

Verwendung von Zensus-
Paradaten unter
besonderer
Berücksichtigung von
Namensverteilungen

Verwendung von Zensus-Paradaten unter besonderer Berücksichtigung von Namensverteilungen

Rainer Schnell

Universität Duisburg-Essen
20. Dezember 2017

Abstract: Bei der Durchführung von Zensen werden Daten durch den Erhebungs- und Verarbeitungsprozess selbst generiert. Solche Daten werden in der neueren Literatur als Paradaten bezeichnet. Paradaten sind für die statistische Beurteilung der Qualität jeder Datenerhebung unverzichtbar. Daneben werden durch einen Zensus Hilfsdateien erzeugt, die für zahlreiche Fragestellungen und administrative Aufgaben von eigenem Interesse sind, so z.B. Straßen- und Geburtsortverzeichnisse. Schließlich werden durch einen Zensus Namensverteilungen generiert, die sowohl für die Migrationsforschung als auch für die Verbesserung der Verfahren der amtlichen Statistik und anwendungsbezogenen Informatik von erheblicher Bedeutung sind. Dieses Papier beschreibt stichwortartig nur die aus Sicht des Verfassers wichtigsten Verwendungsmöglichkeiten dieser Daten.

Keywords: Zensus, Para-Daten, Record-Linkage, Namen, Onomastik

1 Vorbemerkung

Der Verfasser hat der Zensusgruppe des Statistischen Bundesamtes Vorschläge zur Dokumentation und der Hilfsdateien, die über normale Para-Daten hinaus gehen, unterbreitet. Im Folgenden werden einige solcher Merkmale, vor allem in Hinsicht auf Pseudo-Identifikatoren, beschrieben. Der Schwerpunkt liegt dabei auf Verwendungsmöglichkeiten solcher Merkmale, die über die direkte statistische Analyse hinausgehen.

Sollte der Gesetzgeber auf solche weitergehenden Analysen verzichten, werden nicht nur wertvolle Forschungsressourcen nicht genutzt, sondern in erheblichem Umfang finanzielle Einbußen durch erhöhten Aufwand in anderen administrativen Bereichen in Kauf genommen. Dies gilt nicht nur für die zukünftigen Erhebungen der amtlichen Statistik sondern für die alltäglichen Aufgaben der öffentlichen Verwaltung vom Gesundheitssystem bis hin zur öffentlichen Sicherheit.

2 Traditionelle Para-Daten

Der Vollständigkeit halber müssen traditionelle Para-Daten erwähnt werden, die für die Beurteilung jeder Datenerhebung aus Sicht der Survey-Methodology unentbehrlich sind (Kreuter 2013).

Dazu gehören zunächst die Dokumentation einer Interviewer-ID, ohne die keine Berechnung der Workloads und der Design-Effekte möglich sind (Wolter 2007). Neben einem Merkmal, das Auskunft über die Art der Datenerhebung (Erhebungsmodus) gibt, gehört dies zu den unverzichtbaren Merkmalen jeder Datenerhebung.

Daneben werden für jeden Schritt der Datenerhebung (Wohnungseigentümer, Begehungen etc.) explizite Nonresponse-Records benötigt (Schnell 1997). Es muss erkennbar sein, ob zwischen der Kooperation der Eigentümer bzw. Bewohner und Gebäude- oder Bewohnermerkmalen systematische Zusammenhänge bestehen. Insbesondere muss erkennbar sein, wann die Zahl von Wohnungseinheiten geschätzt oder imputiert wurde. Idealerweise liegt dieses Merkmal für jedes Record vor.

Entsprechend werden für jeden Bearbeitungsschritt Imputationsflags benötigt, d.h. für jedes Feld muss erkennbar sein, ob dies erhoben, editiert oder imputiert wurde (de Waal, Pannekoek und Scholtus 2011). Selbstverständlich muss das Imputationsverfahren und alle dafür notwendigen Codierungen und Merkmale bekannt sein.

3 Programm-Codes

Für eine Beurteilung der Verarbeitung der erhobenen Daten ist ein Zugang zu allen Programmcodes für alle Gewichtungen, Imputationen und Record-Linkage-Operationen erforderlich. Dazu gehören insbesondere alle Details der Doublettenerkennung und Bereinigung sowie die technische Dokumentation der Haushaltegenerierung mit allen Codes, Edit-Regeln und Schwellenbestimmungen.¹

Ebenso ist eine Zählung der Anwendung der Edit- und Matchregeln erforderlich. Sollte die Anwendung der Record-Linkage-Verfahren wie üblich innerhalb von Blöcken (z.B. Häusern, Straßen oder Gemeinden) erfolgen, wäre eine Auswertung auf der kleinstmöglichen Blockebene sinnvoll. Damit könnten räumliche Einheiten identifiziert werden, bei denen entweder der Datenerhebungsprozess nicht hinreichend genau durchgeführt wurde oder die Merkmals-träger (Wohnungen oder Personen) besondere erhebungstechnische Herausforderungen darstellen.² Damit hätte man unmittelbare, kleinräumige Datenqualitätsindikatoren.

4 Hilfsmerkmale

Zwei wesentliche Hilfsdateien im Zensus sind der Straßenthesaurus und das Geburtsortverzeichnis. Darüber hinaus ergibt die Wohnungszählung interessante Hilfsmerkmale.

¹Die technische Dokumentation der Haushaltegenerierung allein wird mehrere hundert Seiten zusätzlich zum Programmcode erfordern (vgl. dagegen die Dokumentation zum Zensus 2011: Kreuzmair und Reisch 2013).

²Genau dieses Problem stellte sich beim Zensus 2011 in Berlin (Schnell 2013).

4.1 Straßentheseaurus

Ein Straßentheseaurus ist die Sammlung aller Adressendaten eines Landes. Für den Zensus 2011 wurde ein solcher Datenbestand aufgebaut (Kleber, Maldonado, Scheuregger und Ziprik 2009). Ein solcher Straßentheseaurus ist für alle Datenerfassungsroutinen personenbezogener Daten in Deutschland in Ämtern und Verwaltungen vom KBA bis zu den Krankenkassen von erheblicher praktischer Bedeutung. Es gibt keinen Grund, warum das Straßenverzeichnis des Zensus nicht öffentlich und unentgeltlich zugänglich sein sollte.

Besonders wertvoll würde dieser durch die Möglichkeiten der Georeferenzierung: Damit würde auch eine Historik der Straßennamen möglich. Mithilfe einer solchen Datei könnten all diejenigen Adressen identifiziert werden, die im Laufe der Jahre umbenannt werden. Derzeit ist dies nur durch lokale Recherchen möglich. Allein diese Möglichkeit würde bei Adressbereinigungen aller Art im Rahmen der öffentlichen Verwaltung langfristig zu erheblichen Einsparungen führen.

4.2 Geburtsortverzeichnis

Ebenso ist ein Geburtsortverzeichnis samt Häufigkeiten nach Geburtsjahr für zahlreiche praktische Fragestellungen ein kostenloses und sehr hilfreiches Nebenergebnis (Weißer 2012). Es gibt keinen Datenschutzgrund, der einer uneingeschränkten Veröffentlichung im Wege stünde. Auch hier besteht das wesentliche Anwendungsfeld aus Datenbereinigungsroutinen von Behörden und Verwaltungen allgemein. Schließlich sind diese Daten von nicht unerheblicher Bedeutung für alle Fragestellungen im Zusammenhang mit Bevölkerungsmigration, sowohl innerhalb Deutschlands als auch über die Staatsgrenzen hinweg.

4.3 Gebäudezählung

Im Rahmen der Gebäudezählung entstehen Hilfsdateien, die für die Durchführung zukünftiger amtlicher Erhebungen sinnvoll sind. Dazu gehören alle Gebäudezusatzbezeichnungen, die im Rahmen der Adressaufbereitungen bekannt werden. Diese Bezeichnungen sind für alle Adressbereinigungen von Interesse.

Das Gleiche gilt für eine Liste aller Sonderbereiche, wobei nur der Name und die Anschrift benötigt werden. Beschränkt man sich auf diejenigen Bereiche, die auch in Telefonbüchern verzeichnet sind, entsteht kein Datenschutzproblem. Der Vorteil dieser Datei besteht in der Unterstützung der Erzeugung von Sampling-Frames der empirischen Sozialforschung.

Ebenso hilfreich wäre eine Liste der Gebäude, die keinen Wohnraum enthalten (exakte Anschrift und GPS-Koordinaten). Auch hier entsteht kein Datenschutzproblem und auch diese Datei ist für weitere Aufgaben der Stichprobenziehung hilfreich.

5 Homogenitätsmaße

Über die kleinräumige Verteilung von Populationsmerkmalen ist in Deutschland wenig bekannt. Dies gilt nicht nur für die kleinräumige Berechnung von Mittelwerten und Anteilen,

sondern vor allem für die räumliche Homogenität von Merkmalen. Alters-, Einkommens- und Migrantensegregation sind in den Sozialwissenschaften bekannte Phänomene, die für die Entwicklung von Stadtteilen, Kriminalität und ethnischen Enklaven von erheblicher Bedeutung sind. Es wäre daher mehr als wünschenswert, wenn Analysen der kleinräumigen Homogenität für alle Zensusmerkmale berechenbar blieben und für wissenschaftliche Analysen zur Verfügung gestellt würden. Einzelheiten finden sich in einer Projektskizze zur Zensusbegleitforschung 2011 (Schnell 2012).

6 Namen als inhaltlicher Forschungsgegenstand

Der Zensus bietet über die Erhebung von Vor- und Nachnamen gemeinsam mit regionalen Identifikatoren und inhaltlichen Merkmalen eine einzigartige Datenbasis sowohl für die Grundlagenforschung als auch für praktische Anwendungen im Bereich der Bevölkerungsstatistik, Sozialwissenschaften und Informatik.

6.1 Migrationshintergrund und onomastische Klassifikationen

Die Bestimmung des Migrationshintergrunds spielt für die amtliche Statistik (VDSt Verband Deutscher Städtestatistiker 2013) und die empirische Sozialforschung im Allgemeinen eine zunehmend wichtigere Rolle. Dafür werden neben der Staatsangehörigkeit, dem Geburtsort und direkten Fragen zunehmend namensbasierte Verfahren immer wichtiger.

Namensbasierte Verfahren zur Bestimmung des Migrationshintergrunds sind in Deutschland vor allem mit den Arbeiten von Humpert und Schneiderheinze (2000) verknüpft. Es existieren allerdings nur wenige Studien zur prinzipiellen Qualität, dem Deckungsgrad der verwendeten Dateien und eventuellen Verzerrungen (Mateos 2007; Schnell, Trappmann und Gramlich 2014).

Die Daten des Zensus bieten hier eine einmalige Gelegenheit, die Vollständigkeit und Güte konkurrierender Verfahren der Bestimmung des Migrationsstatus zu untersuchen. Dazu gehört z.B. die Prüfung, ob die existierenden Datenbestände onomastischer Verfahren alle tatsächlich vorkommenden Namen abdecken oder welche Verzerrungen durch unvollständige Abdeckungen hervorgerufen werden.

Solche Untersuchungen könnten in den kontrollierten Datenverarbeitungsgebungen des Statistischen Bundesamtes durchgeführt werden. Vermutlich wäre ein solcher Abgleich auch – unter erheblich erschwerten Bedingungen – durch den Einsatz fehlertolerant verschlüsselter Identifikatoren durch Bloom-Filter (Schnell, Bachteler und Reiher 2009) möglich.

6.2 Namensbasierte Forschung in den Sozialwissenschaften im Allgemeinen

In den Sozialwissenschaften wurde die Forschung auf der Basis von kulturspezifischen Vornamenswahlen von Lieberson begründet (Lieberson 2000) und in Deutschland durch viele andere – häufig unter Verzicht auf den Hinweis auf den Ursprung der Idee – popularisiert

(Gerhards 2003). Seit der Publikation von Lieberson werden Vor- aber auch Nachnamen (sowie deren Kombination untereinander als auch mit anderen Merkmalen wie Bildung oder Migrationsstatus) für zahlreiche fachspezifische Anwendungen benutzt.

Die Anwendungen reichen dabei von der Geographie und Soziologie (Mateos 2014) über die Sozialökonomie (Clark 2014; Collado, Ortuno-Ortin und Romeu 2012) bis hin zur Biologie. Diese zunächst verblüffende Anwendung basiert darauf, dass Nachnamen in vielen Gesellschaften über den Vater – ähnlich Teilen des Y-Chromosoms – vererbt werden. Daher sollten sich Zusammenhänge zwischen Nachnamen und speziellen DNA-Sequenzen zeigen. Dies wurde tatsächlich empirisch nachgewiesen (King und Jobling 2009) und hat zu zahlreichen neuen Arbeiten auf der Basis von Namensverteilungen geführt (Rossi 2013; Boattini und Pettener 2013).

Bei einer Anwendung in Deutschland wäre schon die regionale Namensverteilung für viele inhaltliche und praktische Zwecke von Interesse. Da sich Nachnamensverteilungen nur langsam verändern, wäre hier eine Verteilung nach Geburtskohorten und Region für Forschungen der regionalen Mobilität über Generationen hinweg von Bedeutung (McElduff, Mateos, Wade und Borja 2008). Für viele dieser Analysen benötigt man kaum mehr als die Häufigkeiten von Vor- und Nachname, eventuell Namen vor der Heirat und deren Übergangshäufigkeiten, idealerweise nach Alter, Geschlecht, Region und Migrationsstatus. Es muss darauf hingewiesen werden, dass nahezu alle Statistikämter in Großbritannien, den skandinavischen Ländern und den USA zumindest einfache Häufigkeitsauszählungen auf ihren Homepages kostenlos zum Download anbieten. In Deutschland existiert für diese Art der Forschung bislang keine geeignete Datenbasis.

6.3 Familiensoziologie und Assimilationsforschung

Andere Beispiele für die Verwendung von Namen im Zusammenhang mit anderen Merkmalen finden sich in der Familiensoziologie (Nave-Herz 1999). Die Scheidungswahrscheinlichkeiten in Abhängigkeit von der Namenswahl nach einer Heirat wurde in Deutschland mangels geeigneter Daten bislang nicht untersucht. Ähnliches gilt für die Häufigkeit und Stabilität interethnischer Ehen (Haug 2011). Familienstand und Häufigkeiten (eventuell bereits onomastisch klassifizierter) Vor- und Nachnamen nach Altersgruppe und Kohorte würden wertvolle Daten für diese Forschung liefern. Ließe sich dies mit der Namensgebung für die Kinder verbinden (Becker 2009), hätte man weitere Assimilationsindikatoren, die über reine Befragungsdaten zu Einstellungen weit hinausgingen.

7 Die Rolle von Namen bei der Verknüpfung mehrerer Datenbanken: Record-Linkage

Da es in Deutschland keine einheitliche Personenkennziffer gibt, erfolgt die Identifikation von Personen in fast allen Kontexten (Krankenhäuser, Verwaltung des Gesundheitswesens, öffentliche Verwaltung etc.) über Vor- und Nachnamen zusammen mit Geburtsdatum, Adresse und Geburtsort.

Die Literatur zur Datenzusammenführung von Datenbanken mit Namen betont den hohen Fehleranteil in Namen durch Tippfehler, verschiedene Schreibweisen des gleichen Namens sowie veränderte Zusammensetzung von Namensbestandteilen (Winkler 2009). Für den Umgang mit solchen Fehlern gibt es spezielle Verfahren in der Statistik bzw. Informatik, die in der Regel als „Record-Linkage-Verfahren“ bezeichnet werden (Christen 2012). Am häufigsten wird hierbei das sogenannte Fellegi-Sunter-Modell des probabilistischen Record-Linkages verwendet (Fellegi und Sunter 1969). Diese Verfahren benötigen aber Informationen über die Häufigkeiten von Namen in der Population.

In der Regel basieren diese Verfahren auf großen Stichproben, allerdings werden die Verfahren umso besser, je zuverlässiger auch die Häufigkeit seltener Namen geschätzt werden können. Das gilt insbesondere dann, wenn die Namen nicht auf einer homogenen Population stammen, sondern Subpopulationen mit völlig unterschiedlichen Namenszusammensetzungen existieren. Dies sind die Bedingungen, die durch die Migrationsbewegungen der letzten Jahre in Deutschland gelten.

Für die korrekte Durchführung von Datenzusammenführungen auf der Basis von Namensdaten sind korrekte Informationen über die Häufigkeit auch seltener Namen in der Population daher von großer praktischer Bedeutung nicht nur für die amtliche Statistik, sondern für alle Anwendungen bei denen Datenverknüpfungen verwendet werden. Um den zentralen Punkt hier zu betonen: Hier sind gerade Fälle mit sehr geringer Häufigkeit ($n = 1$) bedeutsam, da die Korrektur eines Namens die Kenntnis einer amtlich registrierten Schreibweise voraussetzt. Hier kann aber auf weitere Kovariate verzichtet werden.

Die Anwendung solcher Daten liegt neben dem Record-Linkage in der amtlichen Statistik und der Forschung in der Adressbereinigung allgemein. Diese ist von kaum zu überschätzender praktischer Bedeutung für die öffentliche Verwaltung. Entsprechend sind z.B. Tabellen der Übergangswahrscheinlichkeiten von einer Ausgangsschreibweise auf die dann im Zensus verwendete Schreibweise von großem praktischen Interesse. Das gilt auch für dazugehörige Tabellen der Distanzmatrizen (der Levenshtein-Distanz, Christen 2012).

Der Zensus bietet darüber hinaus eine einzigartige Möglichkeit, technische Datenschutzverfahren für die Verknüpfung personenbezogener Daten zu testen. Damit könnte bei Erhebungen der amtlichen Statistik zukünftig mit verschlüsselten Identifikatoren anstelle von Klarnamen gearbeitet werden.

In der technischen Literatur werden diese Verfahren als ‘Privacy Preserving Record Linkage’ (PPRL) bezeichnet (Vatsalan, Christen und Verykios 2013). Hierzu existieren mehrere Verfahren, wobei das von der Arbeitsgruppe des Verfassers entwickelte Verfahren auf der Basis von Bloom-Filtern (Schnell et al. 2009) in der Informatik derzeit als ‘de-facto standard’ bezeichnet wird (Smith 2017). Ein Test von PPRL-Verfahren im Rahmen des Zensus bei dem die Ergebnisse der Haushaltgenerierung und Doublettenbereinigung mit Klarnamen mit den Ergebnissen von PPRL-Verfahren verglichen werden, könnte die Datenschutzdiskussion bei zukünftigen registerbasierten Zensen deutlich vereinfachen, da das Gebot der Europäischen Datenschutzrichtlinie nach weitgehender Nutzung von Pseudonymisierungsverfahren (Council of the European Union 2016) befolgt würde.

Literatur

- Becker, B. (2009). Immigrants' emotional identification with the host society: The example of Turkish parents' naming practices in Germany. *Ethnicities*, 9(2), 200–225.
- Boattini, A. & Pettener, D. (2013). What can we do with surnames today? An anthropologist point of view: Comment on "Surname distribution in population genetics and in statistical physics" by Paolo Rossi. *Physics of Life Reviews*, 10(4), 418–419.
- Christen, P. (2012). *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection*. Data-Centric Systems and Applications. Berlin: Springer.
- Clark, G. (2014). *The son also rises: surnames and the history of social mobility*. Princeton University Press.
- Collado, M. D., Ortuno-Ortin, I. & Romeu, A. (2012). Intergenerational linkages in consumption patterns and the geographical distribution of surnames. *Regional Science and Urban Economics*, 42(1), 341–350.
- Council of the European Union. (2016). Council regulation (EU) no 679/2016: On the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).
- de Waal, T., Pannekoek, J. & Scholtus, S. (2011). *Handbook of Statistical Data Editing and Imputation*. Hoboken: John Wiley & Sons.
- Fellegi, I. P. & Sunter, A. B. (1969). A Theory for Record Linkage. *Journal of the American Statistical Association*, 64(328), 1183–1210.
- Gerhards, J. (2003). *Die Moderne und ihre Vornamen*. Springer.
- Haug, S. (2011). Binationale, interethnische und interreligiöse Ehen in Deutschland. *Familie, Partnerschaft, Recht*, 10, 417–422.
- Humpert, A. & Schneiderheinze, K. (2000). Stichprobenziehung für Telefonische Zuwandererumfragen. *ZUMA-Nachrichten*, 47(24), 36–64.
- King, T. E. & Jobling, M. A. (2009). What's in a name? Y chromosomes, surnames and the genetic genealogy revolution. *Trends in Genetics*, 25(8), 351–360.
- Kleber, B., Maldonado, A., Scheuregger, D. & Ziprik, K. (2009). Aufbau des Anschriften- und Gebäuderegisters für den Zensus 2011. *Wirtschaft und Statistik*, 7, 629–640.

- Kreuter, F. (Hrsg.). (2013). *Improving Surveys with Paradata: Analytic Uses of Process Information*. Hoboken: Wiley.
- Kreuzmair, I. & Reisch, M. (2013). Zensus 2011: Ablauf der Haushaltegenerierung. *Statistisches Monatsheft Baden-Württemberg*, 61(4), 13–21.
- Lieberson, S. (2000). *A matter of taste: How names, fashions, and culture change*. Yale University Press.
- Mateos, P. (2007). A review of name-based ethnicity classification methods and their potential in population studies. *Population, Space and Place*, 13(4), 243–263.
- Mateos, P. (2014). *Names, Ethnicity and Populations: Tracing Identity in Space*. Berlin: Springer.
- McElduff, F., Mateos, P., Wade, A. & Borja, M. C. (2008). What's in a name? The frequency and geographic distributions of UK surnames. *Significance*, 5(4), 189–192.
- Nave-Herz, R. (1999). Auswirkungen des neuen Namensrechts. In W. Glatzer & I. Ostner (Hrsg.), *Deutschland im Wandel: Sozialstrukturelle Analysen* (S. 265–274). Opladen: Leske und Budrich.
- Rossi, P. (2013). Surname distribution in population genetics and in statistical physics. *Physics of Life Reviews*, 10(4), 395–415.
- Schnell, R. (1997). *Nonresponse in Bevölkerungsumfragen: Ausmaß, Entwicklung und Ursachen*. Opladen: Leske+Budrich.
- Schnell, R. (2012). Projektskizzen zur Zensusbegleitforschung: Pseudonymisierung der Hilfsmerkmale, Vorprodukte der Regionalisierung, Größe geographischer Mikrozellen und Datenqualitätsindikatoren. Diskussionspapier der Gruppe 'Universitäre Zensusforschung', Universität Duisburg-Essen, Research Methodology Group.
- Schnell, R. (2013). *Derzeitige Dokumentationslücken im Zensus 2011* (Techn. Ber. Nr. 2013-1; Gutachten im Auftrag der Berliner Senatsverwaltung). Methodology Research Group, Universität Duisburg-Essen. Duisburg.
- Schnell, R., Bachteler, T. & Reiher, J. (2009). Privacy-Preserving Record Linkage Using Bloom Filters. *BMC Medical Informatics and Decision Making*, 9(41), 1–11.
- Schnell, R., Trappmann, M. & Gramlich, T. (2014). A Study of Assimilation Bias in Name-Based Sampling of Migrants. *Journal of Official Statistics*, 30(2), 231–249.
- Smith, D. (2017). Secure Pseudonymisation for Privacy-preserving Probabilistic Record Linkage. *Journal of Information Security and Applications*, 34, 271–279.

Literatur

- Vatsalan, D., Christen, P. & Verykios, V. S. (2013). A Taxonomy of Privacy-Preserving Record Linkage Techniques. *Information Systems*, 38(6), 946–969.
- VDSSt Verband Deutscher Städtestatistiker. (2013). Migrationshintergrund in der Statistik. Definitionen, Erfassung und Vergleichbarkeit. Materialien zur Bevölkerungsstatistik, Heft 2. Köln: VDSSt, 15, 2016.
- Weißer, C. (2012). Auswertung der Geburtsorte im Zensus 2011. *Wirtschaft und Statistik*, 10, 843–855.
- Winkler, W. E. (2009). Record Linkage. In D. Pfeffermann & C. Rao (Hrsg.), *Handbook of Statistics 29A, Sample Surveys: Design, Methods and Applications* (S. 351–380). Amsterdam: Elsevier, North-Holland.
- Wolter, K. M. (2007). *Introduction to Variance Estimation* (2. Aufl.). New York: Springer.

IMPRINT

Publisher

German Record-Linkage Center
Regensburger Str. 100
D-90478 Nuremberg

Editors

Stefan Bender, Rainer Schnell

Template layout

Christine Weidmann

All rights reserved

Reproduction and distribution in any form, also in parts,
requires the permission of the German Record-Linkage Center

Download

www.record-linkage.de

**The German Record Linkage Center is funded
by the German Research Foundation (DFG).**