



Rat für Sozial- und
WirtschaftsDaten (RatSWD)

www.ratswd.de

RatSWD

Working Paper Series

Working Paper

Nr. 130

Zehn Jahre European Data Watch

Dokumentation von Datensätzen für die empirische Wirtschafts- und Sozialforschung und Zugangswege zu den Daten

Joachim Wagner

Dezember 2009

GEFÖRDERT VOM

Working Paper Series des Rates für Sozial- und Wirtschaftsdaten (RatSWD)

Die *RatSWD Working Papers* Reihe startete Ende 2007. Seit 2009 werden in dieser Publikationsreihe nur noch konzeptionelle und historische Arbeiten, die sich mit der Gestaltung der statistischen Infrastruktur und der Forschungsinfrastruktur in den Sozial-, Verhaltens- und Wirtschaftswissenschaften beschäftigen, publiziert. Dies sind insbesondere Papiere zur Gestaltung der Amtlichen Statistik, der Ressortforschung und der akademisch getragenen Forschungsinfrastruktur sowie Beiträge, die Arbeit des RatSWD selbst betreffend. Auch Papiere, die sich auf die oben genannten Bereiche außerhalb Deutschlands und auf supranationale Aspekte beziehen, sind besonders willkommen.

RatSWD Working Papers sind nicht-exklusiv, d. h. einer Veröffentlichung an anderen Orten steht nichts im Wege. Alle Arbeiten können und sollen auch in fachlich, institutionell und örtlich spezialisierten Reihen erscheinen. Die *RatSWD Working Papers* können nicht über den Buchhandel, sondern nur online über den RatSWD bezogen werden.

Um nicht deutsch sprechenden Nutzer/innen die Arbeit mit der neuen Reihe zu erleichtern, sind auf den englischen Internetseiten der *RatSWD Working Papers* nur die englischsprachigen Papers zu finden, auf den deutschen Seiten werden alle Nummern der Reihe chronologisch geordnet aufgelistet.

Einige ursprünglich in der *RatSWD Working Papers* Reihe erschienen empirischen Forschungsarbeiten, sind ab 2009 in der RatSWD Research Notes Reihe zu finden.

Die Inhalte der *RatSWD Working Papers* stellen ausdrücklich die Meinung der jeweiligen Autor/innen dar und nicht die des RatSWD.

Herausgeber der RatSWD Working Paper Series:

Vorsitzender des RatSWD (2007/ 2008 Heike Solga; 2009 Gert G. Wagner)

Geschäftsführer des RatSWD (Denis Huschka)

Zehn Jahre European Data Watch

**Dokumentation von Datensätzen für die empirische Wirtschafts- und Sozialforschung
und Zugangswege zu den Daten**

Joachim Wagner¹

Leuphana Universität Lüneburg (wagner[at]leuphana.de)

Zusammenfassung

Dieser Beitrag stellt die seit 2000 in *Schmollers Jahrbuch / Journal of Applied Social Science Studies* erscheinende Artikelserie *European Data Watch* vor. Die Beiträge zu dieser Serie stellen Datensätze für Personen und Firmen vor, die von empirisch arbeitenden Wirtschafts- und Sozialwissenschaftlern genutzt werden können, und zeigen Zugangswege zu diesen Daten auf.

Stichwörter: Forschungsdatenzentren, Datenservicezentren, Personen- und Haushaltsdaten, Betriebs- und Unternehmensdaten

Abstract

This paper is a portrait of the series *European Data Watch* that is published in *Schmollers Jahrbuch / Journal of Applied Social Science Studies* since 2000. Contributions to this series describe data for persons or firms that can be used in empirical research in economics and the social sciences, and document how these data can be accessed.

Keywords: Research data centres, Data service centres, Individual data and household data, establishment data and enterprise data

JEL classification: C8

1 Danksagung: Im Namen aller, die von Beiträgen in der Data Watch Serie profitiert haben und noch davon profitieren werden, danke ich den Autorinnen und Autoren, die durch ihre Mitarbeit einen Beitrag dazu leisten, empirische Arbeiten mit Mikrodaten von Personen und Firmen einfacher möglich zu machen und damit zu fördern. Dank der finanziellen Förderung durch den Rat für Sozial- und Wirtschaftsdaten sind alle Beiträge der Serie kostenlos im Netz verfügbar unter <http://www.ratswd.de/publ/datawatch.php>.

1. Einleitung

Empirische Forschung in den Wirtschafts- und Sozialwissenschaften ist “datenhungrig”, und dieser Hunger wächst. In zunehmendem Umfang werden in empirischen Untersuchungen Mikrodaten für Personen und Haushalte bzw. Betriebe und Unternehmen verwendet. Fortschritte in den technischen Möglichkeiten der Arbeit mit solchen Mikrodaten in Form immer leistungsfähigerer (und preiswerterer!) Computer und Auswertungsprogramme tragen hierzu bei. Umfangreiche Stichproben aus vielen Beobachtungsperioden für Personen oder Firmen selbst zu erheben ist jedoch sehr zeitaufwendig und in der Regel prohibitiv teuer. Ein kostengünstiger Ausweg besteht in der Nutzung vorhandener Datensätze. Hierzu müssen zwei Bedingungen erfüllt sein:

- Die Daten müssen auch für externe Nutzer zugänglich sein.
- Potenzielle Nutzer müssen wissen, dass die Daten die benötigten Informationen enthalten und dass sie von ihnen genutzt werden können.

Die erste Bedingung ist gerade bei den hier angesprochenen Mikrodaten in vielen Fällen nicht erfüllt. Gründe hierfür können Bestimmungen des Datenschutzes sein. Oft liegt dies aber auch an der fehlenden Möglichkeit (oder der mangelnden Bereitschaft) der Datenbesitzer zur Bereitstellung eines leichten Zugangs zu diesen Mikrodaten. In den vergangenen Jahren sind in Deutschland große Fortschritte zu verzeichnen. Wichtige Anstöße lieferten das Gutachten der *Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik* (2001) und die Arbeit des vom Bundesministerium für Bildung und Forschung (BMBF) berufenen und finanzierten *Rates für Sozial- und Wirtschaftsdaten* (RatSWD) (siehe www.ratswd.de). Heute verfügen wir in Deutschland über eine Infrastruktur aus Forschungsdatenzentren (FDZ) bei den wichtigen Datenproduzenten und Datenservicezentren (DSZ), die den Zugang von Wissenschaftlern gerade auch zu vertraulichen Mikrodaten kostengünstig und ohne großen bürokratischen Aufwand möglich macht (vgl. Kaiser und Wagner (2008) für eine Überblick).

Zur besseren Erfüllung der zweiten Bedingung will das Projekt *European Data Watch* beitragen². Seit 2000 erscheinen in *Schmollers Jahrbuch / Journal of Applied Social Science Studies* Beiträge, in denen Datensätze für Personen und Firmen vorgestellt werden, die von empirisch arbeitenden Wirtschafts- und Sozialwissenschaftlern genutzt werden können, und es werden die Zugangswege zu diesen Daten aufgezeigt. Die Verfasser der Beiträge haben oft

² Der Titel dieses Projekts lehnt sich bewusst an die Data Watch Serie an, die seit vielen Jahren unregelmäßig im *Journal of Economic Perspectives* erscheint.

die Daten selbst mit erhoben bzw. aufbereitet oder sie sind durch ihre Arbeiten mit den Datensätzen ausgewiesene Experten hierfür. Die Beiträge behandeln für jeden Datensatz folgende Aspekte:

- Welche Informationen enthalten die Daten? Wo sind sie dokumentiert?
- Welche Themen wurden in (ausgewählten) Studien mit den Daten behandelt?
- Wie sind die Daten für externe Forscher zugänglich?
- An wen müssen sich Interessenten wenden?
- Wie hoch sind die Kosten der Datennutzung?
- Welche Möglichkeiten gibt es, die Daten in Lehrveranstaltungen zu nutzen?

Für die Forschung können die Beiträge in der Serie wichtige Impulse liefern. Sie können Hinweise geben auf geeignete Datensätze für empirische Studien – insbesondere auch dann, wenn die Ergebnisse, die unter Verwendung eines Datensatzes erzielt wurden, einer Robustheitsprüfung mit weiteren Datensätzen unterzogen werden sollen. Hier können auch ältere Datensätze weiter helfen, die vielleicht schon als „ausgeforscht“ und damit zu Unrecht als uninteressant gelten. Darüber hinaus kann ein Beitrag auch den Anstoß geben, über ein empirisches Projekt überhaupt erst nachzudenken – wer weiß denn schon von all den Datens(ch)ätzen, die verborgen hinter dicken (realen oder juristischen) Mauern bei den Datenproduzenten und Datenbesitzern lagern?

Für die Lehre können die Beiträge Hinweise darauf geben, welche Daten für Veranstaltungen nutzbar sind und an Studierende weitergegeben werden dürfen. Schwimmen lernt man im See, Schlittenfahren im Schnee ... und Empirische Wirtschafts- und Sozialforschung am PC! „Learning by doing“ unter Verwendung von echten Mikrodaten mit all ihren Ecken und Kanten ist nach meiner Erfahrung durch keine Lehrbuchlektüre und keine Arbeit mit „Spieldaten“ zu ersetzen.

2. Die Beiträge im Überblick

Die Serie European Data Watch erscheint seit 2000. Bis heute wurden rund 70 Beiträge publiziert; sie alle sind kostenlos auf der Web-Seite des Rates für Sozial- und Wirtschaftsdaten unter <http://www.ratswd.de/publ/datawatch.php> verfügbar. Tabelle 1 enthält eine Zusammenstellung, in der die Titel der Beiträge, ihre Verfasser und das Heft von Schmollers Jahrbuch, in dem der jeweilige Beitrag erschienen ist, aufgelistet sind.

Tabelle 1: Beiträge in der European Data Watch Serie 2000 – 2010 [Stand: 2. Dezember 2009 mit Planung bis Heft 2/2010]

I. Infrastruktureinrichtungen: Forschungsdatenzentren und Datenservicezentren

The German Microdata Lab at ZUMA: Services Provided to the Scientific Community	Lüttlinger, P. et al.	2004 - 3
The research data centres of the Federal Statistical Office and the statistical offices of the Länder	Zühlke, S. et al.	2004 - 4
The Research Data Centre of the Federal Employment Service in the Institute for Employment Research	Kohlmann, A.	2005 - 3
The Research Data Centre of the German Statutory Pension Insurance (FZ-RV)	Rehfeld, U. G., Mika, T.	2006 - 1
The Research Data Centre of the Federal Institute of Vocational Education and Training	Alda, H., Rohrbach-Schmidt, D.	2010 - 2

II. Datensätze für Personen und Haushalte

Deutschland

ALLBUS: A German General Social Survey	Terwey, M.	2000 - 2
The IAB Employment Subsample 1975 – 1995	Bender, S., Haas, A., Klose, C.	2000 - 3
Micro Data on Entrepreneurship: Munich and Leipzig Founder Studies	Brüderl, J., Hinz, T., Ziegler, R.	2001 - 3
Individual tax statistics data and their evaluation possibilities for the scientific community	Zwick, M.	2001 - 4
The German Microcensus	Schwarz, N.	2001 - 4
Consumer Panel Data: A ZUMA research service	Papastefanou, G.	2002 - 4
The 1991/92 German Time Use Microdata	Holz, E.	2003 - 2
The Employment Panel of the German Federal Employment Agency	Koch, I., Meinken, H.	2004 - 2
The 14th "Shell Jugendstudie"	Albert, M. et al.	2004 - 3
German Media-Analyse (MA): A Large Scale Commercial Data Source Available for Secondary Analyses on Media Use and Social Change	Hagenah, J., Meulemann, H., Akinci, H.	2006 - 1
De Facto Anonymised Microdata File on Income Tax Statistics 1998	Merz, J., Vorgrimler, D., Zwick, M.	2006 - 2
Retirement Pension Provision Schemes in Germany 1996 and 2005	Frommert, D., Heien, T.	2006 - 2
Health Data in Germany	Böhm, K.	2006 - 3
The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements	Wagner, G. G., Frick, J. R., Schupp, J.	2007 - 1
The German Integrated Employment Biographies Sample IEBS	Jakobebbinghaus, P., Seth, S.	2007 - 2
The German Taxpayer-Panel	Kriete-Dodds, S., Vorgrimler, D.	2007 - 3
New Possibilities for Socio-Economic Research through Longitudinal Data from the Research Data Centre of the German Federal Pension Insurance (FDZ-RV)	Himmelreicher, R. K., Stegmann, M.	2008 - 4

Österreich		
Longitudinal Data from Social Security Records in Austria	Hofer, H., Winter-Ebmer, R.	2003 - 4
Schweiz		
"Living in Switzerland" Swiss Household Panel 1999 – 2003 (SHP)	Budowski, M. et al.	2002 - 2
The Swiss Environmental Survey 1994	Franzen, A.	2002 - 4
The Swiss Labor Market Survey 1998 (SLMS 98)	Jann, B.	2003 - 2
Swiss Unemployment Insurance Micro Data	Gast, J., Lechner, M., Steiger, H.	2004 - 1
International		
The International Social Survey Programme (ISSP)	Uher, R.	2000 - 3
The European Community Household Panel (ECHP)	Wirtz, C., Mejer, L.	2002 - 1
The LIS / LES Project Databank: Introduction and Overview	Smeeding, T., Jesuit, D. K., Alkemade, P.	2002 - 3
Quality of Life in Europe	Litske, H.	2006 - 3
The Cross-National Equivalent File (CNEF) and its Member Country Household Panel Studies	Frick, J. R. et al.	2007 - 4
III. Datensätze für Firmen (Betriebe und Unternehmen)		
Deutschland		
Firm Panel Data from German Official Statistics	Wagner, J.	2000 - 1
The IAB-Establishment Panel	Kölling, A.	2000- 2
The ZEW Foundation Panels and the Mannheim Enterprise Panel (MUP) of the Centre for European Economic Research (ZEW)	Almus, M., Engel, D., Prantl, S.	2000 - 2
The Mannheim Innovation Panels (MIP and MIP-S) of the Centre for European Economic Research (ZEW)	Janz, N. et al.	2001- 1
Deutsche Bundesbank's Corporate Balance Sheet Statistics and Areas of Application	Stöss, E.	2001 - 1
The German Mechanical Engineering Industry and the NIFA-Panel	Widmaier, U.	2001 - 2
The RWI Data Base on the Globalisation of German Companies	Döhrn, R.	2001 - 3
The 'Service Sector Business Survey' (SSBS) collected by the Centre for European Economic Research and Creditreform	Kaiser, U.	2002 - 2
The EPOC Survey	Pekruhl, U.	2003 - 3
The Hannover Firm Panel (HFP)	Gerlach, K., Hübler, O., Meyer, W.	2003 - 3
The Deutsche Bundesbank's FDI micro database	Lipponer, A.	2003 - 4
The Establishment File of the German Social Insurance Statistics	Fritsch, M., Brixy, U.	2004 - 1
Cost Structure Surveys for Germany	Fritsch, M. et al.	2004 - 4
New Possibilities for Economic Research through Integration of Establishment-level Panel Data of German Official Statistics	Konold, M	2007 - 2

Micro Data at the Ifo Institute for Economic Research – The “Ifo Business Survey” Usage and Access	Becker, S. O., Wohlrabe, K.	2008 - 2
The Establishment History Panel	Spengler, A.	2008 - 3
The German Turnover Tax Statistics Panel	Vogel, A., Dittrich, S.	2008 - 4
The IAB Establishment Panel – Things Users Should Know	Fischer, G. et al.	2009 - 1
Official Firm Data in Germany	Malchin, A., Voshage, R.	2009 - 3
The German Business Services Statistics Panel	Vogel, A.	2009 - 3
The KfW/ZEW Start-up Panel: Design and Research Potential	Fryges, H., Gottschalk, S., Kohn, K.	2010 - 1
The Research Potential of New Types of Enterprise Data based on Surveys from Official Statistics in Germany	Wagner, J.	2010 - 1
Österreich		
--		
Schweiz		
Firm Panel Data from the Swiss Innovation Survey	Arvanitis, S., Hollenstein, H.	2004 - 2
International		
Ifo World Economic Survey Micro Data	Stangl, A.	2007 - 3

IV. Kombinierte Datensätze für Personen und Firmen

Deutschland		
The Lower Saxonian Salary and Wage Structure Survey – Linked Employer-Employee Data from Official Statistics	Stephan, G.	2001 - 2
The linked employer-employee dataset created from the IAB establishment panel and the process-produced data of the IAB (LIAB)	Alda, H., Bender, S., Gartner, H.	2005 - 2
The German Structure of Earnings Survey: Methodology, Data Access and Research Potential	Hafner, H.-P., Lenz, R.	2008 - 3
WeLL – Unique Linked Employer-Employee Data on Further Training In Germany.	Bender, S. et al.	2009 - 4
Österreich		
--		

Schweiz		
--		

V. Weitere Datensätze

A Real-Time Data Set for German Macroeconomic Variables	Gerberding, C. et al.	2005 - 2
The Deutsche Bundesbank’s Large Credit Database (<i>BAKIS-M and MiMiK</i>)	Schmieder, C.	2006 - 4
The Deutsche Bundesbank’s Prudential Database (<i>BAKIS</i>)	Memmel, C., Stein, I.	2008 - 2
The Bundesbank’s macroeconomic real-tiume database for the German Economy (<i>Gerda</i>)	Knetsch, T.	2010 - 2

VI. Datensätze zur Verwendung in der Lehre

CAMPUS Files – Free Public Use Files for Teaching Purposes	Zwick, M.	2007 - 4
--	-----------	----------

VII. Sonderthemen

Anonymising Business Micro Data – Results of a German Project	Lenz, R. et al.	2006 - 4
Computation of Standard Values for Physical and Mental Health Scale Scores Using the SOEP Version of SF-12v2	Andersen, H. H. et al.	2007 - 1
Assessing Cognitive Capacities in Computer-Assisted Survey Research: Two Ultra-Short Tests of Intellectual Ability in the German Socio-Economic Panel (SOEP)	Lang, F. R. et al.	2007 - 1
The German Socio-Economic Panel (SOEP) as Reference Data Set	Siedler, T. et al.	2009 - 2
Business Microdata in Germany: Linkage and Anonymisation	Lenz, R., Zwick, M.	2009 - 4

Anmerkungen: Innerhalb der Rubriken sind die Beiträge chronologisch nach dem Erscheinungsdatum angeordnet; 2000-1 verweist dabei auf das Heft 1 des Jahrgangs 2000 von Schmollers Jahrbuch / Journal of Applied Social Science Studies. Sämtliche erschienenen Beiträge stehen kostenlos zum Download bereit unter <http://www.ratswd.de/publ/datawatch.php>.

Neben den unter I. aufgeführten Aufsätzen zu den Forschungsdatenzentren (FDZ) und Datenservicezentren (DSZ) und den unter VII. genannten Sonderthemen mit engem Bezug zu Mikrodaten listet Tabelle 1 unter II. bis IV. alle Beiträge auf, die jeweils einzelne Datensätze vorstellen. Hierbei wird unterschieden zwischen Datensätzen für Personen und Haushalte, Datensätzen für Firmen (Betriebe und Unternehmen) und kombinierten Datensätzen für Personen und Firmen (linked employer-employee Daten). Die Angaben sind jeweils gegliedert nach Datensätzen für Deutschland, Österreich und die Schweiz sowie internationalen Datensätzen. Außerdem werden unter V. Beschreibungen weiterer Datensätze aufgeführt, die für geldpolitische Analysen verwendet werden können. Eine Sonderstellung nehmen die ausdrücklich für eine Verwendung in Lehrveranstaltungen konzipierten Daten ein, die unter VI. zu finden sind.

Der Schwerpunkt der Serie liegt bisher bei Beiträgen zu deutschen Daten für Personen und Haushalte sowie für Betriebe und Unternehmen. Darüber hinaus sind die verfügbaren deutschen kombinierten Datensätze für Beschäftigte und ihre Arbeitgeber dokumentiert; ferner finden sich Aufsätze zu wichtigen internationalen Datensätzen mit Angaben für Personen.

Viele der Beiträge in der *European Data Watch* Serie sind Standardreferenzen zu den beschriebenen Datensätzen; sie werden daher in zahlreichen Publikationen, die empirische Studien mit diesen Datensätzen enthalten, zitiert. Die Eingabe des Beitrags von Arnd Kölling

zum IAB-Betriebspanel aus Heft 2/2000 in Google Scholar lieferte am 2. Dezember 2009 z.B. Hinweise auf 200 Publikationen, die diesen Aufsatz zitieren. Dies zeigt beispielhaft, dass die *European Data Watch* Serie in der Scientific Community große Aufmerksamkeit findet.

3. Statt eines Schlusswortes – Aufforderung zur Mitarbeit!

Das *European Data Watch* Projekt ist auf die Mitarbeit von Datenexperten – Datenproduzenten, Primärforschern, erfahrenen Nutzern – angewiesen. Das Schreiben von Beiträgen zur Serie ist die Produktion eines öffentlichen Gutes. Die Zeitkosten sind hoch, der persönliche Ertrag ist sehr gering: Es gibt kein Honorar - und keine Punkte in Publikations- und Zitationsrankings. In Zeiten wie diesen, in denen der Wettbewerb auch in der Wissenschaft schärfer wird, ist die Mitarbeit an einem solchen Projekt keineswegs selbstverständlich. Dass sich seit über zehn Jahren immer wieder Mitstreiter finden stimmt mich zuversichtlich hinsichtlich der nächsten Jahre. Wer immer sich als potenzieller Verfasser eines Beitrags angesprochen fühlt, sende mir eine Mail (wagner[at]leuphana.de).

Literatur:

- Kaiser, Ulrich und Joachim Wagner (2008), Neue Möglichkeiten zur Nutzung vertraulicher amtlicher Personen- und Firmendaten. Perspektiven der Wirtschaftspolitik 9(3): 329-349.
- Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (2001), Wege zu einer besseren informationellen Infrastruktur. Baden-Baden: Nomos.