

Diese Synopse wurde mit Unterstützung von Heribert Engstler, Jan Goebel, Tobias Koberg, Tatjana Mika, Philipp vom Berge und Heike Wirth erstellt

Inhalt

1	Empfohlene Anonymisierungsmaßnahmen bei Müller et al (1991)	2
2	Mikrozensus	3
3	Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB)	5
4	Rentenversicherungsdaten der DRV	7
5	National Educational Panel Study (NEPS)	9
6	DZA Freiwilligensurvey	10
7	DZA Alterssurvey	11
8	Sozio-oekonomisches Panel (SOEP)	12

1 Empfohlene Anonymisierungsmaßnahmen bei Müller et al (1991)

Grundsätzliche Empfehlungen von Müller et al. (1991) zur Anonymisierung (Mikrozensus und EVS)	
<ul style="list-style-type: none"> • Vertragliche Bindung • Geheimhaltung der lokalen Umsetzung der Stichprobenpläne • Systemfreie Anordnung der Daten 	
Alle in dieser Synopse vorgestellten Datenbestände folgen diesen Richtlinien	

Empfehlungen von Müller et al. (1991) zur Anonymisierung des Mikrozensus - Grundfiles	
Stichprobe	SUF als 70% Substichprobe des Mikrozensus (→ 0,7% der Grundgesamtheit)
Regionalangaben	<p>Bundesland (außer für Bremen und Saarland, da hier die 500.000er Regel - s. unten - nicht eingehalten werden kann)</p> <p>Grundsätzlich: Regionaltyp-Indikatoren oder vergrößerte Gemeindegrößenklasse. Dabei gilt:</p> <ul style="list-style-type: none"> • Es darf keine einzelne Gemeinde identifizierbar sein, die weniger Einwohner als 500.000 umfasst. • Ein Gemeindetyp (z.B. Gemeindegrößenklasse), dem mehrere Gemeinden zugehören, darf nicht weniger als 400.000 Einwohner umfassen. <p>anderenfalls Zusammenfassung von Kategorien</p>
Staatsangehörigkeit	Jede ausgewiesene Kategorie muss mindestens 50.000 Einwohner in der Grundgesamtheit umfassen (→ ca. 500 Fälle in den Mikrozensus-Originaldaten), anderenfalls Zusammenfassung von Kategorien
alle übrigen Merkmale	Jede ausgewiesene Merkmalsausprägung muss in der Grundgesamtheit mindestens 5.000 Fälle umfassen (→ ca. 50 Fälle in den Mikrozensus-Originaldaten), anderenfalls Zusammenfassung von Kategorien

Empfehlungen von Müller et al. (1991) zur Anonymisierung des Mikrozensus - Regionalfiles	
Stichprobe	SUF aus mindestens 85% Substichprobe des Mikrozensus (mind. → 0,85% der Grundgesamtheit). Der Auswahlsatz ist höher als beim Grundfile-SUF, weil die Auswirkungen einer Substichprobe auf das Analysepotential beim Regionalfile stärker ist.
Regionalangaben	Durch die Kombination von Regionalklassifikationen soll keine Regionaleinheit ermittelbar sein, die eine Einwohnerzahl von weniger als 100.000 Personen aufweist, anderenfalls Zusammenfassung von Kategorien
Staatsangehörigkeit, Beruf, Wirtschaftszweig, Alter	<ul style="list-style-type: none"> • Jede ausgewiesene Kategorie muss mindestens 50.000 Einwohner in der Grundgesamtheit umfassen (→ ca. 500 Fälle in den Mikrozensus-Originaldaten) • Jede ausgewiesene Kategorie muss pro übermittelter Regionaleinheit in den Mikrozensus-Originaldaten mindestens 3 Fälle enthalten <p>anderenfalls Zusammenfassung von Kategorien</p>
alle übrigen Merkmale	Jede ausgewiesene Merkmalsausprägung muss in der Grundgesamtheit mindestens 5.000 Fälle umfassen (→ ca. 50 Fälle in den Mikrozensus-Originaldaten), anderenfalls Zusammenfassung von Kategorien

2 Mikrozensus

Mikrozensus - jährliche Querschnittsfiles	
Datenmerkmale (Originaldaten)	<p>jährliche Befragung mit Auskunftspflicht</p> <p>1% Haushaltsstichprobe, N= 830 000 Personen in etwa 370 000 privaten Haushalten und Gemeinschaftsunterkünften (→ https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/Bevoelkerung/Mikrozensus.html, angesehen: 15.10.2015)</p> <p>Klumpenstichprobe mit partieller Rotation: Ein Auswahlbezirk verbleibt vier Jahre in der Stichprobe, jährlich wird ein Viertel der Auswahlbezirke durch neu in die Stichprobe aufzunehmende Auswahlbezirke ersetzt.</p>
off-site nutzbarer SUF	Mikrozensus SUF YYYY (jährlich)
Produzent des SUF	Fachabteilung Mikrozensus, Statistisches Bundesamt; GESIS stellt u.a. Syntaxen und Datendokumentationen bereit und führt Datenprüfungen durch
Anonymisierungsmaßnahmen für den Mikrozensus SUF (Querschnittsbefragungen)	
Quellen:	
<ul style="list-style-type: none"> • Missy-Mikrodaten-Informationssystem http://www.gesis.org/missy/missy-home/ • Studienbeschreibung: http://www.gesis.org/missy/studie/erhebung/studienbeschreibung/ • Leitfaden zur faktischen Anonymisierung statistischer Einzelangaben des Mikrozensus: http://www.gesis.org/missy/studie/erhebung/rechtliche-und-allgemeine-informationen/leitfaden-zur-faktischen-anonymisierung 	
Orientierung an Müller et al. (1991)	ja
Stichprobenauswahlsatz (Substichprobe)	70% der Originaldaten, N (ungewichtet) = 486.100 Personen (2011), damit steht ein Teilnehmer des Mikrozensus 2011 im Durchschnitt für 168 Personen der Grundgesamtheit
kleinste regionale Gliederung	Bundesland
Merkmal "Gemeindegrößenklasse"	Regel lt. Empfehlung von Müller et al. (1991): Einzelne Gemeinden müssen mind. 500.000 EW haben, zusammengefasste Gemeinden mind. 400.000 EW
Staatsangehörigkeit	Regel lt. Empfehlung von Müller et al.(1991): Jede ausgewiesene Nationalität bzw. Nationalitätengruppe umfasst in der Grundgesamtheit mind. 50.000 EW
weitere Merkmalsvergrößerungen	Regel lt. Empfehlung von Müller et al.(1991): Jede ausgewiesene Ausprägung muss in den univariaten Verteilungen mindestens 5.000 Fälle in der Grundgesamtheit aufweisen

Mikrozensus - Regionalfile	
Datenmerkmale (Originaldaten),	wie vorstehend
off-site nutzbarer SUF	Mikrozensus Regionalfile SUF für das Jahr 2000
Produzent des SUF	Fachabteilung Mikrozensus, Statistisches Bundesamt; GESIS stellt u.a. Syntaxen und Datendokumentationen bereit und führt Datenprüfungen durch
Anonymisierungsmaßnahmen für den Mikrozensus Regionalfile 2000 SUF	
Quellen:	
<ul style="list-style-type: none"> • Helga Christians, Heike Wirth (2009): Mikrozensus-Regionalfile: Hochrechnung, Stichprobenfehler und Anonymisierung http://www.forschungsdatennetzwerk.de/bestand/mikrozensus_regionalfile/fdz_1_mz-regionalfile_dokumentation.pdf • Heike Wirth: Kapitel 8.2 in: Regionale Standards Ausgabe 2013, GESIS-Schriftenreihe Bd. 12, 229-254 https://www.destatis.de/DE/Methoden/Methodenpapiere/Download/RegionaleStandards_Ausgabe2013.pdf?__blob=publicationFile 	
Orientierung an Müller et al. (1991)	ja
Stichprobenauswahlsatz (Substichprobe)	85% der Originaldaten, entspr. 0,85% der Grundgesamtheit
kleinste regionale Gliederung	Regionalgliederung entspricht der Empfehlung von Müller et al. (1991) Kreise bzw. Mikrozensus-Kreisregionen (MZKR) mit mindestens 100.000 EW je Regionaleinheit
Staatsangehörigkeit, Beruf, Wirtschaftszweig, Alter bzw. Geburtsjahr	Regel lt. Empfehlung von Müller et al.(1991): Falls erforderlich, regionenspezifische Aggregation: <ul style="list-style-type: none"> • pro Ausprägung mind. 500 Fälle im Original Mikrozensus (entspr. 50.000 Fällen in der Grundgesamtheit) • sowie mind. 3 Fälle pro ausgewiesener Region im Original Mikrozensus (entspr. mind. 300 Fällen in der Region) Zugunsten der Auswertbarkeit der Daten differenzierte Umsetzung der Regeln (detailliert beschrieben in o.g. Quellen)
weitere Merkmalsvergrößerungen	Regel lt. Empfehlung von Müller et al.(1991): Jede ausgewiesene Ausprägung muss in den univariaten Verteilungen mindestens 5.000 Fälle in der Grundgesamtheit aufweisen

3 Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB)

Stichprobe der Integrierten Arbeitsmarktbiografien 1975 – 2010 (SIAB-7510)	
Datenmerkmale (Originaldaten)	<p>Prozessproduzierte Daten, tagesgenau oder zu bestimmten Stichtagen. Zusammenführung aus mehreren Quellen → Integrierte Erwerbsbiografien (IEB) Vollregister: alle Personen in der Bundesrepublik, die</p> <ul style="list-style-type: none"> - sozialversicherungspflichtig oder geringfügig beschäftigt waren, - arbeitssuchend gemeldet waren, - Leistungen nach SGB II und SGB III bezogen haben. - an Maßnahmen der aktiven Arbeitsmarktpolitik nach SGB III und SGB II teilgenommen haben, <p>ab 1975/1997/1999/2007 (je nach Sachverhalt und Datenquelle, s. vom Berge/Burghardt/Trenkle (2013: 9) bis 2010 Die SIAB (Stichprobe der Integrierten Arbeitsmarktbiografien) ist eine schwach anonymisierte 2%-Stichprobe aus diesen Daten (N Personen =1.639.325) Die SIAB Regionalfile ist die faktisch anonymisierte Datenversion, auf die sich die nachfolgende Beschreibung bezieht.</p>
off-site nutzbarer SUF	<p>SIAB Regionalfile SIAB-R-7510, N Personen = 1.594.466. Besonderheit: Um tiefer gegliederte Regionalmerkmale beibehalten zu können, wurden andere Merkmale stärker vergrößert, bzw. aus dem SUF entfernt: Starke Reduktion der Merkmale zu Leistungsempfang und Arbeitssuche, unter weitgehender Beibehaltung der Merkmale zu Beschäftigung</p>
Produzent des SUF	FDZ der BA im IAB
Anonymisierungsmaßnahmen für den SUF	
Quellen:	
<ul style="list-style-type: none"> • vom Berge / Burghardt / Trenkle (2013): Stichprobe der Integrierten Arbeitsmarktbiografien Regionalfile 1975 - 2010 (SIAB - R 75 10), FDZ-Datenreport 9/2013, S. 20ff: http://doku.iab.de/fdz/reporte/2013/DR_09-13.pdf 	
Orientierung an Müller et al. (1991)?	ja, jedoch teilw. darüber hinausgehend. Allerdings ist der Stichprobenauswahlsatz höher als beim Mikrozensus-Regionalfile
Stichprobenauswahlsatz (Substichprobe)	<p>2% der Ursprungsdaten, entspr. 2% der Grundgesamtheit Wie beim schwach anonymisierten SIAB beträgt der Stichprobenauswahlsatz 2%. Aufgrund von Anonymisierungsmaßnahmen ist die Fallzahl jedoch etwas kleiner.</p>
kleinste regionale Gliederung	<p>Regionalgliederung entspricht der Empfehlung von Müller et al. (1991): Kreise, jedoch Zusammenfassung, wenn sie weniger als 100.000 EW umfassen. → Reduktion der 412 kreisfreien Städte und Landkreise auf 332 Kreisregionen</p>
Staatsangehörigkeit	<p>über die Empfehlung von Müller et al. hinausgehend: Im SIAB Regionalfile nur dichotom als "Deutsch" - "Nichtdeutsch" ausgewiesen.</p>
Beruf, Wirtschaftszweig, Alter	<p>Regel lt. Empfehlung von Müller et al.(1991): Falls erforderlich, regionenspezifische Aggregation:</p> <ul style="list-style-type: none"> • pro Ausprägung mind. 50.000 Fällen in der Grundgesamtheit • sowie mind. 3 Fälle pro ausgewiesener Region <p>→ Beruf, ausgeübte Tätigkeit: Zusammenfassung zu 120 Berufsklassen → Wirtschaftszweig: Zusammenfassung zu 14 Wirtschaftszweigen → Alter: Nur Episoden, bei denen das Alter der Person zwischen 17 und 62 Jahren liegt, sind im SUF enthalten. Episoden, die vor dem Alter von 17 enden, bzw. die nach dem Alter von 62 beginnen, sind ausgeschlossen.</p>

Stichprobe der Integrierten Arbeitsmarktbiografien 1975 – 2010 (SIAB-7510)	
weitere Merkmalsvergrößerungen	Regeln lt. Empfehlung von Müller et al.(1991): Alle übrigen im Regionalfile enthaltenen Merkmale werden so weit aggregiert, dass jede ausgewiesene Merkmalsausprägung in der Grundgesamtheit mindestens 5.000 Personen umfasst.
Spezifische Anonymisierungsmaßnahmen	
Zusammenfassung von Datenquellen	keine Unterscheidung nach kommunalem Träger / Bundesagentur für Arbeit im Merkmal "Quelle (der Prozessdaten)" → 6 Quellen werden zu 4 Quellen zusammengefasst.
Beginn und Ende der Episoden	Anpassungen für Zeiträume/Regionen mit dünner Besetzung (Einzelheiten s. Abschnitt 3.4.2 des o.g. Datenreports 9/2013)
Geldbeträge	Tagesentgelt sowie die Leistungssätze werden auf volle Euro gerundet
Geburtsdatum	Es wird nur das Geburtsjahr ausgewiesen

4 Rentenversicherungsdaten der DRV

Rentenversicherungsdaten der DRV	
Datenmerkmale (Originaldaten),	<p>Prozessproduzierte Daten</p> <p>Theoretisch Vollregister, jedoch bei einigen Datenformaten ist nur eine Stichprobe realisiert, z.B. bei der Versicherungskontenstichprobe. Grundgesamtheit: alle Personen in der Bundesrepublik, die ein Rentenkonto bei der DRV als allgemeiner Rentenversicherung, oder der Knappschaft Bahn See haben. Dies schließt lebenslange Beamte, Beamte, deren Konto gelöscht wurde und verkammert Gesicherte wie Ärzte, Notare, Rechtsanwälte aus, deren Konto mit Beitragserrstattung gelöscht wurde. Eingeschlossen sind Künstler, durch die Künstlersozialversicherung und versicherungspflichtige Selbständige (Handwerker, Hebammen, Logopäden usw.).</p>
off-site nutzbarer SUF	<p>zahlreiche Querschnitt- und Längsschnitt-SUFs, als Basisfiles und als Themenfiles, siehe http://www.fdz-rv.de/FdzPortalWeb/discontent.do?id=main_fdz_forschung&chmenu=ispvwNavEntriesByHierarchy27</p>
Produzent des SUF	FDZ-DRV
<p>Anonymisierungsmaßnahmen für die SUFs</p> <p>Quellen:</p> <ul style="list-style-type: none"> • Querschnittsdaten Stegmann / Luckert / Mika (2005): Die Bereitstellung prozessproduzierter Daten der GRV im Forschungsdatenzentrum der Rentenversicherung (FDZ-RV): Grundsätze zur faktischen Anonymisierung von Mikrodaten und zu Gastwissenschaftler Arbeitsplätzen, Kap. 4.2 http://www.fdz-rv.de/FdzPortalWeb/resDisplay.do?id=128&tabelle=3 • Längsschnittsdaten (Vollendete Versicherungsleben VVL): Stegmann, Michael (2006): Aufbereitung der Sondererhebung „Vollendete Versichertenleben 2004“ als Scientific Use File für das FDZ-RV http://forschung.deutsche-rentenversicherung.de/FdzPortalWeb/getRessource.do?key=stegmannm.pdf Stegmann, Michael (2013): Methodische Umsetzung des SK79 in einen anonymisierten Datensatz fester Satzlänge: Sequentielle Biografiedaten http://forschung.deutsche-rentenversicherung.de/FdzPortalWeb/getRessource.do?key=stegmannm.pdf 	
Orientierung an Müller et al. (1991)?	ja, plus weitergehende Maßnahmen
Stichprobenauswahlsatz (Substichprobe)	Je nach SUF zwischen 1% und 35% aus den Originaldaten
kleinste regionale Gliederung	Bundesland, bei Einhaltung der Gruppengrößen-Kriterien (s.u.) maximale Tiefe Kreise, kreisfreie Städte und weitere Raumordnungsmerkmale
Staatsangehörigkeit	<p>Regel lt. Empfehlung von Müller et al.(1991):</p> <p>Jede ausgewiesene Nationalität bzw. Nationalitätengruppe umfasst in der Grundgesamtheit mind. 50.000 EW</p>
Alter	<p>Altersbegrenzung: datensatzspezifisch an den Rändern der Verteilung, z.B. im Rentenbestand 2014: geboren bis einschließlich 1920 oder ab 1979 geboren. Im Längsschnitt in der Biografiedatensätzen: Information vor dem 14ten und nach dem 67ten Lebensjahr.</p>
weitere Merkmalsvergrößerungen	<p>Regel lt. Empfehlung von Müller et al.(1991):</p> <p>Jede ausgewiesene Ausprägung muss in den univariaten Verteilungen mindestens 5.000 Fälle in der Grundgesamtheit aufweisen</p>

Rentenversicherungsdaten der DRV	
Spezifische Anonymisierungsmaßnahmen	
Zusammenfassung von Längsschnittinformationen	Reduktion der detaillierten rentenrechtlichen Längsschnittinformationen auf 15 Zustände - sogenannte SES (Soziale Erwerbssituationen), siehe Stegmann (2013) sowie die Methodenbeschreibungen zu den Biografiedatensätzen.
Beginn und Ende der Episoden	Längsschnittfiles: Verdichtung der tagesgenauen Angaben auf Monatsangaben, Umrechnung der Jahresangaben auf Monatsangaben.
Vergrößerung von Datumsangaben und Zeitdauern	Querschnittsfiles: Datumsangaben werden aufgeteilt in Jahres- und Monatsangaben. Die Jahresangaben werden bei einer Besetzung von 5.000 Fällen und weniger in der Grundgesamtheit gruppiert. Bei Monatsangaben, die weitergegeben werden, wird auf das Quartal oder Halbjahr verdichtet. Bei der Umrechnung von Datumsangaben in Alters- oder Dauervariablen wird auf Vierteljahre gerundet und werden schwach besetzte Ausprägungsbereiche zusammengefasst.
Vergrößerung von Geldbeträgen	Geldbeträge werden gruppiert oder zumindest ganzzahlig gerundet. Sowie 'Top-Coding': Zusammenfassung der Werte am oberen Rand der Verteilung zum arithmetischen Mittelwert. Weitergabe der Bruttoeinkommen als Entgeltpunkte der Rentenversicherung, d.h. als relative Bruttolohnposition in dem gegebenen Jahr.
Vergrößerung der Entgeltpunkte	"Angaben zu Entgeltpunkten werden ganzzahlig gerundet oder bis zur ersten Nachkommastelle weitergegeben. Sowie 'Top-Coding': Zusammenfassung der Werte am oberen Rand der Verteilung.
weitere Anonymisierungsmaßnahmen	Werte zur Rentenberechnung (Entgeltpunkte, Angaben zur Dauer rentenrechtlicher Zeiten) werden nicht differenziert nach Versicherungszweig und Ost/West weitergegeben und nach oben begrenzt

5 National Educational Panel Study (NEPS)

NEPS Erwachsenenkohorte (Startkohorte 6)	
Datenmerkmale (Originaldaten),	jährliches Personenpanel
off-site nutzbarer SUF	ja
Produzent des SUF	FDZ LIfBi
Anonymisierungsmaßnahmen für den SUF (siehe (1))	
Quellen:	
<ul style="list-style-type: none"> Koberg, Tobias (2011): Starting Cohort 6 of the National Educational Panel Study: Anonymisation procedures and statistical disclosure control. Technical Report https://www.neps-data.de/Portals/0/Neps/Datenzentrum/Forschungsdaten/SC6/1-0-0/SC6_1-0-0_DM_SUP_E_Anonymization_EN.pdf 	
Orientierung an Müller et al. (1991)?	nein
Stichprobenauswahlsatz (Substichprobe)	nicht erforderlich: z.Zt. (5. Welle) 17.140 Personen der Geburtsjahrgänge 1944-1986, davon ca. 10% im Ausland Geborene Stichprobengröße ca. 3 % der Gesamt-Zielpopulation (gemessen am Mikrozensus 2011)
kleinste regionale Gliederung	Bundesland
Länderangaben (z.B. Staatsangehörigkeit)	Reduktion auf 17 Ausprägungen
Sprachen-Angaben	Reduktion auf 21 Ausprägungen
weitere Merkmalsvergrößerungen	Bei Selbständigen: Zahl der Angestellten Top Coded ab 20
Spezifische Anonymisierungsmaßnahmen	
offene Textangaben	nicht im SUF enthalten
Inhalte besuchter (Weiterbildungs-)Kurse	codiert (Kurskennziffern)

6 DZA Freiwilligensurvey

DZA Freiwilligensurvey (Welle 2009)	
Datenmerkmale (Originaldaten)	Deutschsprachige Wohnbevölkerung ab 14 Jahren. Bisher 4 Wellen in den Jahren 1999, 2004, 2009, 2014. Kein Panel, sondern jeweils unabhängige Stichproben. Die wissenschaftliche Leitung liegt seit 2009 beim DZA.
off-site nutzbarer SUF	ja (vom FDZ-DZA betreute Version 3.1)
Produzent des SUF	DZA (Version 3.1)
Anonymisierungsmaßnahmen für den SUF	
Quellen:	
<ul style="list-style-type: none"> • Kurzbeschreibung der Scientific Use Files http://www.dza.de/fileadmin/dza/pdf/fdz/FWS_1999_2004_2009_Kurzbeschreibung_3.1.pdf • Deutscher Freiwilligensurvey 2009 - Dokumentation des Erhebungsdesigns und Instruments http://www.dza.de/fileadmin/dza/pdf/fdz/Diskussionspapier_53_FWS_2009.pdf 	
Orientierung an Müller et al. (1991)?	nein
Stichprobenauswahlsatz (Substichprobe)	nicht erforderlich (100% der Originaldaten) Realisierte Stichprobengröße 2009: 20.005 Personen → circa 0,03% der Grundgesamtheit (Wohnbevölkerung Deutschlands ab 14 Jahren).
kleinste regionale Gliederung	Bundesland zusätzlicher Regionalindikator: Kreistyp
Staatsangehörigkeit und Länderangaben	In den Originaldaten ist Staatsbürgerschaft nur als deutsch/nichtdeutsch erfasst. Das offen erhobene Geburtsland der Zielperson und der Eltern (teilweise mit Nennung innerstaatlicher Regionen) wurde jeweils zu einer numerischen Einzelländervariable und einer Zusammenfassung zu Ländergruppen verarbeitet (verschiedene Ländergruppen innerhalb Europas und Asiens sowie eine sonstige Ländergruppe). Geburtsland-Codes mit einer Besetzung von unter 5 Fällen wurden im SUF auf Missing gesetzt, womit für die betroffenen Personen nur die Ländergruppe ausgewiesen wird.
weitere Merkmalsvergrößerungen	Zusammenfassung seltener Geburtsjahrgänge, Top-Coding Ausübungsjahre des Engagements, Vergrößerung Zuzugsjahr nach Deutschland.
Spezifische Anonymisierungsmaßnahmen	
offene Textangaben	nicht im SUF enthalten

7 DZA Alterssurvey

Alterssurvey (DEAS)	
Datenmerkmale (Originaldaten),	Panel mit Neustichproben im 6-Jahres-Abstand; Panel-Befragungswellen seit 2008 im 3-Jahres-Abstand (1996, 2002, 2008, 2011, 2014 noch nicht veröffentlicht): In Privathaushalten lebende Bevölkerung ab 40 Jahren (Startalter der Neustichproben jeweils 40-85 Jahre)
off-site nutzbarer SUF	ja
Produzent des SUF	DZA
Anonymisierungsmaßnahmen für den SUF	
Quellen:	
<ul style="list-style-type: none"> • DEAS-Projektwebsite http://www.dza.de/forschung/deas.html • Website des FDZ-DZA zum DEAS http://www.dza.de/fdz/deutscher-alterssurvey.html 	
Orientierung an Müller et al. (1991)?	nicht explizit, jedoch Anwendung des Kriterium: Mindesthäufigkeit = 3 (allerdings nicht: pro Regionaleinheit, sondern univariat)
Stichprobenauswahlsatz (Substichprobe)	nicht erforderlich (100% der Originaldaten). Neustichproben jeweils disproportional nach Alter, Geschlecht und Region (alte/neue Bundesländer) geschichtete Melderegister-Zufallsstichproben. Realisierte Neustichprobe 2014: 6002 Personen (40-85 Jahre) → 0,013% der Grundgesamtheit. Realisierte Gesamtstichprobe 2014: 10.325 Personen (40 und mehr Jahre) → 0,022% der Grundgesamtheit
Vergrößerung oder Codierung von Klartextangaben	<ul style="list-style-type: none"> • Berufe: Ersetzung durch ISCO-Codes (3-Steller) • sensible offene Angaben: seltene Krankheiten, spezifische Ausbildungen, Land der Kindheit, Staatsangehörigkeit, Mitgliedschaften in Vereinen mit regionalen Benennungen, spezifische Handlungsgründe mit regionalen oder personenbezogenen Angaben
kleinste regionale Gliederung	Bundesland zusätzliche Regionalindikatoren: Kreistyp, fiktive Kreiskennziffer
Staatsangehörigkeit und Länderangaben	Zusammenfassung gering besetzter (n<3) und offener Angaben zur Staatsangehörigkeit und dem Land des Aufwachsens zu Ländergruppen od. Kontinenten
Vergrößerung von seltenen Angaben	Zusammenfassung seltener Jahresangaben (Kriterium: Mindesthäufigkeit = 3) in mehreren Varianten, je nach Merkmal Top-Coding der Merkmale Haushaltsgröße, Kinder- und Geschwisterzahl
Weitere Merkmalsvergrößerungen	Familienstand: Eingetragene Lebenspartnerschaft → Ehe
Spezifische Anonymisierungsmaßnahmen	
Nicht in den SUF übernommene Merkmale	<ul style="list-style-type: none"> • Gemeindgrößenklasse • offene Angaben, die Rückschluss auf den den Wohnort ermöglichen könnten (z.B. Lektüre regionaler Medien) • Geburtstag und -monat, Monatsangaben zu familiendemografischen Ereignissen • Vornamenkürzel von Mitgliedern der Familie und des Netzwerks (dem DZA liegen nur die auf 3 Buchstaben gekürzten Nennungen vor). • alle anderen offenen Textangaben

8 Sozio-oekonomisches Panel (SOEP)

Soep-core	
Datenmerkmale (Originaldaten), Quelle:	jährliches Haushalts- und Personenpanel seit 1983
off-site nutzbarer SUF	ja, als Querschnitt-Datensätze und als longfile über alle Wellen
Produzent des SUF	Infrastruktureinrichtung Sozio-oekonomisches Panel im DIW Berlin
Anonymisierungsmaßnahmen für den SUF	
Quellen:	
<ul style="list-style-type: none"> Frick, Joachim et al. (2010): Verfahren für den Datenschutz beim Zugang zu den SOEP-Daten innerhalb und außerhalb des DIW Berlin https://www.diw.de/documents/dokumentenarchiv/17/diw_01.c.347090.de/soep_datenschutzverfahren.pdf 	
Orientierung an Müller et al. (1991)?	nein
Stichprobenauswahlsatz (Substichprobe)	nicht erforderlich: Stichprobengröße der Welle 2013: 18.093 Haushalte von knapp 40 Mio Haushalten in der Bundesrepublik, → 0,045 %-Stichprobe. (Dies ist ein Durchschnittswert, in der Stichprobe sind einige Gruppen mit Absicht überrepräsentiert).
kleinste regionale Gliederung	Bundesland Nach Vorlage eines erweiterten, vom Datenschutzbeauftragten des DIW Berlin geprüften Datenschutzkonzeptes ist off-site-Nutzung der Merkmale Gemeindegrößenklassen, BIK-Regionen und von Regionalinformationen auf der Ebene der 97 Raumordnungsregionen möglich
Spezifische Anonymisierungsmaßnahmen	
Nicht in den SUF übernommene Merkmale	Vornamen, Geburtsort, Entfernung bzw. Kodierung aller Klartextangaben, Regionalangaben unterhalb des Bundeslands (für erweit. Vertrag: unterhalb der Raumordnungsregionen)