

Meet-the-Data

@Amtliche

Statistik

13. Februar 2025,
13:15–14:00 Uhr,
online



Intro zur Reihe

- Jeden zweiten Donnerstag im Monat, 13:15-14:00 Uhr
- Programm 2024/2025: @Amtliche Statistik, Datenzugang, Datennutzung ...
- Welche Forschungsdaten kann ich für eigene Forschungsfragen nutzen?
- Wie erhalte ich Zugriff?
- Welche Unterstützungsangebote gibt es zusätzlich?

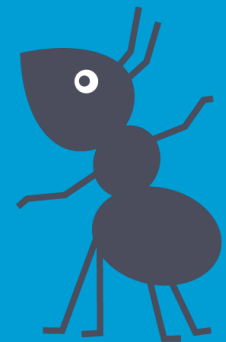
Forschungsdatenzentren





Datenbestand

Dr. Daniel Brüggmann (FDZ der Deutschen Rentenversicherung Bund)



Einleitung

- Das Forschungsdatenzentrum der Rentenversicherung (FDZ-RV) stellt der Wissenschaft und Forschung Mikrodatensätze aus dem Bestand ihrer prozessproduzierten Daten zur Verfügung.
- Einrichtung auf Empfehlung der Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI)
- Aufbauförderung durch das Bundesministerium für Bildung und Forschung (BMBF) und dem Forschungsnetzwerk Alterssicherung (FNA)

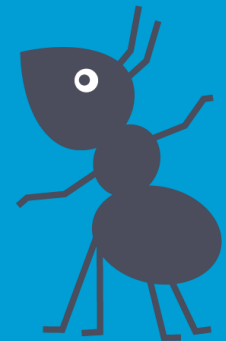
DRV Bund und Forschung





Datenzugang & Datennutzung

Online-Antrag, Dokumentationen sowie Beispieldatensätze via
<https://fdz-rv.de/> (nur für wissenschaftliche Einrichtungen)



Datenformate

- Campusfiles
 - Studenten und Lehrende
 - Digitaler Versand
 - Nicht geeignet für wissenschaftliche Forschung
- Scientific-Use-Files
 - Für Forschungsinstitute und Universitäten
 - Projektantrag zwingend erforderlich
 - Digitaler Versand
 - Für wissenschaftliche Forschung geeignet, jedoch Datensätze mit stark reduzierten Fallzahlen sodass seltene Ereignisse oder die Bestimmung heterogener Effekte schwierig werden kann

Datenformate

- Onsite-Files
 - Für Forschungsinstitute und Universitäten
 - Projektantrag zwingend erforderlich
 - Zugangswege:
 - Fernrechnen
 - Geschützte Räume der Forschungsdatenzentren
 - Berlin, Würzburg und Mannheim
 - Extrem wertvoll für die wissenschaftliche Forschung, da nicht anonymisiert aber pseudonymisiert
 - Sehr hohe Fallzahlen bis hin zu Vollerhebungen
 - Beamte, Selbständige und Mitglieder verkammerter Berufe sind jedoch nicht erfasst
 - Nur Scheidungen mit Versorgungsausgleich
 - Sehr anspruchsvolle Datensätze, da Erfassungen und Merkmale mit dem Rentenrecht korrelieren (bspw. 4 Rechtskreise)
 - Data preparation and data cleaning zwingend nötig

Zugangswege zu pseudonymisierten Daten

- Fernrechnen
 - Sehr anspruchsvoll und hoher Frustrationsgrad
 - Arbeiten im „Blindflug“
- FDZ im FDZ
 - GESIS FDZ in Mannheim
- FDZ der DRV-Bund in Würzburg
- FDZ der DRV-Bund in Berlin

Datenprodukte

- Datensätze zu Rentnern und Erwerbsminderungsrentnern
- Datensätze zu medizinischen und beruflichen Rehabilitationen
- Datensätze zu den aktiv Versicherten
- Datensätze zum Versorgungsausgleich
- Datensätze die die komplette (rentenrechtliche) Biographie spiegeln
- Datensätze aus Kooperationen:
 - Survey of Health, Ageing and Retirement in Europe (SHARE-RV)
 - Sozio-oekonomische Panel (SOEP-RV)
 - Biografiedaten ausgewählter Sozialversicherungsträger in Deutschland (BASiD)



Was haben Forschende mit diesen Daten schon herausgefunden?



Forschungspotenziale pseudonymisierter Daten

- Alle vorher genannten Datenprodukte lassen sich anhand des Pseudonyms verlinken
- Welche Informationen sind in den Datensätzen enthalten und was ist rentenrechtlich relevant?
 - Ehen im Falle des Versorgungsausgleich
 - Geburten
 - Schul- und Ausbildungszeiten
 - Pflegezeiten
 - Arbeitslosigkeit
 - Wehr- und Ersatzzeiten
 - Beschäftigungszeiten und Bruttoeinkommen
 - Wohnort und regionale Merkmale
 - Berufsklassifikationen
 - Arbeitsunfähigkeitszeiten ab 6 Wochen
 - Rehabilitationen mit Diagnosen und Entlassungsbescheinigung

Forschungspotenziale

- Daten der Rentenversicherung eignen sich nicht nur zur Analyse von Rentenaskpekten und Rentenreformen
- Evaluationen von Reformen, Gesetzen, aktiven Arbeitsmarktprogrammen
- Analysen zu individuellen Erwerbsbiographien von Müttern
- Lebenseinkommen
- Effekte von Ehen, Scheidungen oder Geburten
- Effekte von Arbeitslosigkeit
- Modellierung individueller Erwerbsverläufe und deren Heterogenität



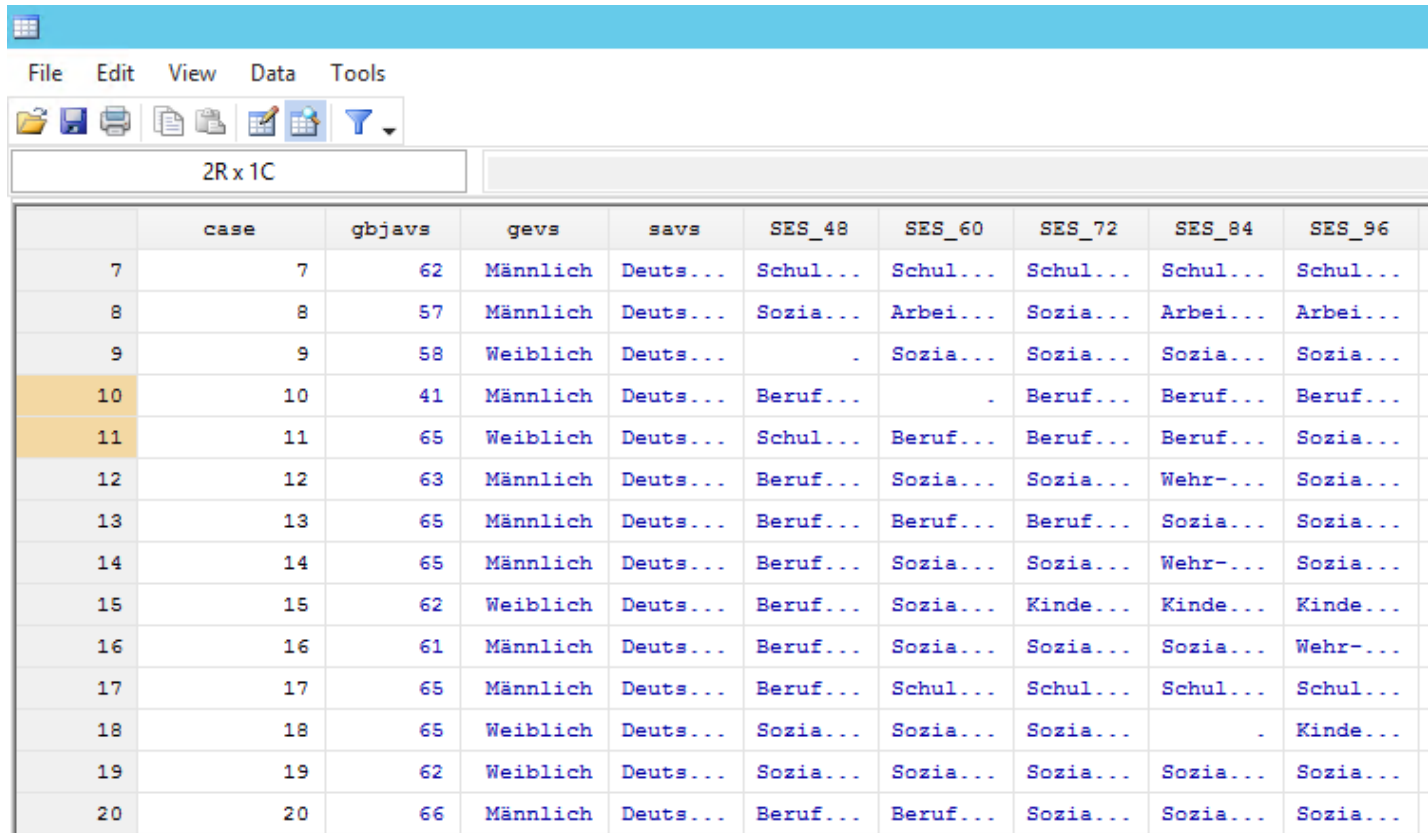
Was ist das Besondere an diesen Daten?



Mikrodaten des FDZ-RV

Blick in die Längsschnitt-Datensätze

Wide Format des SUF



	case	gbjavs	gevs	savs	SES_48	SES_60	SES_72	SES_84	SES_96
7	7	62	Männlich	Deuts...	Schul...	Schul...	Schul...	Schul...	Schul...
8	8	57	Männlich	Deuts...	Sozia...	Arbei...	Sozia...	Arbei...	Arbei...
9	9	58	Weiblich	Deuts...	.	Sozia...	Sozia...	Sozia...	Sozia...
10	10	41	Männlich	Deuts...	Beruf...	.	Beruf...	Beruf...	Beruf...
11	11	65	Weiblich	Deuts...	Schul...	Beruf...	Beruf...	Beruf...	Sozia...
12	12	63	Männlich	Deuts...	Beruf...	Sozia...	Sozia...	Wehr-...	Sozia...
13	13	65	Männlich	Deuts...	Beruf...	Beruf...	Beruf...	Sozia...	Sozia...
14	14	65	Männlich	Deuts...	Beruf...	Sozia...	Sozia...	Wehr-...	Sozia...
15	15	62	Weiblich	Deuts...	Beruf...	Sozia...	Kinde...	Kinde...	Kinde...
16	16	61	Männlich	Deuts...	Beruf...	Sozia...	Sozia...	Sozia...	Wehr-...
17	17	65	Männlich	Deuts...	Beruf...	Schul...	Schul...	Schul...	Schul...
18	18	65	Weiblich	Deuts...	Sozia...	Sozia...	Sozia...	.	Kinde...
19	19	62	Weiblich	Deuts...	Sozia...	Sozia...	Sozia...	Sozia...	Sozia...
20	20	66	Männlich	Deuts...	Beruf...	Beruf...	Sozia...	Sozia...	Sozia...

Quelle: FDZ-RV SUF VVL 2014

Mikrodaten des FDZ-RV

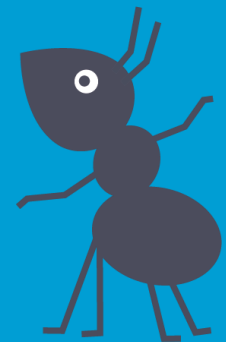
Blick in die Längsschnitt-Datensätze

Long Format der OSV Daten

32	11	AR, ab 2005_	Pflichtbeit_	0	keine Kinde_	keine beitr_	ohne Zuordn_	keine Anwen_	Schlüssel 1_	19740101	19741231	12	25469	1.2496	Sozialversi_
33	12	AR, ab 2005_	Pflichtbeit_	0	keine Kinde_	keine beitr_	ohne Zuordn_	keine Anwen_	Schlüssel 1_	19750101	19751231	12	24133	1.1066	Sozialversi_
34	13	AR, ab 2005_	Pflichtbeit_	0	keine Kinde_	keine beitr_	ohne Zuordn_	keine Anwen_	Schlüssel 1_	19760101	19761231	12	26526	1.1367	Sozialversi_
35	14	AR, ab 2005_	Pflichtbeit_	0	keine Kinde_	keine beitr_	ohne Zuordn_	keine Anwen_	Schlüssel 1_	19770101	19771231	12	28302	1.1346	Sozialversi_
36	15	AR, ab 2005_	Pflichtbeit_	0	keine Kinde_	keine beitr_	ohne Zuordn_	keine Anwen_	Schlüssel 1_	19780101	19781231	12	29416	1.121	Sozialversi_
37	16	AR, ab 2005_	Pflichtbeit_	0	keine Kinde_	keine beitr_	ohne Zuordn_	keine Anwen_	Schlüssel 1_	19790101	19791231	12	25084	.9061	Sozialversi_
38	17	AR, ab 2005_	Pflichtbeit_	0	keine Kinde_	keine beitr_	ohne Zuordn_	keine Anwen_	Schlüssel 1_	19800101	19801231	12	32581	1.105	Sozialversi_
39	18	AR, ab 2005_	Pflichtbeit_	0	keine Kinde_	keine beitr_	ohne Zuordn_	keine Anwen_	Schlüssel 1_	19810101	19811231	12	37243	1.2053	Sozialversi_
40	19	AR, ab 2005_	Pflichtbeit_	0	keine Kinde_	keine beitr_	ohne Zuordn_	keine Anwen_	Schlüssel 1_	19820101	19821231	12	38873	1.2073	Sozialversi_
41	20	AR, ab 2005_	Pflichtbeit_	0	keine Kinde_	keine beitr_	ohne Zuordn_	keine Anwen_	Schlüssel 1_	19830101	19831231	12	38197	1.1473	Sozialversi_
42	21	AR, ab 2005_	Pflichtbeit_	0	keine Kinde_	keine beitr_	ohne Zuordn_	keine Anwen_	Schlüssel 1_	19840101	19841231	12	38855	1.1331	Sozialversi_
43	22	AR, ab 2005_	Pflichtbeit_	0	keine Kinde_	keine beitr_	ohne Zuordn_	keine Anwen_	Schlüssel 1_	19850101	19851231	12	42791	1.2127	Sozialversi_
44	23	AR, ab 2005_	Pflichtbeit_	0	keine Kinde_	keine beitr_	ohne Zuordn_	keine Anwen_	Schlüssel 1_	19860101	19861231	12	49328	1.3468	Sozialversi_
45	24	AR, ab 2005_	Pflichtbeit_	0	keine Kinde_	keine beitr_	ohne Zuordn_	keine Anwen_	Schlüssel 1_	19870101	19871130	11	48043.74	1.2735	Sozialversi_
46	25	AR, ab 2005_	Anrechnungs_	1	keine Kinde_	Beitragsfre_	ohne Zuordn_	.	Schlüssel 1_	19871207	19871211	1	Wochenbeitr_	0	Arbeitsunfä_
47	26	AR, ab 2005_	Pflichtbeit_	0	keine Kinde_	Beitragsfre_	ohne Zuordn_	keine Anwen_	Schlüssel 1_	19871201	19871206	1	873.52	.0232	Sozialversi_
48	27	AR, ab 2005_	Pflichtbeit_	0	keine Kinde_	Beitragsfre_	ohne Zuordn_	.	Schlüssel 1_	19871207	19871211	0	456.15	.0121	Arbeitsunfä_
49	28	AR, ab 2005_	Pflichtbeit_	0	keine Kinde_	Beitragsfre_	ohne Zuordn_	keine Anwen_	Schlüssel 1_	19871212	19871231	0	2911.74	.0772	Sozialversi_
50	29	AR, ab 2005_	Pflichtbeit_	0	keine Kinde_	keine beitr_	ohne Zuordn_	keine Anwen_	Schlüssel 1_	19880101	19880831	8	35899.55	.923	Sozialversi_



Ausblick: Was gibt es noch?



Vorteile prozessproduzierter administrativer Daten (im Vergleich zu Umfragedaten)

- Keine Erinnerungsfehler, keine Panelmortalität, keine Auskunftsverweigerung, unabhängige Informationsquellen zu Krankheiten, Zeiten der Arbeitslosigkeit, Bruttoentgelte
- Geprüfte Daten mit ausführlicher Dokumentation, weil die einzelnen Versicherungen zu den gemeinsam vereinbarten Statistik-Merkmalen melden
- Hohe Fallzahlen
- Seit 2011: Möglichkeit, verschiedene Datensätze zu verlinken (aber strenge Datenschutzbestimmungen)

Nachteile prozessproduzierter Daten (im Vergleich zu Umfragedaten)

- Datenverfügbarkeit für Forschungsfrage
- Auffinden geeigneter Daten
- Datenzugang
- Besonderheiten in Datenerhebung und -produktion
 - Verzerrung durch Entstehungszusammenhang: Beschränkter / komplexer Merkmalskanon durch rechtliche Vorgaben (auch Datenschutzrecht!); Eigeninteressen der Beteiligten (Sachbearbeiter, Verwaltung, Bürger usw.)
 - Überlieferung / Datenaufbewahrung (Verlust / Vernichtung)
 - Datenaufbereitung (Transformation für wissenschaftlichen Nutzung)

Ansprechpartner / Kontakt für Rückfragen

Dr. Daniel Brüggmann

Deutsche Rentenversicherung Bund
Forschungsdatenzentrum

Daniel.brueggmann@drv-bund.de
+49 30 865-89357

KonsortSWD wird Rahmen der NFDI durch die Deutsche
Forschungsgemeinschaft (DFG) gefördert - Projektnummer: **442494171**



Meet-the-Data: das Programm ab März

Datum	Thema
13. März 2025	IQB-Bildungstrend 2021 im Primarbereich
10. April 2025	NEPScomp – Daten des Nationalen Bildungspanels für Studierende und Lehre
8. Mai 2025	EUROSTUDENT – Daten zur Hochschulbildung in Europa
12. Juni 2025	Die Volkshochschul-Statistik: Daten und Forschung

**Gute Daten.
Bessere Forschung!**

Meet-the-Data@Amtliche Statistik kompakt

12. Mai 2025, 13:00-16:15 Uhr	13. Mai 2025, 13:00-15:30 Uhr
Forschungsdatenzentrum der Rentenversicherung	Forschungsdatenzentrum der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin
Forschungsdatenzentrum des Deutschen Zentrums für Altersfragen	Forschungsdatenzentrum des Bundesamtes für Migration und Flüchtlinge
Forschungsdatenzentren der statistischen Ämter des Bundes und der Länder	Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung
Forschungsdatenzentrum im Kraftfahrtbundesamt	

Mögliche Fragen um die Diskussion zu starten

- Fallen Kosten für den Gebrauch der Daten an?
- Sie erwähnten, dass man die verschiedenen On-Site Datensätze miteinander verlinken kann. Gilt das auch für die Scientific-Use-Files oder die Public-Use-Files?
- Was ist der Unterschied zwischen den On-Site Datensätzen und den Scientific-Use-File und Public-Use-Files?
- Was ist der Unterschied zu den administrativen Daten des IAB und kann man generell die Daten der DRV mit den Datensätzen der anderen FDZen verknüpfen?
- Wieviel Arbeitsplätze gibt es um mit den On-Site-Daten arbeiten zu können?
- Welche Programme sind auf den Rechner installiert um die Daten auswerten zu können?
- Wie läuft der Transfer meiner Auswertungen, können die mir zugeschickt werden?
-