

Etablierung und Weiterentwicklung des Datenangebots des FDZ-RV

Dr. Michael Stegmann / Dr. Ralf K. Himmelreicher***

** Leiter des Forschungsdatenzentrums der Rentenversicherung, Würzburg*

*** Leiter des Forschungsdatenzentrums der Rentenversicherung, Berlin*

Etablierung und Weiterentwicklung

In den Jahren seit der Gründung des Forschungsdatenzentrums der Rentenversicherung (FDZ-RV) im Jahr 2004 hat sich nicht nur in der gesetzlichen Rentenversicherung vieles ereignet. Auch das FDZ-RV hat sich in seinen Strukturen und Inhalten weiterentwickelt. Ziel der ersten Förderphase bis Ende 2006 war es, ein Basisangebot an Daten für die Wissenschaft in Form von Scientific Use Files und Public Use Files aufzubauen. Auf Grundlage einer Forscherbefragung (*Himmelreicher 2004*) lag der erste Schwerpunkt hierbei auf der Erschließung der Daten zum Renten- und Verrentungsgeschehen, dem Rentenbestand, dem Rentenzugang und dem Rentenwegfall. Diese Ziele wurden erreicht und das Ergebnis der ersten Datenerschließungen wurde von einer vom Rat für Sozial- und Wirtschaftsdaten¹ eingesetzten wissenschaftlichen Evaluationsgruppe ausdrücklich gewürdigt (*Rehfeld 2006*).

Nachdem der Zug im Jahr 2006 mit dem Ausbau des Datenangebots und des Datenservices weiter Fahrt aufgenommen hatte, sind seit 2007 die Weichen auf den Kurs Weiterentwicklung und Etablierung gestellt. Besondere Innovationen sind die erste Erschließung der Längsschnittdaten, die Erarbeitung von Konzepten für Regionalauswertungen und die Stärkung der wissenschaftlichen Arbeit des FDZ-RV u. a. durch die Schaffung befristeter Stellen für wissenschaftliche Mitarbeiter.

Unter diesem Zeichen stand auch der nunmehr vierte Workshop, der in Kooperation mit dem Wissenschaftszentrum Berlin für Sozialforschung (WZB) durchgeführt wurde. Zahlreiche Wissenschaftler und Fachexperten der Rentenversicherung diskutierten Forschungsergebnisse zur sozialen Sicherung, die auf Basis der Daten des FDZ-RV gewonnen worden waren. Über den Stand der Datenerschließung und die geplanten Neuerschließungen sowie über die Zugangswege informiert der vorliegende Beitrag.

Mikrodaten der gesetzlichen Rentenversicherung

Die Erschließung der Daten der Rentenversicherung findet in einem festgelegten Rahmen statt. Um zu verstehen, welche Daten grundsätzlich erhoben werden und damit auch zur Verfügung gestellt werden können, ist es wichtig die Bedingungen der Datenentstehung und deren Ziele zu benennen.

¹ Zum Rat für Sozial- und Wirtschaftsdaten siehe www.ratswd.de.

Bei den Mikrodaten der gesetzlichen Rentenversicherung handelt es sich um Informationen, die im Verwaltungsprozess erhoben werden; sie werden im Folgenden prozessproduzierte Daten genannt. Wichtig dabei ist allerdings, dass diese statistischen Daten nicht direkte Prozessdaten sind, sondern nach Absprache zwischen den Trägern der Rentenversicherung festgelegt werden (Luckert 2004). Welche Inhalte insbesondere für das Statistische Berichtswesen gemeldet werden, ist in einer Verwaltungsvorschrift (RSVwV)² nach § 79 SGB IV geregelt. Wesentliche Richtschnur sind dabei die Prämissen der Datensparsamkeit, des Datenschutzes und der Wirtschaftlichkeit. Die Rentenversicherung ist eingebunden in das Korsett der RSVwV und gehalten nur solche Daten zu erheben, die sie für ihr Berichterstattung und Planung auch tatsächlich benötigt.

Rahmenbedingungen durch die RSVwV

Die Statistiken der Deutschen Rentenversicherung sind als einheitliches Berichtswesen konzipiert, um alle erforderlichen Informationen zur sozialen Sicherung insbesondere des Sektors Alterssicherung für die Träger der Deutschen Rentenversicherung und ihre Vorstände, den Grundsatz- und Querschnittsbereich der Deutsche Rentenversicherung Bund sowie für Ministerien auf Bundes- und Länderebene zu erheben, zu sammeln, aufzubereiten, auszuwerten und zu archivieren (Bütefisch 2004: 20).

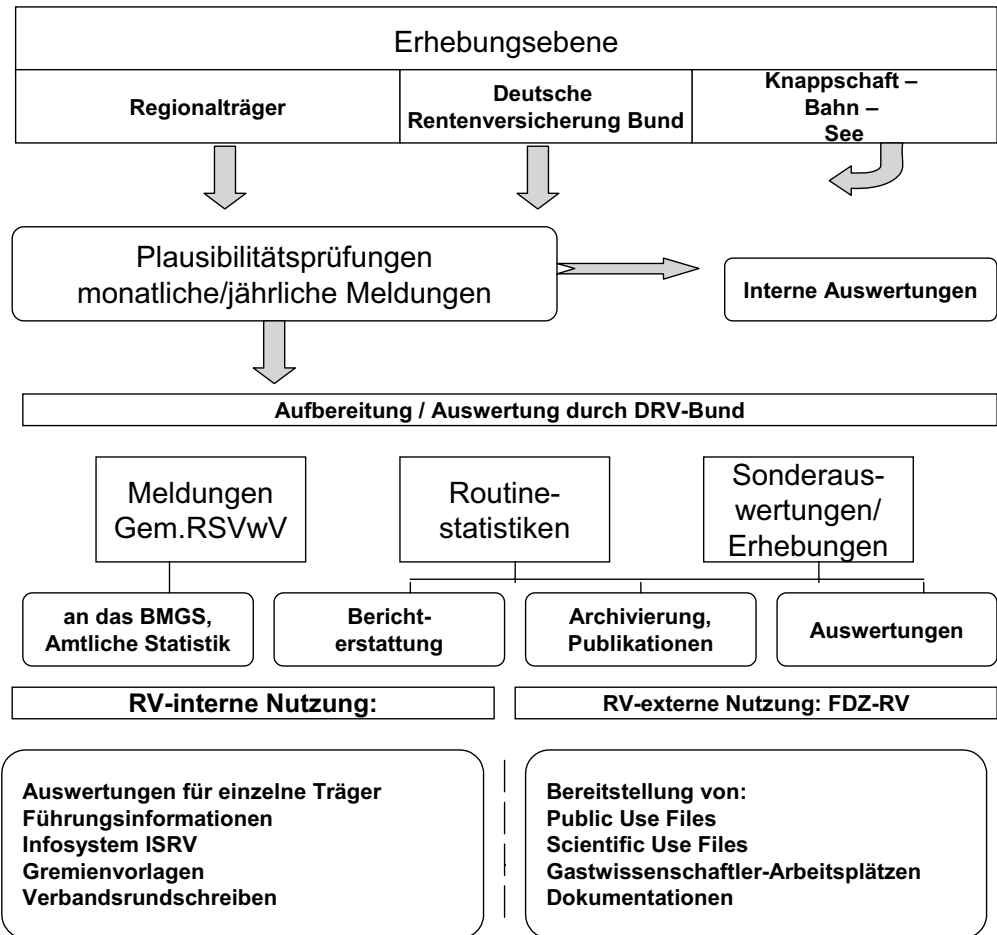
Methodik der Datenmeldung

Dies bedeutet, dass die Statistiken an Verwaltungsvorgänge bei den Versicherungsträgern gekoppelt sind und ihre Inhalte (Merkmale und deren Ausprägungen) abgeleitete Produkte der Verwaltungsdaten sind. Die Aufbereitung erfolgt in der Regel in Form anonymer Datensätze, die mit Extrakten aus den Verwaltungsvorgängen an den Grundsatz- und Querschnittsbereich der Deutschen Rentenversicherung Bund übermittelt werden (vgl. **Abbildung 1**). Zur Prüfung der Statistik-Daten werden einheitliche Plausibilitätsprogramme auf Trägerebene eingesetzt, um weitgehende formale Korrektheit und Auswertbarkeit der erhobenen Daten sicherzustellen. In der Regel werden, um zusätzlichen Aufwand zu vermeiden, die Statistiken als Vollerhebungen durchgeführt. Dieses Prinzip wird lediglich in jenen Fällen durchbrochen, in denen erhebliche Verarbeitungszeiten und eine Arbeitsbelastung auftreten und daher Stichprobenziehungen und anschließende Hochrechnungen wirtschaftlicher sind.

Die Methodik der Datenerhebung hat zwei entscheidende Konsequenzen für die Gestaltung von Scientific Use Files. Es handelt sich insbesondere bei den Querschnittsdatsätzen in aller Regel um Fallstatistiken und nicht um Personenstatistiken. Im Rentenbestand kann eine Person damit beispielsweise zweimal auftauchen, als Bezieher einer Versichertenrente (z. B. Altersrente) und als Bezieher einer Hinterbliebenenrente (z. B. Witwenrente). Die Anonymisierung der Daten bringt es mit sich, dass es bis auf eine Ausnahme (Statistik der Rentenzahlungen, siehe Abbildung 2) nicht möglich ist, diese Fall- in eine Personenstatistik zu überführen.

² Allgemeine Verwaltungsvorschrift über die Statistik in der Rentenversicherung.

Abbildung 1: Datenmeldung und Berichterstattung



Quelle: Bütefisch (2004), eigene Überarbeitung.

Das Datenangebot des FDZ-RV

Das FDZ-RV erschließt sukzessive die Daten der gesetzlichen Rentenversicherung aus den Bereichen Versicherung, Rente und Rehabilitation. Neben der Unterscheidung dieser drei wichtigen Themen des statistischen Berichtswesens ist bei der Konzeption des Datenangebots auch die Gliederung der einzelnen Statistiken nach bestimmten Rentenarten bzw. Personenkreisen von Bedeutung. Die Gestaltung des Datenangebots versucht dem Rechnung zu tragen und nutzerfreundliche und themenadäquate Daten für Wissenschaft und Forschung zur Verfügung zu stellen.

Grundlegende Philosophie und Zugangswege

Grundsätzlich stellt das FDZ-RV Daten über drei verschiedene Zugangswege zur Verfügung. Wichtig dabei ist, dass diese Angebote sich ergänzen. Die Scientific Use Files – Basisfiles bilden das tragende Gerüst der FDZ-Daten. Zu jedem Datensatz bzw. Datengebiet wird ein solcher Basisfile erschlossen, der ab der Erstauflage in jedem Folgejahr neu aufgelegt wird. Die Eingrenzung der Personengesamtheit und die Auswahl der Merkmale ist möglichst breit angelegt und hat zum Ziel, einen ersten Zugang zu den Daten zu finden. Aus Gründen des Datenschutzes werden sensible Merkmale in diesen Files relativ stark modifiziert (siehe dazu *Stegmann* u. a.: 2005). Ergänzend kommen sogenannte Themenfiles mit besonderen thematischen Schwerpunkten hinzu. Eine Reihe von Themenfiles wird in ähnlicher Weise jährlich aktualisiert. Hinzu kommen allerdings eine Fülle von Themenfiles, die mit den Wissenschaftlern abgesprochen und auf deren Fragestellung zugeschnitten werden können.

Die Zugangswege, die ein Forscher findet, hängen von der Fragestellung ab, sollten aber in der Regel ein mehrstufiges Verfahren bilden. Am Anfang sollte immer die Beschäftigung mit einem thematisch geeigneten Basisfile stehen, insbesondere dann, wenn auch noch ein thematischer Zugang zur sozialen Sicherung gefunden werden muss. Diese Auswertungen können ergänzt werden durch maßgeschneiderte Themenfiles, die entweder in der „Off-Site“-Nutzung dem Forscher zur Verfügung gestellt werden oder an einem der Gastwissenschaftler-Arbeitsplätze in Würzburg oder Berlin genutzt werden können. Welche Zugangsform gewählt wird, hängt davon ab, welchen Anonymisierungsgrad die Daten besitzen und davon, ob ein Beratungsbedarf durch das FDZ-RV besteht.

Mit dem kontrollierten Fernrechnen besteht ein dritter Weg der Datennutzung. Im aktuellen Stadium bietet er eine sinnvolle Ergänzung zu einem Gastwissenschaftleraufenthalt. Der Forscher kann geprüfte SPSS-Syntaxen oder STATA-DO-Files an das FDZ-RV übermitteln, wo die Berechnungen durchgeführt werden. Die geprüften Ergebnisse werden dann per E-Mail zugestellt.

Überblick über das Datenangebot

Einen ersten Überblick über das Datenangebot der Rentenversicherung im Hinblick auf die Wissenschaft mit den aktuell über das FDZ-RV verfügbaren Scientific Use Files gibt **Abbildung 2**. Unter den in den Spalten genannten Themenbereichen sind die jeweiligen Mikrodatensätze und ihre Eigenschaften, ob Quer- und/oder Längsschnittdaten zur Verfügung stehen, genannt.

Im Rahmen des Datenangebotes des FDZ-RV sind die Querschnittsdatsätze der Rentenzugangs- und Rentenbestandsstatistik sowie die Versichertenstatistik bereits für mehrere Berichtsjahre aufbereitet. Im Jahr 2006 wurde das Angebot um den Längsschnittdatensatz „Vollendete Versichertenleben 2004“ erweitert. Die aufwändige Entwicklung dieses ersten Datenangebots im Sektor der Längsschnittdaten wurde durch eine wissenschaftliche Testnutzung begleitet (*Stegmann* 2006).

Das erweiterte Datenangebot des FDZ-RV, das nun auch Längsschnittdaten umschließt, verbessert die Nutzungsmöglichkeiten für die Wissenschaft erheblich. So ist es auf Basis der Längsschnittdaten möglich, biografiebezogene Analysen durchzuführen und kohorten- bzw. gruppenvergleichende Entgeltprofile oder Quotenverläufe (vgl. *Fachinger* und *Himmelreicher*

Abbildung 2: Prozessproduzierte Mikrodaten der gesetzlichen Rentenversicherung in thematischer und methodischer Gliederung und deren Verfügbarkeit im FDZ-RV

| Prozessproduzierte Mikrodaten der gesetzlichen Rentenversicherung eines Jahrgangs in thematischer und methodischer Gliederung | | | | | |
|--|---|---|---|---|--|
| Gegenstand der Statistik | | | | | |
| Renten-zugangs-statistik | Renten-bestands-statistik | Rentenzahl-bestands-statistik | Versicherten-statistik | Reha-Statistik | Versorgungsausgleichs-statistik |
| Querschnittsdatensätze | | | | | |
| Rentenzugang Rentenwegfall Renten- änderung (Stichtag 31.12.) | Rentenzah- lungen aus den Träger- konten (Stichtag 31.12.) | Mehrfach- bezug von Renten- zahlungen (Stichtag 30.06) | Statistik der aktiv Versicherten (Stichtag 31.12. / Berichtszeit- raum) | Abgeschlos- sener Reha- Leistungen (Stichtag 31.12. / Berichtszeit- raum) | Versorgung- ausgleich (Stichtag 31.12.) |
| Längsschnittsdatensätze | | | | | |
| Vollendete Versicherten- leben | | | Versiche- rungskonten- stichprobe | RSD-Verlaufs- erhebung | |

= verfügbar im FDZ-RV, zum Teil für die Berichtsjahre 2003, 2004, 2005, 2006 (Stand: Januar 2008).

Quelle: Eigene Darstellung, in Anlehnung an *Himmelreicher und Radl (2006)*.

2006) zu untersuchen. Da die FDZ-RV-Daten mittlerweile in zahlreichen wissenschaftlichen Disziplinen von mehr als einhundertsechzig Wissenschaftlerinnen und Wissenschaftlern genutzt und zunehmend mehr Vorträge und Publikationen auf Basis dieser Daten erstellt werden, erhöht sich der Bekanntheitsgrad des Forschungsdatenzentrums zusehends. Das aktuelle Datenangebot sowie weitere Informationen über und um das Forschungsdatenzentrum sind unter www.fdz-rv.de zu finden.

Neuentwicklungen und Ausblick

Die zweite Förderphase des FDZ-RV beinhaltet vor allem den Ausbau des Datenangebots von Längsschnittdaten und den Aufbau eines ersten Datenangebots im Themengebiet Rehabilitation. Weitere Bereiche sind die Entwicklung von Regionalmerkmalen, die in die Scientific Use Files einfließen sollen und die Optimierung des Online-Angebots durch verbesserte Informations- und Recherchemöglichkeiten (*Dippold 2006*).

Im Laufe des Jahres 2008 wird die Versicherungskontenstichprobe (VSKT) als Scientific Use File veröffentlicht werden. Die VSKT ist eine ständige Sondererhebung der gesetzlichen Ren-

tenversicherung. Sie erfasst die Versicherten der deutschen gesetzlichen Rentenversicherung und den Stand ihrer Rentenanwartschaften und liefert Informationen über sämtliche (gespeicherten) rentenrelevanten Tatbestände. Da die komplette Versicherungsbiografie der Untersuchungspersonen abgebildet wird, wäre eine jährliche Untersuchung in Form einer Stichprobe nicht leistbar. Deswegen wurde für die Ersterhebung eine geschichteten Zufallsstichprobe aus den Versicherungskonten gezogen und in den Folgejahren als Panel fortgeführt und gepflegt. Diese Vorgehensweise wurde gewählt, um einerseits auf geklärte Konten zurückgreifen zu können, andererseits aber den Aufwand der Kontenklärung möglichst gering zu halten.

Die Erhebung umfasst alle Informationen zu rentenrechtlichen Zeiten und Beiträgen, die nach einer Kontenklärung im Rentenversicherungskonto der Person gespeichert sind. Dieser Ursprungsdatensatz ist variabel lang und enthält für einen Versicherten tagesgenau alle rentenrechtlichen Ereignisse, die sich in seiner Biografie finden. So lässt sich beispielsweise kontrollieren, wann eine Person sozialversicherungspflichtig erwerbstätig war oder wann Kinder erzo-gen wurden (vgl. *Stegmann 2007*).

Diese prozessproduzierten Längsschnittdaten der Rentenversicherung sind von besonderem Nutzen, da erwerbsbiografische Verläufe analysiert werden können und die typischen Probleme von Längsschnitterhebungen der Sozialforschung vermieden werden. Panelmortalität und Erinnerungfehler wie bei retrospektiven Befragungen können nicht auftreten. Bei den Daten, die aus dem geklärten Rentenversicherungskonto stammen, handelt es sich um Meldedaten der Sozialversicherung, um Daten über andere Beitragszeiten und um nachgewiesene und geprüfte Daten zu beitragsfreien Zeiten. Damit wird eine hohe Datenqualität und Genauigkeit erreicht, die von Befragungen nicht erzielt werden können (vgl. *Stegmann 2006*).

Im Bereich der Rehabilitation wird im Laufe des Jahres 2008 der erste Scientific Use File veröffentlicht werden. Dieser wird nach dem Stand der Planungen Auswertungen zu den im Berichtsjahr abgeschlossenen Rehabilitationsleistungen ermöglichen. Eingeschlossen sind hierbei die medizinische Rehabilitation und die Leistungen zur Teilhabe am Arbeitsleben.

Bei der Neuauflage der Basisfiles für das Berichtsjahr 2007 wird das neue Regionalisierungskonzept Berücksichtigung finden. Ausgewählte Scientific Use Files werden mit Regionalinformationen angereichert werden, die eine sozio-geografische Typisierung beinhalten. Auf diese Weise wird es möglich sein, Angaben zum Lebensraum in die Analysen zu integrieren, die kleinräumig auf Basis des Wohnortes (Landkreis, kreisfreie Stadt) gewonnen wurden.

Zusammenfassend konnte das Forschungsdatenzentrum der Rentenversicherung in den ersten Jahren nach seiner Gründung wesentliche Mikrodatensätze der Rentenversicherung aufbereiten und nutzerfreundlich kommentieren. Diese Datensätze wurden inzwischen zu einem wichtigen Bestandteil quantitativer Forschungsarbeiten vor allem in Soziologie, Ökonomie und Demografie. Durch die Erschließung der Reha-Daten sowie der Versicherungskontenstichprobe können sicherlich weitere Forschergruppen angesprochen werden.

Literatur

- Bütefisch, Th.* (2004): Datenwege und praktischer Datenzugang, Band 55/2004, S. 24-41. WDV: Bad Homburg.
- Dippold, C.* (2006): Wissensverwaltung und Wissensrecherche durch Integration eines Metadatenkonzepts im FDZ-RV. In: Deutsche Rentenversicherung Heft 9-10/2006, S. 622-640. WDV: Bad Homburg.
- Fachinger, U. & Himmelreicher, R. K.* (2006): Die Bedeutung des Scientific Use Files Vollendete Versichertenleben 2004 (SUFVVL2004) aus der Perspektive der Ökonomik. In: Deutsche Rentenversicherung Heft 9-10/2006, S. 562-582. WDV: Bad Homburg.
- Himmelreicher, R.* (2004): Daten der Rentenversicherung und Bedarf der Wissenschaft. Befunde einer Befragung unter potentiellen Nutzern des Forschungsdatenzentrums der gesetzlichen Rentenversicherung. Deutsche Rentenversicherung 8/2004, S. 472-485.
- Himmelreicher, R. K. & Radl, J.* (2006): Zusammenfassung und Ausblick auf die weitere Entwicklung des FDZ-RV. In: Forschungsrelevante Daten der Rentenversicherung. Bericht vom zweiten Workshop des Forschungsdatenzentrums der Rentenversicherung vom 27. bis 29. Juni 2005 in Würzburg. DRV-Schriften Band 55/2005, S. 281-286. WDV: Bad Homburg.
- Luckert, H.* (2004): Statistikdaten der gesetzlichen Rentenversicherung – ein grober Überblick. In: DRV-Schriften, Band 55/2004, S. 24-41. WDV: Bad Homburg.
- Rehfeld, U.* (2006): Aktueller Stand und weitere Perspektiven im Forschungsdatenzentrum der Rentenversicherung. In: DRV-Schriften, Band 55/2006, S. 13-22. WDV: Bad Homburg.
- Stegmann, M., Luckert, H. & Mika, T.* (2005): Die Bereitstellung prozessproduzierter Daten im Forschungsdatenzentrum der Rentenversicherung. Grundsätze zur faktischen Anonymisierung von Mikrodaten und zu Gastwissenschaftler-Arbeitsplätzen. In: Deutsche Rentenversicherung 2-3/2005, S. 203-215. WDV: Bad Homburg.
- Stegmann, M.* (2006): Aufbereitung der Sondererhebung „Vollendete Versichertenleben 2004“ als Scientific Use File für das FDZ-RV. In: Deutsche Rentenversicherung Heft 9-10/2006, S. 537-553. WDV: Bad Homburg.
- Stegmann, M.* (2007): Biografiedaten der Rentenversicherung und die Aufbereitung der Sondererhebung „Vollendete Versichertenleben 2004“ als Scientific Use File. DRV-Schriften Band 55/2006, S. 82-95.

Dr. Ralf K. Himmelreicher, Diplom-Soziologe, absolvierte eine Ausbildung zum Elektriker und studierte Soziologie, Volkswirtschaftslehre und Statistik in Berlin und Bremen. Zurzeit leitet er das Forschungsdatenzentrum der Rentenversicherung am Standort Berlin und ist Lehrbeauftragter am Institut für Soziologie der Freien Universität Berlin.

Dr. Michael Stegmann, M.A. Soziologe, absolvierte eine Ausbildung zum Rettungsassistenten und studierte Soziologie, Psychologie und Pädagogik in Würzburg und Frankfurt am Main. Beim VDR war er bis 2004 in der Abteilung Volkswirtschaft und Statistik im Referat Altersvorsorge in Deutschland (AVID) und Sondererhebungen tätig. Zurzeit leitet er das Forschungsdatenzentrum der Rentenversicherung in Würzburg.