel“)

Gefüllt für	Alle Quellen
Art der Datenveränderung	Technisches Merkmal
Beschreibung	
<p>Der Spellzähler je Episode ist ein laufender Zähler je parallel-liegender Episoden (Episoden-Paket) eines Kontos. Die Variable „nlevel2“ weist dagegen die maximale Anzahl der parallel-liegenden Episoden je Konto aus. Pro Konto können mehrere solcher „Episoden-Pakete“ existieren. Je „Episoden-Paket“ wird dabei neu gezählt.</p>	

6.3.5. Spellzähler und Anzahl der Spells pro Episode und Quelle („level1“ und „nlevel1“)

Gefüllt für	Alle Quellen
Art der Datenveränderung	Technisches Merkmal
Beschreibung	
<p>Die Variable level1 ist ein Spellzähler pro Episode und Quelle und nummeriert parallele Sätze einer Episode innerhalb einer Quelle. Die vier Teilverfahren der MTH werden als separate Quellen betrachtet. Die Variable nlevel1 weist hierzu die maximale Anzahl an parallel-liegenden Episoden je Quelle aus.</p>	

6.3.6. Bereinigungskennzeichen („berknz“)

Gefüllt für	Leh und MTH
Art der Datenveränderung	Technisches Merkmal
Beschreibung	
<p>Das Bereinigungskennzeichen markiert Bereinigungen, die erfolgen, wenn in den Quelldateien mehrere Spells zu vermutlich demselben Leistungsbezug bzw. derselben Maßnahmeteilnahme vorliegen.</p>	

6.3.7. Status des Endedatums („stendat“)

Gefüllt für	LeH, MTH, BewA
Art der Datenveränderung	keine
Beschreibung	Das Merkmal zeigt an, ob das Endedatum durch eine abschließende Meldung bestätigt wurde, ob es sich um ein geplantes Endedatum oder den letzten Aktualisierungsstand handelt. Je Quelle enthält dieses Merkmal unterschiedliche Informationen.

6.3.8. Kennzeichen des BewA-Spells („ken_asu“)

Gefüllt für	BewA
Art der Datenveränderung	Neue Variable
Beschreibung	Diese Variable gibt die Art der Vollständigkeit / Nicht-Vollständigkeit eines BewA-Satzes wieder (vgl. hierzu S. 7).

	Ausprägung Scientific Use File	Ausprägung IEBS
Kein BewA-Spell	0	-
BewA-Spells der Person liegen ausschließlich vor 2000	1	-
BewA-Spells der Person liegen für alle drei Phasen vor (bis 2000, zwischen 2000 und 08,09/2000 sowie nach 08,09/2000	2	-
BewA-Spells liegen ausschließlich für die Zeit nach 08,09/2000 vor	3	-
BewA-Spells existieren vor 2000 und nach 08,09/2000	4	-

6.4. Personenstatus vor, während und nach dem aktuellen Spell

6.4.1. Erwerbsstatus vor Arbeitsuche („estatvor“)

Gefüllt für	BewA
Art der Datenveränderung	keine
Beschreibung	
Das Merkmal zeigt für BewA-Spells den Erwerbsstatus vor der Arbeitsuche an. Ab 12/2002 erfolgte eine deutliche Reduzierung der Ausprägungen. Die Werte älterer Sätze wurden auf die aktuell gültigen Werte umgeschlüsselt.	

6.4.2. Fortsetzung der Arbeitslosigkeit nach Arbeitsunfähigkeit („krankvor“)

Gefüllt für	BewA
Art der Datenveränderung	keine
Beschreibung	
Die Variable gibt an, ob die Person vor Beginn des aktuellen Arbeitslosigkeits-Spells krank gemeldet war. Die Angabe ist gültig für den ungesplitteten Originalspell und für die erste gesplittete Episode.	

6.4.3. Erwerbsstatus („erwstat“)

Gefüllt für	Alle Quellen
Art der Datenveränderung	Anonymisierungen für BeH und MTH, siehe Tabelle
Beschreibung	
Die Ausprägungen dieser Variable haben in Abhängigkeit der zugrunde liegenden Datenquelle dabei jeweils eine unterschiedliche Bedeutung: <ul style="list-style-type: none"> • für BeH-Datensätze ist die Personengruppe der Beschäftigtenmeldung gemäß der Datenerfassungs- und -übermittlungsverordnung (DEÜV) enthalten, • für LeH-Datensätze ist eine Kategorisierung der Art der Leistung enthalten, die eine Person bezieht, • MTH-Datensätze enthalten eine Differenzierung der jeweiligen Maßnahme, an der eine Person teilnimmt, und • bei BewA-Datensätzen ist der Arbeitssuchendenstatus ausgewiesen. Die Anonymisierung der Variable, ist im Folgenden quellenspezifisch (für die BeH und die MTH) aufgeführt.	

Anonymisierungen BeH-Teil

	Ausprägung Scientific Use File	Ausprägung IEBS
SV-pflichtig Beschäftigte	101	101,140,143,203
Auszubildende	102	102,141
Beschäftigte in Altersteilzeit	103	103,142
Praktikanten	105	105
Werkstudenten	106	106
Bezieher von Vorruhestandsgeld	108	108
Geringfügig entlohnte Beschäftigte	109	109
Unständig und kurzfristig Beschäftigte	110	110,118,205
Mitarbeitende Familie in Landwirtschaft	112	112
Versicherungsfreie Altersvollrentner	119	119
Über Haushaltsscheckverfahren Beschäftigte	201	201,209
Pflegepersonen	207	207,208
Grundwehrdienstleistende	301	301
Zivildienstleistende	303	303
Residualkategorie BeH	-9	-9, 104, 107, 111, 113, 114, 116, 120, 201, 202, 204, 209, 302, 999

Anmerkungen zu BeH-Meldungen:

Zu berücksichtigen ist, dass ab 01.01.1999 die Personengruppen a) Bezieher von Vorruhestandsgeld, b) Pflegepersonen, c) Grundwehrdienstleistende und d) Zivildienstleistende nicht mehr in die IEB integriert werden. Zudem: Mit dem 01.04.2003 werden auch Personen erfasst, die über das Haushaltscheckverfahren (erleichtertes Verfahren zur Erfassung von im Haushalt geringfügig beschäftigter Personen) beschäftigt sind. Soweit möglich wurden die alten Meldewerte auf die neuen Werte umgeschlüsselt.

Anonymisierungen MTH-Teil

	Ausprägung Scientific Use File	Ausprägung IEBS
ABM primärer Sektor, Grund- und verstärkte Förderung	10101	10101,10102,10103,10201,10202,10203
ABM Bauwirtschaft, Grund- und verstärkte Förderung	10105	10104,10105,10106,10204,10205,10206
ABM Dienstleistungssektor, Grund- und verstärkte Förderung	10107	10107,10108,10207,10208
ABM sonstige	10109	10103,10209,101xx,101yy,102xx,102yy
SAM Umweltsanierung und -verbesserung, Ost und West	10301	10301,10302,10303,10304,10401
SAM gemeinnützige Arbeit, Ost und West	10305	10305,10306,10307,10308,10402,10403,10409,10410
SAM kommunale Arbeit, Ost und West	10309	10309,10310,10311,10312,10313,10315,10316,10318,10404,10405,10406,10407,10408,10411,10412,10414
SAM OfW	10314	10314
SAM für Ältere, Ost und West	10317	10317,10413
EGZ bei Einarbeitung	10701	10701
EGZ bei erschwelter Vermittlung	10702	10702
EGZ für Ältere	10703	10703
EGZ §218	10705	10705,10706
Einstellungszuschuss bei Neugründung (EZN)	10801	10801,108xx,108yy
Überbrückungsgeld	11101	11101,111xx,111yy
BHI für LZA 1 b. u. 2 Jahre	11301	11301
BHI für LZA 2 b. u. 3 Jahre	11302	11302
BHI für LZA 3 Jahre +	11303	11303
ExGZ	11801	11801
Beauftragung von Trägern (§421i)	11901	11901
FF sonstige	13001	13001,13008,13009,1300X,1300Z,130xx,130yy
FF für Bildung und Qualifikation	13002	13002,13004,13005
FF - ExGH	13003	13003
FF von Eigenbemühungen, Beratung und Vermittlung durch Dritte	13006	13006
FF regionaler Mobilität	13007	13007
FBW-sonstige	14001	14001,140\$\$,140ZZ,140XX
FBW Feststellungsmaßnahmen – alle Arten	14002	14002,14003,14004,14010

FBW mit Abschluss – alle Arten	14020	14020,14040,14041
FBW ohne Abschluss – alle Arten	14031	14031,14034,14050,14060,14062,14070
Trainingsmaßnahmen – alle Arten	14101	14101,14102,14103,14104,141ZZ
DSL – alle Arten	14201	14201,14202,14203,14204,14210,14220,14231,14232,14233,14234,14240,14241,14250,14260,14261,14262,14270,14280,14290,142ZZ,142XX
ESF-BA – alle Arten	15001	15001,15002,15003,15004,15005,15006,15008,150XX,150ZZ,14090,14290
Residualkategorie MTH	-7	10402,10403,10501,10502,10503,10504,10505,10506,10507,10508,10509,10510,10511,10512,10513,10514,105xx,105yy,10601,10602,10603,10704,10901,10902,10903,10904,10905,11401,11501,11502,11503,11504,11505,11506,11507,11508,11509,11510,11511,11512,11513,115xx,115yy,11601,11602,116xx,116yy

6.4.4. Grund des Spellendes („grund“)

Gefüllt für	BeH, LeH, BewA
Art der Datenveränderung	Anonymisierungen je nach Quelle, siehe Tabelle
Beschreibung	
<p>Die Variable „Grund des Spellendes“ enthält je nach Quelle unterschiedliche Informationen:</p> <ul style="list-style-type: none"> • für BeH-Datensätze ist der Grund der Abgabe der Beschäftigtenmeldung enthalten, • für LeH-Datensätze ist der Grund der Beendigung des Leistungsbezugs enthalten, • bei BewA-Datensätzen ist der Grund enthalten, warum eine Person nicht weiter als arbeitsuchend geführt wird, und • MTH-Datensätze enthalten keinen „Grund des Spellendes“. 	

Das Spektrum an möglichen Gründen für die Beendigung von Arbeitsuchendenzeiten wurde im April 2003 neu definiert. Alte Ausprägungen wurden auf die neuen, seit der Umstellung einheitlich verwendeten Ausprägungen umgeschlüsselt. Meldungen mit den Abgabegründen

- Ende der Beschäftigung,
- Sonstige Gründe/Änderungen im Beschäftigtenverhältnis,
- Tod,
- Jahresmeldung,
- Unterbrechungsmeldung wegen Bezug von bzw. Anspruch auf Entgeltersatzleistungen und
- Meldung eines einmalig gezahlten Arbeitsentgelts, die nach dem alten Meldeverfahren DEVO/DÜVO eingegangen waren, wurden rückwirkend auf die neue Kategorisierung der Abgabegründe der DEÜV umgeschlüsselt.

Die Anonymisierung ist im Folgenden quellenspezifisch aufgeführt

Besonderheiten BeH-Teil

	Ausprägung Scientific Use File	Ausprägung IEBS
Abmeldung wegen: Ende einer Beschäftigung	30	30
Abmeldung wegen: Krankenkassenwechsel	31	31
Abmeldung wegen: Beitragsgruppenwechsel	32	32
Abmeldung wegen: sonstiger Gründe oder Änderungen im Beschäftigtenverhältnis	33	33
Abmeldung wegen: Ende einer sozialversicherungspflichtigen Beschäftigung nach einer Unterbrechung von > 1 Monat	34	34
Abmeldung wegen: Wechsel des Entgeltabrechnungssystems / Währungsumstellung während des Kalenderjahres	36	36
Gleichzeitige An- und Abmeldung wegen: Ende einer Beschäftigung	40	40
Abmeldung wegen: Tod	49	49
Jahresmeldung	50	50
Unterbrechungsmeldung wegen: Bezug von bzw. Anspruch auf Entgeltersatzleistungen	51	51
Unterbrechungsmeldung wegen: Erziehungsurlaub / Elternzeit	52	52
Unterbrechungsmeldung wegen: gesetzlicher Dienstpflicht	53	53
Meldung eines einmalig gezahlten Arbeitsentgelts	54	54
Sonstige (Ab-)Meldungen	99	35,55,56,59

Anmerkungen zu BeH-Meldungen:

Meldungen mit den Abgabegründen „Ende der Beschäftigung“, „Sonstige Gründe/Änderungen im Beschäftigungsverhältnis“, „Tod“, „Jahresmeldung“, „Unterbrechung wegen Entgeltersatzleistungen“, „Meldungen eines einmalig gezahlten Arbeitsentgelts“, die nach dem alten Meldeverfahren (vor 1999) eingegangen sind, konnten im DWH der Bundesagentur für Arbeit identifiziert und rückwirkend auf die Abmeldegründe des neuen Meldeverfahrens (nach 1999) umgeschlüsselt werden.

Besonderheiten LeH-Teil

	Ausprägung Scientific Use File	Ausprägung IEBS
Bezug von Eingliederungsgeld bei beruflicher Bildung (Wohnort-Arbeitsagentur)	1000	1000
Arbeitsaufnahme	1001	1001
Mutterschaftsgeld	1004	1004
Sonstige Gründe	1005	1003,1005,1010,1014, 1021, 1034
Ablauf Bewilligungs-Abschnitt Arbeitslosenhilfe	1006	1006
Abbruch der Maßnahme	1008	1008
Anspruch Arbeitslosengeld erschöpft	1012	1012
Programminterne Einstellung	1015	1015
Gesetzliche Dienstpflicht	1018	1018, 1026 (mit Endedatum nach 1.1.1990), 1028
Ortsabwesenheit / Urlaub	1019	1019
Bezug von Unterhaltsgeld (Wohnort-Arbeitsagentur)	1020	1020
Anspruch auf Krankengeld (für gesetzlich Versicherte)	1023	1023
Säumniszeit	1024	1024
Nichterscheinen zur Meldung	1025	1025
Bezug von Übergangsgeld	1027	1027
Umzug in anderen Arbeitsagentur-Bezirk	1030	1030
Anspruch auf Altersrente	1031	1031
Bezug von Unterhaltsgeld (Arbeitsort-Arbeitsagentur)	1032	1032
Kuraufenthalt	1033	1033
Sperrzeit	1035	1035
Ablauf der Maßnahme bzw. Leistung (z.B. Eingliederungsgeld Sprachkurs)	1036	1036
Eigene Abmeldung	1037	1037
Bezug sonstiger Leistungen	1040	1011,1022,1026 (mit Endedatum vor 01.01.1990), 1029
keine Abmeldung – noch laufender Leistungsbezug	1050	-

Besonderheiten BewA-Teil

	Ausprägung Scientific Use File	Ausprägung IEBS
Krank (Arbeitsunfähigkeit, Kur, Mutterschutz)	41	41
Durch BA/ARGE in Arbeit vermittelt	60	29,30,32,33,34,35,60,80
Vermittlung in ABM/SAM/BSI	61	31,36,61
Wiedereinstellung beim gleichen Arbeitgeber	65	42,65
Selbst gesuchte Arbeitsaufnahme	66	37,38,66
Abgang in Selbständigkeit	67	46,47,67
Abgang in Wehr-/Zivildienst	68	53,68
Aufnahme einer schulischen Ausbildung	69	40,49,69
Aufnahme einer betrieblichen Ausbildung	70	43, 70
Vermittlung in FbW / Deutschsprachlehrgang	71	39,71
Nichterneuerung der Meldung	73	73
Fehlende Verfügbarkeit / Mitwirkung	74	44,51,74
Sonderregelungen (z.B. §§ 125, 428, 429 SGB III)	75	50,75
Ausscheiden aus dem Erwerbsleben	76	52,76
Betreuung durch andere Verwaltungseinheit	77	45,77,81
Sonstige Gründe oder unbekannt	78	48,54,64
Sonstige Maßnahmen der aktiven Arbeitsmarktpolitik	79	62,63,79
Keine Zuordnung möglich	-9	-9
keine Angabe	-7	-7

6.5. Persönliche Merkmale

6.5.1. Geburtsjahr ("gebjahr")

Gefüllt für	Alle Quellen
Art der Datenveränderung	Anonymisierung, siehe Tabelle
Beschreibung	<p>Das Geburtsdatum wurde auf das Geburtsjahr reduziert. Bei Personen, die am ersten Tag des ersten abgebildeten Zeitraums ihrer Erwerbsbiographie jünger als 16 Jahre waren, ist nicht das Geburtsjahr enthalten, sondern der Wert 10. Bei Personen, die am letzten Tag des letzten abgebildeten Zeitraums ihrer Erwerbsbiographie älter als 65 Jahre waren, ist nicht das Geburtsjahr enthalten, sondern der Wert 5. Die Kappung an diesen Altersgrenzen entspricht dem Vorgehen in der IABS (vgl. Bender et al. 1996: 38 ff), und orientiert sich an den üblichen Altersgrenzen zum Ein- bzw. Austrittsalter in das Erwerbspersonenpotenzial. Das Geburtsjahr lässt sich dadurch über das Beginndatum der Episode approximieren.</p>

	Ausprägung Scientific Use File	Ausprägung IEBS
im ersten Spell jünger 16	10	Geburtsjahr
im letzten Spell älter 66	5	Geburtsjahr

6.5.2. Geschlecht („sex“)

Gefüllt für	Alle Quellen
Art der Datenveränderung	keine
Beschreibung	Das Geschlecht wird aus der Versicherungsnummer gebildet, sofern diese vorhanden ist. Bei Konten ohne Versicherungsnummer wird auf Basis der BA-Kundennummer die aktuellste Geschlechtsangabe aus der Kundenhistorik der BA zugespielt.

6.5.3. Staatsangehörigkeit („nation“)

Gefüllt für	Alle Quellen
Art der Datenveränderung	Anonymisierung, Siehe Tabelle
Beschreibung	Angaben zur Staatsangehörigkeit sind für alle Datenquellen vorhanden. Die Anonymisierung wurde analog der Kategorisierung im Rahmen der Hartz-Evaluation durchgeführt, nach der sich vier Kategorien ergeben: Deutsche, EU-Staatsangehörigkeit (außer Deutsche), Angehörigkeit von Nicht-EU-Staaten und eine Restkategorie, die nicht zuordenbare Angaben und Staatenlosigkeit umfasst.

	Ausprägung Scientific Use File	Ausprägung IEBS
Deutsch	1	10
EU-Staatsangehörigkeit (außer Deutsch)	2	11 bis 49
Nicht-EU-Staatsangehörigkeit	3	sonstige
staatenlos, keine Angabe, keine Zuordnung möglich	-7	-7

6.5.4. Schwerbehindertenstatus („schweb“)

Gefüllt für	MTH, BewA
Art der Datenveränderung	Anonymisierung Siehe Tabelle
Beschreibung	
<p>Die vier Ausprägungen</p> <ul style="list-style-type: none"> • Anerkannter Schwerbehinderter (Grad der Behinderung größer als 50) • Gleichgestellt (Grad der Behinderung größer als 30 und kleiner als 50) • Zusicherung (Grad der Behinderung muss noch geprüft werden, liegt nach Angabe des Betroffenen über 30) • kein Schwerbehinderter <p>wurden zu einem dichotomen Merkmal zusammengefasst, das unterscheidet, ob ein Schwerbehinderungsgrad vorliegt oder nicht.</p>	

	Ausprägung Scientific Use File	Ausprägung IEBS
Anerkannter Schwerbehinderter / gleichgestellter Behinderter / Gleichstellung wird geprüft	1	1, 2, 3
Nicht schwerbehindert	4	4
Fehlender Wert	-7	-7

6.5.5. Schulabschluss („schbild“)

Gefüllt für	MTH, BewA
Art der Datenveränderung	keine
Beschreibung	
<p>Das Merkmal stammt aus dem BewA. Es wird der Status Schulbildung aus dem BewA zugespielt, der am Tag vor Beginn der Maßnahmeteilnahme bzw. der Arbeitsuche gültig ist.</p>	

6.5.6. Ausbildung („bild“)

Gefüllt für	BeH, MTH, BewA
Art der Datenveränderung	keine
Beschreibung	
<p>Für BeH-Spells ist der Ausbildungsschlüssel der Beschäftigtenstatistik enthalten, dass heißt die Angaben stammen aus dem Meldeverfahren und dort aus den „Angaben zur Tätigkeit“ (B2-schlüssel). Bei den BeH-Meldungen umfasst die Ausprägung zwei Dimensionen: den Schulabschluss und eventuelle berufliche Abschlüsse. Da die Angaben im Meldeverfahren</p>	

jedoch keine hohe Bedeutung haben, ist die Präzision der Angaben unklar. Für MTH- und BewA-Spells stammt die abgeschlossene Berufsausbildung aus dem BewA. Für MTH-Spells wird sie zum Beginn der Maßnahme zugespielt, für BewA-Spells zum Beginn der Arbeitsuche.

Anmerkungen zu BeH-Meldungen:

Die Angaben zur Ausbildung für BeH-Sätze stammen aus den Beschäftigtenmeldungen und dort aus dem sogenannten B2-Schlüssel. Dieser erfasst zwei Dimensionen: Den Schulabschluss und eventuelle berufliche Abschlüsse. Da die Angaben zur Ausbildung keine hohe Bedeutung im Meldeverfahren einnehmen, ist die Präzision der Angaben unklar.

6.5.7. Familienstand („famstand“)

Gefüllt für	BewA
Art der Datenveränderung	Neu aufgenommene Variable, siehe Tabelle
Beschreibung	Die Angaben zum Familienstand einer Person stammen aus der Datenquelle BewA. Es ist zu beachten, dass die ausgewiesenen Kategorien allerdings nicht trennscharf sind. Aus ihnen kann z. B. nicht abgelesen werden, ob <ul style="list-style-type: none"> • eine verheiratete, in häuslicher Gemeinschaft lebende Person Kinder erzieht, • eine verheiratete, in häuslicher Gemeinschaft alleinlebende Person überhaupt Kinder hat, • eine alleinerziehende, in häuslicher Gemeinschaft alleinlebende Person verheiratet ist.

	Ausprägung Scientific Use File	Ausprägung IEB
allein stehend	0	0
nicht allein stehend (lebt mit mindestens einer Person in häuslicher Gemeinschaft)	1	1
allein erziehend	2	2
verheiratet	3	3
keine Angabe	-7	missing, 9997
keine Zuordnung möglich	-9	9999

6.5.8. Geburtsjahr, jüngstes Kind („gebjahr_k“)

Gefüllt für	BewA
Art der Datenveränderung	Neu aufgenommene Variable, anonymisiert für Geburtsjahrgänge vor 1961, siehe Tabelle
Beschreibung	<p>Bei der Anonymisierung wurden unplausible Werte auf 99 gesetzt (Geburtsdatum des Kindes liegt nach dem Enddatum des Arbeitsuchendenzeitraums) und das Geburtsdatum wurde auf das Geburtsjahr reduziert. Zudem wurden die Kohorten vor 1961 zu Gruppen zusammengefasst.</p> <p>Das Geburtsdatum des jüngsten Kindes wird bei der Arbeitsvermittlung erfragt. Von den Arbeitsvermittlern sollen nur Kinder unter 15 Jahren erfasst werden, da eine Einschränkung der Arbeitszeit für die Betreuung dieser Kinder erlaubt ist. Aus den vorliegenden Angaben zum Geburtsjahr kann im Abgleich mit den Beginn- und Enddaten der BewA-Datensätze jedoch abgelesen werden, dass auch Kinder über 15 Jahre ausgewiesen werden. Diese Tatsache kann darin begründet sein, dass</p> <ul style="list-style-type: none"> • die Arbeitsanweisung nicht durchgängig befolgt wird und/oder • die Anweisung auch auf betreuungsbedürftige Kinder über 15 Jahren Anwendung findet (z.B. aufgrund von Krankheiten oder Behinderungen des Kindes) • das Geburtsdatum zu einem Zeitpunkt erfasst wurde, als das Kind noch unter 15 Jahren alt war und keine Löschung erfolgte als das Kind die Altersgrenze überschritt.

	Ausprägung Scientific Use File	Ausprägung IEB
Kohorten nach 1963	Geburtsjahr	Geburtsjahr
Kohorte 1962 und 1963	1962,3	1962 und 1963
Kohorten vor 1961	1959.5	1959 und 1960
	1957.5	1957 und 1958
	1954.3	1953 bis 1956
	1951.3	1951 und 1952
	1945	vor 1950

6.5.9. Kinderzahl im Haushalt („kind“)

Gefüllt für	BewA
Art der Datenveränderung	Neu aufgenommene Variable, anonymisiert gegenüber der IEB-Kodierung, siehe Tabelle
Beschreibung	
<p>Zur Anonymisierung ist die Anzahl der Kinder im Haushalt in vier Kategorien zusammengefasst. Diese Kategorisierung wurde ebenfalls für die Datenlieferung im Rahmen der Hartz I-III-Evaluation vorgenommen. Von den Arbeitsvermittlern sollen nur im Haushalt lebende Kinder unter 15 Jahren gezählt werden, da eine Einschränkung der Arbeitszeit für die Betreuung dieser Kinder erlaubt ist. Es ist allerdings fraglich, ob tatsächlich nur Kinder unter 15 Jahren gezählt werden.</p>	

	Ausprägung Scientific Use File	Ausprägung IEB
kein Kind	0	0
ein Kind	1	1
zwei Kinder	2	2
drei und mehr Kinder	3	3,4,5,6,7,8,9
Fehlender Wert	-7	Leer

6.6. **Angaben zu Beschäftigungsverhältnis und Arbeitsuche**

6.6.1. Stellung im Beruf und Arbeitszeit („stib“)

Gefüllt für	BeH
Art der Datenveränderung	keine
Beschreibung	
<p>Das Merkmal Stellung im Beruf und Arbeitszeit stammt aus den im Rahmen des Meldeverfahrens vom Arbeitgeber übermittelten „Angaben zur Tätigkeit“. Die Differenzierung nach Voll- und Teilzeitbeschäftigten richtet sich nach dem Verhältnis zwischen vertraglich vereinbarter und betriebsüblicher Arbeitszeit. Die Teilzeitbeschäftigten sind in zwei Gruppen unterteilt, wobei die Grenze entlang der Hälfte der üblichen Arbeitsstunden eines Vollzeitbeschäftigten verläuft. Infolge der Erfassung im Meldeverfahren liegt nur für Vollzeitbeschäftigte die Stellung im Beruf vor. Für Teilzeitbeschäftigte wiederum ist lediglich der Teilzeitstatus bekannt, aber nicht die Stellung im Beruf.</p>	

6.6.2. Berufsgruppe („beruf“)

Gefüllt für	BeH, MTH, BewA
Art der Datenveränderung	Anonymisierung
Beschreibung	
Die Ausprägungen der Variable entsprechen der Berufsgruppe (2Steller) der Berufsklassifikation der Bundesagentur für Arbeit, ¹³ die der Berufsklassifikation des Statistischen Bundes-Amtes von 1988 entspricht.	

Lediglich drei zusätzliche Kategorisierungen wurden vorgenommen:

- Behinderte (Berufsgruppe 55) und Rehabilitanden (66) wurden sonstigen Arbeitskräften (99) zugewiesen,
- Mithelfende Familienangehörige in der Landwirtschaft (Berufsordnung 971) wurden der Berufsgruppe 1 (Landwirte) zugeordnet.

6.6.3. Tagesentgelt / täglicher Leistungssatz („tentgelt“)

Gefüllt für	BeH, LeH
Art der Datenveränderung	Anonymisierung, siehe Tabelle
Beschreibung	
Diese Variable weist für die Quelle BeH das berechnete Tagesentgelt einer Person aus, für die Quelle LeH den täglichen Leistungssatz, beide Angaben sind in Euro zu verstehen. Die Angabe von 0 Euro für die Quelle BeH ist auf „Unterbrechungsmeldungen“ zurückzuführen, die z. B. bei Mutterschaftszeiten abgegeben werden. Die Kategorisierung der Ausprägungen erfolgte entsprechend den Vorgaben, die für die Hartz-Evaluierung I-III gemacht wurden.	

	Ausprägung Scientific Use File	Ausprägung IEBS
(Angaben in Euro)		
0	0	0
unter 19,5	1	unter 19,5
19,5 bis unter 29,5	2	19,5 bis unter 29,5
29,5 bis unter 39,5	3	29,5 bis unter 39,5
39,5 bis unter 49,5	4	39,5 bis unter 49,5
49,5 bis unter 59,5	5	49,5 bis unter 59,5

¹³ Bundesanstalt für Arbeit (1988): Klassifizierung der Berufe. Systematisches und alphabetisches Verzeichnis der Berufsbenennungen. Nürnberg: Bundesanstalt für Arbeit.

	Ausprägung Scientific Use File	Ausprägung IEBS
59,5 bis unter 69,5	6	59,5 bis unter 69,5
69,5 bis unter 79,5	7	69,5 bis unter 79,5
79,5 bis unter 89,5	8	79,5 bis unter 89,5
89,5 bis unter 99,5	9	89,5 bis unter 99,5
99,5 bis unter 109,5	10	99,5 bis unter 109,5
109,5 bis unter 119,5	11	109,5 bis unter 119,5
über 119,5	12	über 119,5
keine Angabe	-7	7

6.6.4. Wirtschaftszweig 93-Klassifizierung („w93“)

Gefüllt für	BeH
Art der Datenveränderung	Anonymisierung, siehe Tabelle
Beschreibung	
Angaben zum Wirtschaftszweig 93 sind in der BeH in den Jahren 1999-2003 vorhanden. Der fünfstellige NACE-Schlüssel, der in den IEB enthalten ist, wurde auf den 3-stelligen Schlüssel (Wirtschaftsgruppen) reduziert. Darüber hinaus wurden einzelne Wirtschaftsgruppen zusätzlich kategorisiert.	

	Ausprägung Scientific Use File	Ausprägung IEBS
	3-stellige NA- CE-Wirtschafts- gruppe	5-stellige NACE- Wirtschaftsklas- se (dargestellt über 3Steller
Abweichungen von der 3-stelligen NACE-Wirtschaftsgruppe		
Forstwirtschaft und gewerbliche Jagd	20	14, 20
Tierhaltung, Fischerei und Fischzucht	12	12, 50
Gewinnung von Salz, Steinen und Erden a.n.g., Eisenerzbergbau, NE-Metallerzbergbau (ohne Bergbau auf Uran- und Thoriumerze), sonstiger Bergbau	145	131 , 132 , 144, 145
Druckgewerbe, Vervielfältigung von bespielten Ton-, Bild- und Datenträgern	222	222, 223
Herstellung von chemischen Grundstoffen, Schädlingsbekämpfungs- und Pflanzenschutzmitteln	241	241, 242
Keramik, Herstellung von keramischen Wand- und Bodenfliesen und -platten	262	262, 263
Herstellung von Spielwaren und Sportgeräten	365	364, 365
sonstiger Landverkehr, Transport in Rohrfernleitungen	602	602, 603
Datenverarbeitungsdienste, Datenbanken	723	723, 724
neu gebildete 3-stellige Schlüssel		
Torfgewinnung und -veredelung, (Erbringung von Dienstleistungen bei der) Gewinnung von Erdöl und Erdgas, Bergbau auf Uran- und Thoriumerze	100	103 , 111 , 112 , 120
Herstellung von Lederbekleidung, gewirkten und gestrickten Fertigerzeugnissen, Zurichtung und Färben von Fellen, Herstellung von Pelzwaren	180	177 , 181 , 183
Kokerei, Mineralölverarbeitung, Herstellung und Verarbeitung von Spalt- und Brutstoffen	230	231 , 232 , 233
Linienflugverkehr, Gelegenheitsflugverkehr, Raumtransport	620	621, 622 , 623

6.6.5. Dauer der Arbeitslosigkeit („daualo“)

Gefüllt für	MTH, BewA
Art der Datenveränderung	Anonymisiert, siehe Tabelle
Beschreibung	
Das Merkmal wird zu Beginn eines jeden MTH- und BewA-Spells berechnet, auch der gesplitteten Episoden. Es handelt sich um die Anzahl der Tage seit Beginn der Arbeitslosigkeit.	

	Ausprägung Scientific Use File	Ausprägung IEBS
0	0	0
1 bis 30	30	1 bis 30
31 bis 61	61	31 bis 61
62 bis 91	91	62 bis 91
92 bis 122	122	92 bis 122
123 bis 152	152	123 bis 152
153 bis 182	182	153 bis 182
183 bis 213	213	183 bis 213
214 bis 243	243	214 bis 243
244 bis 274	274	244 bis 274
275 bis 304	304	275 bis 304
305 bis 335	335	305 bis 335
336 bis 365	365	336 bis 365
366 bis 396	396	366 bis 396
397 bis 426	426	397 bis 426
427 bis 457	457	427 bis 457
458 bis 487	487	458 bis 487
488 bis 517	517	488 bis 517
518 bis 548	548	518 bis 548
549 bis 578	578	549 bis 578
579 bis 609	609	579 bis 609
610 bis 639	639	610 bis 639
640 bis 669	669	640 bis 669
670 bis 700	700	670 bis 700
701 bis 730	730	701 bis 730
länger als 730	777	731 bis 7992

6.6.6. Geplantes Enddatum („endplan“)

Gefüllt für	MTH, BewA
Art der Datenveränderung	keine
Beschreibung	Das geplante Enddatum der Maßnahmeteilnahme entspricht dem Austrittsdatum aus der Maßnahme, wie es von der Fachkraft der Arbeitsagentur bei der Ersteingabe angegeben wurde. Abweichungen gegenüber dem Enddatum des Originalspells können sich z. B. ergeben durch ein Verschieben der Maßnahme oder einen vorzeitigen Maßnahmeaustritt.

6.6.7. Berufserfahrung („erfahrung“)

Gefüllt für	BewA
Art der Datenveränderung	Neu aufgenommene Variable, siehe Tabelle
Beschreibung	
Das Merkmal sagt aus, ob ein Bewerber im angegebenen Beruf Berufserfahrung aufweist oder nicht. Die Berufserfahrung steht in Bezug zur angegebenen Qualifikationsstufe.	

	Ausprägung Scientific Use File	Ausprägung IEB
vorhanden	1	2
nicht vorhanden	0	3
keine Angabe	-7	missing, 9997

6.6.8. Qualifikationsstufe („quali“)

Gefüllt für	BewA
Art der Datenveränderung	Neu aufgenommene Variable, siehe Tabelle
Beschreibung	
<p>Die vom Vermittler angegebene Qualifikationsstufe soll mit dem angegebenen Beruf („BKZ“-Berufskennziffer der Arbeitsuchendenmeldung, nicht im Scientific Use File der IEB enthalten) und der angegebenen Berufserfahrung abgeglichen werden. Im Handbuch des BA-Erfassungssystems für Arbeitsuchendenmeldungen (coArb) steht: „In dieses Feld ist die Qualifikationsstufe einzugeben, mit der der Bewerber vermittelt werden kann. Hierbei sind Berufserfahrung, Prüfungen, etc. zu berücksichtigen. Hat ein Bewerber geraume Zeit berufsfremd gearbeitet bzw. war entsprechend lange arbeitslos, ist ggf. die Qualifikationsstufe abzusenken. [...]</p> <p>Bei der Festlegung der Qualifikationsstufe ist wichtig, dass diese mit der vergebenen BKZ „harmonieren“ muss (z. B. BKZ 7840-101 harmoniert nicht mit Qualifikationsstufe L: Fachkräfte) bzw. dass nicht allein der Berufsabschluss ausschlaggebend ist (z.B. langjährige fundierte Berufserfahrung kann fehlenden Berufsabschluss ersetzen).“</p>	

	Ausprägung Scientific Use File	Ausprägung IEB
Spitzenkräfte (Top-Management)	1	1
Kräfte mit Hochschulniveau	2	2
Kräfte mit Fachhochschulniveau	3	3
Kräfte mit Fachschulniveau	4	4
Fachkräfte	5	5
Kräfte mit und ohne Fachkenntnisse	6	6
Sonstige Kräfte	9	12
keine Angabe	-7	missing, 9997

6.6.9. Mobilität („mobil“)

Gefüllt für	BewA
Art der Datenveränderung	Neu aufgenommene Variable, siehe Tabelle
Beschreibung	<p>Die Bezeichnung dieser Variable im Rahmen der Hartz-Evaluation war „Ausgleichsfähigkeit einer Behinderung“. Im Zuge der Recherchen der Machbarkeitsstudie ergab sich jedoch, dass kein Bezug zu einer Behinderung vorliegt, sondern die „Ausgleichsfähigkeit“ lediglich mobilitätsbezogen zu interpretieren ist. Hintergrund: Bei Meldungen zur Arbeitsuche muss der/die Arbeitsuchende angeben, ob bundesweit oder regional eingeschränkt Arbeit gesucht wird. Wenn nur ausgewählte Regionen angegeben werden, erfolgt die Nachfrage, ob trotzdem auch andere Regionen in Betracht gezogen werden können. Wenn dies der Fall ist, gilt eine Person als „regional ausgleichsfähig“, andernfalls nicht. Bei bundesweiter Suche gilt eine Person dementsprechend automatisch als „ausgleichsfähig“. In der Machbarkeitsstudie wird dieser Variable der sprechendere Name „Mobilität“ zugewiesen.</p>

	Ausprägung Scientific Use File	Ausprägung IEB
bundesweit mobil	1	J
nur regional mobil	2	N
fehlende Angabe	-7	missing

6.6.10. Gesundheitliche Einschränkungen („gesund ein“)

Gefüllt für	BewA
Art der Datenveränderung	Neu aufgenommene Variable, siehe Tabelle
Beschreibung	
Angaben zu „Gesundheitlichen Einschränkungen“ gibt es nur in der Quelle BewA. Als dichotome Variable informieren die Ausprägungen lediglich darüber, ob (keine) gesundheitlichen Einschränkungen vorliegen.	

	Ausprägung Scientific Use File	Ausprägung IEB
gesundheitliche Einschränkungen liegen vor	1	J
gesundheitliche Einschränkungen liegen nicht vor	2	N
Fehlender Wert	-7	leer

6.6.11. Gesundheitliche Einschränkungen – Auswirkung auf Vermittlung („auswirkung“)

Gefüllt für	BewA
Art der Datenveränderung	Neu aufgenommene Variable, siehe Tabelle
Beschreibung	
Angaben zu den Auswirkungen auf die Vermittlung der gesundheitlichen Einschränkungen gibt es nur in der Quelle BewA. Als dichotome Variable informieren die Ausprägungen lediglich darüber, ob (keine) Auswirkungen auf die Vermittlung gegeben sind.	

	Ausprägung Scientific Use File	Ausprägung IEB
Auswirkung auf Vermittlung gegeben	1	J
Auswirkung auf Vermittlung nicht gegeben	2	N
Fehlender Wert	-7	Leer

6.6.12. Arbeitszeit des Stellengesuchs („arbeitszeit“)

Gefüllt für	BewA
Art der Datenveränderung	Neu aufgenommene Variable, anonymisiert gegenüber der IEB-Kodierung, siehe Tabelle
Beschreibung	Arbeitsuchende geben beim Vermittlungsvorgang an, welchen zeitlichen Umfang die gesuchte Stelle annehmen sollte. Diese Angaben gehen in die Datenquelle BewA ein und werden für den Scientific Use File der IEB auf drei Kategorien verdichtet.

	Ausprägung Scientific Use File	Ausprägung IEB
Vollzeit	1	1,5,7
Teilzeit und Heimarbeit	2	2,3,6,8
Teil- oder Vollzeit	3	4
keine Angabe	-7	missing

6.6.13. Restanspruchsdauer Arbeitslosengeld („restanspruch“)

Gefüllt für	LeH
Art der Datenveränderung	Neu aufgenommene Variable, siehe Tabelle
Beschreibung	Die Restanspruchsdauer für Arbeitslosengeldzahlungen wird in der LeH ausgewiesen. Sie bezieht sich auf das Endedatum einer Leistungsempfangszeit. Es ist davon auszugehen, dass die Angabe „Missing“ eine Restanspruchsdauer von 0 Kalendertagen ausdrückt.

	Ausprägung Scientific Use File	Ausprägung IEB
unzulaessiger Wert	-7	-
kein Restanspruch	0	missing, 0
1 bis 90	90	1 bis 90
91 bis 120	120	91 bis 120
121 bis 182	182	121 bis 182
183 bis 243	243	183 bis 243
244 bis 304	304	244 bis 304
305 bis 364	364	305 bis 364
365 bis 425	425	365 bis 425
426 bis 486	486	426 bis 486

	Ausprägung Scientific Use File	Ausprägung IEB
487 bis 546	546	487 bis 546
547 bis 607	607	547 bis 607
608 bis 668	668	608 bis 668
669 bis 728	728	669 bis 728
729 bis 971	971	729 bis 971
ALHI/UHG (erwstat = 1, 2)	Missing	Missing

6.7. Ortsangaben

6.7.1. Wohnort – Regionaltyp der Arbeitsagentur („wo_aatyp“)

Gefüllt für	Alle Quellen (BeH ab 1999)
Art der Datenveränderung	keine
Beschreibung	Die 12 Regionaltypen, die auf Basis der Arbeitsagenturbezirke gebildet werden und für alle Quellen der IEB enthalten sind, wurden im Rahmen der Hartz-Evaluation zu 5 übergeordneten Gruppen zusammengefasst. Diese Kategorisierung entspricht den „Strategietypen“ (vgl. Blien/Hirschenauer, 2005)

	Ausprägung Scientific Use File	Ausprägung IEBS
Bezirke in Ostdeutschland mit dominierendem Arbeitsplatzdefizit	10	Ia, Ib, Ic
Großstädtisch geprägte Arbeitsagenturen, West, mit hoher Arbeitslosigkeit	20	IIa, IIb
Mittelstädtische und ländliche Gebiete West mit durchschnittlicher Arbeitslosigkeit	30	IIIa, IIIb, IIIc
Zentren in West mit günstiger Arbeitsmarktlage und hoher Dynamik	40	IV
Bezirke in Westdeutschland mit guter Arbeitsmarktlage und hoher Dynamik	50	Va, Vb, Vc
Fehlender Wert	-7	missing, 7, 9

6.7.2. Wohnort – Ost-West („wo ost west“)

Gefüllt für	Alle Quellen
Art der Datenveränderung	Neu aufgenommene Variable, siehe Tabelle
Beschreibung	Auf Basis der Angaben zur BA-Geschäftsstelle am Wohnort einer Person wurde die Variable „Wohnort – Ost-West“ gebildet. Sie liegt mit dichotomer Ausprägung für alle Quellen vor.

	Ausprägung Scientific Use File	Ausprägung IEBS (erste 3 Stellen von wo_aa)
Wohnort in einem neuen Bundesland	1	31-39, 42-79, 92-99, 901-965
Wohnort in einem alten Bundesland	2	81, 111-139, 211-281, 311-391, 411-459, 511-563, 611-687, 711-863
Fehlender Wert	-7	-7
keine Zuordnung möglich	-9	-9

Literatur

Bender, S. et al. (1996): Die IAB-Beschäftigtenstichprobe 1975-1990. In: Institut für Arbeitsmarkt- und Berufsforschung (Hrsg.): Beiträge zur Arbeitsmarkt- und Berufsforschung. Nürnberg: IAB.

Bernhard, S.; Dressel, C.; Fitzenberger, B.; Schnitzlein, D.; Stephan, G. (2006): Überschneidungen in der IEBS: Deskriptive Auswertung und Interpretation. Nürnberg: Reihe / Serie: FDZ Methodenreport Nr. 04/2006.

Blien, Uwe; Hirschenauer, Franziska (2005): Vergleichstypen 2005 - Neufassung der Regionaltypisierung für Vergleiche zwischen Agenturbezirken. Nürnberg: Reihe / Serie: IAB-Forschungsbericht Nr. 24/2005.

Bundesanstalt für Arbeit (1988): Klassifizierung der Berufe. Systematisches und alphabetisches Verzeichnis der Berufsbenennungen. Nürnberg: Bundesanstalt für Arbeit.

Domingo-Ferrer, J., Torra, V. (Hrsg.) (2004): Privacy in statistical databases. Berlin: Springer.

Gnoss, R., Sturm, R. (1999): Möglichkeiten und Grenzen der Bereitstellung wirtschaftsstatistischer Einzeldaten der amtlichen Statistik der Wissenschaft“. In: Statistisches Bundesamt (Hrsg.): Kooperation zwischen Wissenschaft und amtlicher Statistik – Praxis und Perspektiven. Band 34 der Schriftenreihe „Forum der Bundesstatistik“, S. 176ff.

Hamann, S. et al. (2004): Die IAB-Regionalstichprobe 1975-2001: IABS-R01. In: Zentralarchiv für Empirische Sozialforschung (Hrsg.), ZA-Information 55, 34 - 59

Helmcke, T., Knoche, P. (1992): Zur faktischen Anonymität von Mikrodaten. In Statistisches Bundesamt (Hrsg.): Wirtschaft und Statistik 3. Stuttgart: Metzler-Poeschel, 139 – 144.

Höhne, I. (2003): Methoden zur Anonymisierung wirtschaftsstatistischer Einzeldaten. In: Gnoss, R., Ronning, G. (Hrsg.): Anonymisierung wirtschaftsstatistischer Einzeldaten – Nutzerworkshop in Tübingen am 20./21. März 2003. Band 42 der Schriftenreihe „Forum der Bundesstatistik“, 69 - 94.

Höhner, J., Sturm, R., Vorgrimler, D. (2003): Konzept zur Beurteilung der Schutzwirkung faktischer Anonymisierung. In: Statistisches Bundesamt (Hrsg.): Wirtschaft und Statistik 4. Stuttgart: Metzler-Poeschel, 287 – 292.

Hummel, E., Jacobebbinghaus, P., Kohlmann, A., Oertel, M., Wübbeke, C., Ziegerer, M. (2005): FDZ Datenreport: Stichprobe der Integrierten Erwerbsbiographien IEBS 1.0. Nürnberg: Forschungsdatenzentrum des IAB in der BA.

Jaenichen, Ursula; Kruppe, Thomas; Stephan, Gesine; Ullrich, Britta; Wießner, Frank (2005): You can split it if you really want * Korrekturvorschläge für ausgewählte Inkonsistenzen in IEB und MTG. (FDZ Datenreport, 04/2005), Nürnberg.

Kaiser, U.; Wagner, J. (2007): Neue Möglichkeiten zur Nutzung vertraulicher amtlicher Personen- und Firmendaten. Nürnberg: Reihe / Serie: FDZ Methodenreport Nr. 02/2007.

Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (Hrsg.) (2001): Wege zu einer besseren informationellen Infrastruktur. Baden Baden: Nomos.

Kruppe, T.; Oertel, M. (2003): Von Verwaltungsdaten zu Forschungsdaten die Individualdaten für die Evaluation des ESF-BA-Programms 2000 bis 2006. IAB-Werkstattbericht, 10/2003, Nürnberg.

Lüttinger et al. (2003): Ergebnisse der zweiten Befragung von Nutzern der Mikrozensus-Scientific Use Files. In: Statistisches Bundesamt (Hrsg.): Wirtschaft und Statistik 11. Stuttgart: Metzler-Poeschel, 1055 – 1058.

Merz, J., Vorgrimler, D., Zwick, M. (2004): Faktisch anonymisiertes Mikrodatenfile der Lohn- und Einkommensteuerstatistik 1998. In: Statistisches Bundesamt (Hrsg.): Wirtschaft und Statistik 10. Stuttgart: Metzler-Poeschel, 1079 – 1090.

Müller, E. ; Schneider, F.-G. ; Wilke, R. A. (2006): Vergleich der Randverteilungen ausgewählter Variablen in IABS und IEBS. Mannheim: Reihe / Serie: ZEW-Dossier Nr. 12.

Müller, W. et al. (1991): Die faktische Anonymität von Mikrodaten. Stuttgart: Metzler-Poeschel.

Osikominu, A.; Fitzenberger, B.; Völter, R. (2005): Imputation Rules to Improve the Education Variable in the IAB Employment Subsample, Schmollers Jahrbuch 126 (3), 405-436.

Ronning, G., Sturm, R., Höhne, J., Lenz, R., Rosemann, M., Scheffler, M., Vorgrimler, D. (2005): Handbuch zur Anonymisierung wirtschaftsstatistischer Mikrodaten. In: Statistisches Bundesamt (Hrsg.): Statistik und Wissenschaft, Band 4. Wiesbaden: Statistisches Bundesamt.

Rosemann, M., Vorgrimler, D., Lenz, R. (2004): Erste Ergebnisse faktischer Anonymisierung wirtschaftsstatistischer Einzeldaten. In: Allgemeines Statistisches Archiv, Heft 1, S 73-99.

Stegmann, M., Luckert, H., Mika, T. (2005): Die Bereitstellung prozessproduzierter Daten der GRV im Forschungsdatenzentrum der Rentenversicherung (FDZ-RV). In: Verband Deutscher Rentenversicherungsträger (Hrsg.): Deutsche Rentenversicherung 2-3. S. 203-215.

Team Pallas (IAB-ITM) (2004): IEB-Benutzerhandbuch V3.01.

Wirth, H., 2003: Szenarien für Angriffe auf wirtschaftsstatistische Einzeldaten - Ein Überblick. Forum der Bundesstatistik, Band 42. Metzler- Poeschel. Stuttgart, S. 11-24.

Jahresweise Anonymisierung

Für MTH-Meldungen enthält der Erwerbsstatus die Art der Maßnahme, an der eine Person teilgenommen hat. Wie oben bereits beschrieben, werden bei den MTH-Meldungen nur jahresweise Anonymisierungen auf Grundlage von quartalsweisen Häufigkeitsauszählungen vorgenommen. Hierzu sind jeweils Jahres- und Quartalsplittings erforderlich. Die nachfolgende Abbildung zeigt am Beispiel eines Jahresplittings das entsprechende Aufsplitten der Daten. Die in der Abbildung enthaltenen roten Pfeile illustrieren, wie diese Maßnahmeteilnahmezeiten im Zuge der Anonymisierung aufgeteilt werden. Insgesamt lassen sich für das Splitting vier verschiedene Szenarien unterscheiden.

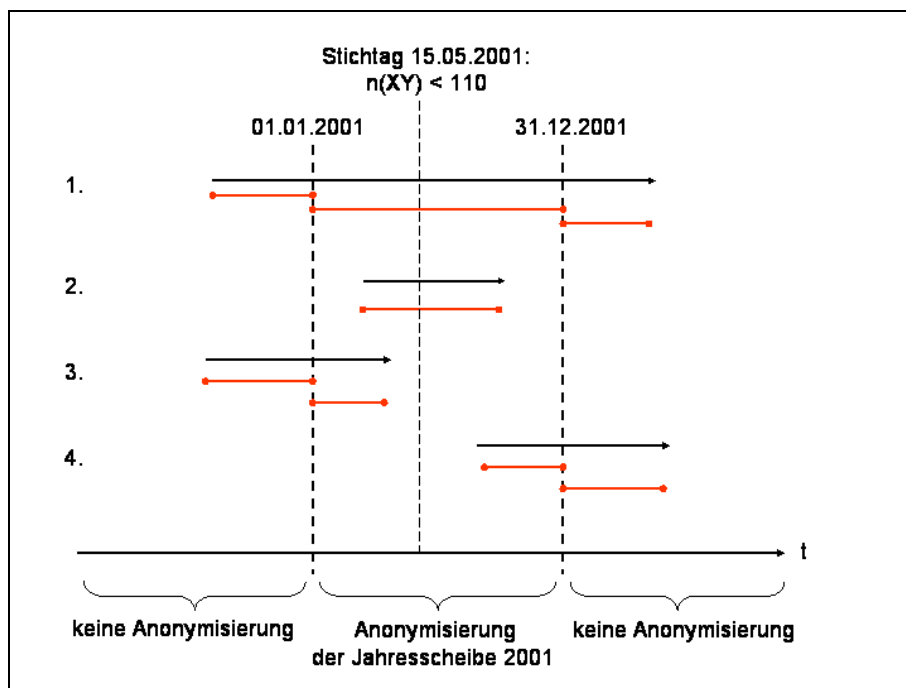


Abbildung 1: Jahresweise Anonymisierung von unterschiedlichen Verläufen von Maßnahmeteilnahmen

1. Im ersten Fall beginnt der ursprüngliche Spell vor der Jahresscheibe. Dieser wird in drei neue „jahresgesplittete“ Spells zerlegt – der erste (nicht anonymisierte) neue Spell geht vom ursprünglichen Beginndatum bis zum 31.12.2000, der zweite (anonymisierte) neue Spell geht vom 01.01.2001 bis zum 31.12.2001, der dritte (nicht anonymisierte) neue Spell verläuft dann vom 01.01.2002 bis zum Endedatum des ursprünglichen Spells.
2. Im zweiten Fall liegt der ursprüngliche Spell komplett in der zu anonymisierenden Jahresscheibe – er wird lediglich anonymisiert und nicht gesplittet.
3. Im dritten Fall beginnt der ursprüngliche Spell vor der zu anonymisierenden Jahresscheibe und endet darin. Er wird in zwei neue Spells zerlegt – der erste (nicht anonymisierte) neue Spell geht vom ursprünglichen Beginndatum bis zum 31.12.2000,

der zweite (anonymisierte) neue Spell geht vom 01.01.2001 bis zum ursprünglichen Endedatum.

4. Im vierten Fall beginnt der ursprüngliche Spell in der zu anonymisierenden Jahrescheibe und endet erst danach. Er wird in zwei neue Spells zerlegt – der erste anonymisierte Spell geht vom ursprünglichen Beginndatum bis zum 31.12.2001, der zweite (nicht anonymisierte) neue Spell geht vom 01.01.2002 bis zum ursprünglichen Endedatum.

Danksagung

Das Scientific Use File der IEB konnte nur aufgrund der Vorarbeiten und Mithilfe einer Vielzahl von Personen erzielt werden, denen wir danken möchten.

Elisabeth Roß vom Justizariat des IAB hat mit ihrer freundlichen Kooperationsbereitschaft entscheidend dazu beigetragen, dass dieser Datensatz erstellt werden konnte.

Grundlegende konzeptionelle Vorarbeiten wurden bereits 2003 von der Arbeitsgruppe „Scientific Use File des IAB für die Evaluationsforschung“ geleistet, namentlich Stefan Bender, Herbert Buscher, Jochen Kluge, Wolfram Kempe, Patrick Puhani, Bertram Raum, Barbara Rüstemeier, Hilmar Schneider, Jürgen Schupp und Heike Wirth.

Beratung und Unterstützung leistete das gesamte Team ITM des IAB, allen voran unser lieber Markus Köhler, insbesondere aber auch Volker Ferrari, Silke Hamann, Elisabeth Hummel, Katharina Sadowicz, Karl Schmidt, Christian Seysen, Melanie Stamm, Ulrich Thomsen und Manfred Ziegerer.

Zahlreiche Ratschläge und Hinweise gaben die Mitarbeiter des Forschungsdatenzentrums, insbesondere Stefan Bender, Nils Drews, Peter Jacobebbinghaus und Alexandra Schmucker.

Bedeutende Unterstützung kam auch von Wissenschaftlern des IAB: Udo Brixy, Annette Haas, Hans Kiesel und Thomas Kruppe.

Wichtige Hintergrundrecherchen waren möglich durch die Kooperation der Agentur Weißenburg (Gerhard Durst und Monika Stanek) sowie Arnulf Zitzelsberger (BA-Servicehaus).

Besonderer Dank gebührt schließlich Marco Caliendo (DIW Berlin), der einen ersten Scientific Use File der IEB -Prototypen begutachtet hat.

... und: Großer Dank an all jene Personen, die wir vergessen haben zu erwähnen!

Imprint**FDZ *Datenreport***

No. 01/2007

Publisher

The Research Data Centre (FDZ)
of the Federal Employment Service
in the Institute for Employment Research
Regensburger Str. 104
D-90478 Nuremberg

Editorial staff

Stefan Bender, Dagmar Herrlinger

Technical production

Dagmar Herrlinger

Copyright

Reproduction – also in parts – only with permission of the FDZ

Downloadhttp://doku.iab.de/fdz/reporte/2007/DR_01-07.pdf**Internet**<http://fdz.iab.de/>**Corresponding author**

Dirk Oberschachtsiek, Tel.: 0911/179-1615
E-Mail: dirk.oberschachtsiek@iab.de