



Rat für Sozial- und
WirtschaftsDaten (RatSWD)

www.ratswd.de

RatSWD

Working Paper Series

Working Paper

Nr. 157

Zur Entwicklung der
Dateninfrastruktur in Deutschland

Roland Habich, Ralf K. Himmelreicher und Denis Huschka

September 2010

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; seit 2009 Gert G. Wagner)

Geschäftsführer des RatSWD (Denis Huschka)

Zur Entwicklung der Dateninfrastruktur in Deutschland

Roland Habich, Ralf K. Himmelreicher und Denis Huschka*

* Der Text basiert in Teilen auf: Habich, Roland/Himmelreicher, Ralf K. und Huschka, Denis: „Datenquellen zur Analyse der Lebensverläufe und Lebensbedingungen in Ost- und Westdeutschland“ In: Krause, Peter/Ostner, Ilona (2010): Leben in Ost- und Westdeutschland. Eine sozialwissenschaftliche Bilanz der deutschen Einheit 1990-2010, Campus: Frankfurt am Main/New York.

1. Einleitung

Die Analyse der Statistik und Dynamik sozialer Phänomene anhand von empirischen Daten hat in den Sozialwissenschaften eine lange Tradition (Adorno 1970; Lowe 1926). Am Anfang des 21. Jahrhunderts können Forschende auf eine Vielzahl qualitativ hochwertiger Daten zurückgreifen (Fachinger u. a. 2010), weshalb zum einen zahlreiche Forschungsprojekte erst ermöglicht wurden und zum anderen nicht mehr in jedem Fall eigene Datenerhebungen benötigt werden. So positiv die Entwicklungen hin zu mehr Datenverfügbarkeit zu bewerten sind, so aktuell ist es aber auch, die Daten im Rahmen einer geordneten und transparenten Infrastruktur zur Verfügung zu stellen. Neben den klassischen Datenarchiven sind die vom Rat für Sozial- und Wirtschaftsdaten initiierten Forschungsdatenzentren (FDZ) und Datenservicezentren (DSZ) Teil einer Forschungsinfrastruktur. Die meisten der intensiv ausgewerteten Datensätze sind heute als Mikrodatensätze über diese Forschungsdateninfrastruktur beziehbar.

Je nach dem jeweiligen Modus ihres Entstehens können drei Grundtypen von Mikrodaten unterschieden werden: a) Daten, die auf der Grundlage von Gesetzen durch die amtliche Statistik „produziert“ werden, b) Daten, die durch Verwaltungsakte¹ entstehen und c) Daten, die im Rahmen wissenschaftlicher Erhebungen oder Erhebungsprogramme entstehen. Ferner können die Daten nach ihrer Periodizität beziehungsweise ihren Erhebungsjahren gekennzeichnet werden.

a) *Datensätze der amtlichen Statistik:* Zu den Erhebungen der amtlichen Statistik gehören beispielsweise der jährlich durchgeführte *Mikrozensus* oder die im fünfjährigen Abstand erhobene *Einkommens- und Verbrauchsstichprobe* (EVS). Bei diesen Erhebungen handelt es sich um allgemeine Haushaltsbefragungen, die unter anderem Informationen über soziodemographische Merkmale von Personen, den Haushalts- und Familienzusammenhang und den Bereich Einkommen enthalten. Die Daten der amtlichen Statistik erlauben somit – anders als prozessproduzierte Daten – Aussagen auf der Personenebene und im Haushaltzzusammenhang. Aufgrund der Größe der Stichproben sind auch Aussagen über kleinere Gruppen möglich.

¹ Diese Art der Datenproduktion ist ebenfalls gesetzlich geregelt. So hat der Gesetzgeber in Paragraph 79 SGB IV eine Rahmenregelung für das Berichtswesen aller Sozialversicherungszweige getroffen. Für die Rentenversicherung werden die Inhalte der einzelnen Berichterstattungen in der „Allgemeinen Verwaltungsvorschrift für die Statistiken in der Rentenversicherung (RSVwV)“ geregelt (Bütefisch 2004).

b) *Prozessproduzierte Datensätze*: Diese Daten werden im Rahmen von Verwaltungsprozessen beispielsweise bei den Trägern der sozialen Sicherung erfasst. In diesem Rahmen dürfen lediglich solche Daten erhoben und gespeichert werden, die zur gesetzlichen Aufgabenerfüllung dieser Institutionen erforderlich sind. Die Vorteile prozessproduzierter Daten liegen in ihrer Vollständigkeit und Genauigkeit.² Bei (retrospektiven) Befragungen würde man hier häufig an Grenzen stoßen, insbesondere wenn das Alter der Zielpopulation sehr hoch und der zu memorierende Zeitraum sehr lang ist. Nachteile prozessproduzierter Daten sind, dass wegen der administrativen Zielsetzung dieser Daten kaum erklärende Variablen erfasst werden. Häufig handelt es sich um Fall- und nicht um Personenerhebungen. Aufbereitete Mikrodaten stehen der Wissenschaft zum Beispiel über das Forschungsdatenzentrum der Rentenversicherung (FDZ-RV) und dem Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung zur Verfügung. Noch vor zehn Jahren erschien eine Weitergabe anonymisierter Einzeldaten seitens der Träger der Sozialen Sicherung an die Wissenschaft unrealistisch.³ Von daher ist die inzwischen erreichte Öffnung des Zugangs der Wissenschaft zu prozessproduzierten Mikrodaten ein enormer Fortschritt.

c) *Wissenschaftliche Erhebungen*: Umfragedaten eröffnen die Möglichkeit, inhaltliche Schwerpunkte zu setzen, zum Beispiel aktuelle sozialpolitische Fragestellungen aufzugreifen und einen umfassenden Katalog soziodemografischer Merkmale zu erheben. Umfragedaten haben damit den Vorzug eines stärkeren Theoriebezugs und einer detaillierteren Messung von erklärenden Variablen. Befragungen haben Personen und/oder Haushalte als Untersuchungseinheit. Prominente Beispiele für wissenschaftsgetragene Erhebungen, die Mikrodaten bereitstellen, sind der ALLBUS, das *Sozio-oekonomische Panel* (SOEP) oder der Survey of Health, Ageing and Retirement in Europe (SHARE). Die genannten Datenquellen sind über die Forschungsdatenzentren der jeweiligen Datenproduzenten zugänglich.

2 Weitere Vorteile prozessproduzierter Daten gegenüber Umfragedaten sind die Entlastung der untersuchten Personen und die Senkung der Kosten der Datenbeschaffung (Rolf-Engel 2010).

3 Hierzu konstatiert Rehfeld im Jahr 2001 in seiner Expertise für die KVI über prozessproduzierte Daten der gesetzlichen Rentenversicherung: „Die Weitergabe von anonymen Einzeldatensätzen (...) für wissenschaftliche Zwecke ist in den Institutionen der sozialen Sicherung nicht vorgesehen und scheint derzeit aufgrund der erheblichen datenschutzrechtlichen Problematik sowie möglicher politischer Auswirkungen ausgeschlossen.“ (ebd. 2001: 14)

Ferner sind als Mischform „halbamtliche Erhebungen“ (Rolf-Engel 2010) zu nennen. So wurden zum Beispiel die Untersuchungen *Alterssicherung in Deutschland* (ASID) im Auftrag des *Bundesministeriums für Arbeit und Soziales* (BMAS) beziehungsweise der Deutschen Rentenversicherung Bund durchgeführt.⁴ Die ASID wurde bis heute sechsmal (einmal pro Legislaturperiode für den Alterssicherungsbericht) erhoben. Die ASID liefert detaillierte Angaben zur Art und Höhe der Alterseinkünfte von Personen und Ehepaaren sowie zu den Bestimmungsfaktoren der Alterssicherung.

2. Zur Entwicklung der Dateninfrastruktur in Deutschland

Datenproduktion, Datenzugang und die Entwicklung einer Forschungsinfrastruktur

Die Weitergabe bereits erhobener Daten ist in den Sozialwissenschaften – im Vergleich zu den Naturwissenschaften, aber auch z. B. den Verhaltenswissenschaften – ausgeprägt und wird seit langem praktiziert. Die Regularien der deutschen Forschungsförderer sehen in der Regel die Langfristarchivierung sozialwissenschaftlicher Daten nach Beendigung des jeweiligen Projektes, in dessen Rahmen sie entstanden sind, im früheren Zentralarchiv für Empirische Sozialforschung in Köln (heute: GESIS Datenarchiv für die Sozialwissenschaften) vor.

Während das Kölner Datenarchiv weiterhin ein wichtiges Portal für den Zugang zu sozialwissenschaftlichen Mikrodaten bleibt, bieten mehr und mehr große Produzenten laufender Erhebungsprogramme ihre Daten direkt über eigene Plattformen und Forschungsdatenzentren an. Hier wird zusätzlich ein Service um die Daten direkt vom Produzenten angeboten und ein direkter Austausch zwischen den Datenproduzenten und der Wissenschaftscommunity ermöglicht.⁵

Die erneute Auswertung durch die Wissenschaftlerkollegen, teilweise unter gänzlich neuen Auswertungsgesichtspunkten, bezeichnet man als Sekundäranalyse.

Anders stellte sich die Situation im Bereich der sozialwissenschaftlich interessanten amtlichen und prozessproduzierten Daten dar. Noch in den 1990er

⁴ Sie können deshalb auch als „halbamtliche“ Daten bezeichnet werden.

⁵ Die Landschaft der Datenarchive wird in Zukunft noch differenzierter werden, da auch die zentralen Fachbibliotheken beginnen, sich mit der Archivierung von Forschungsdaten zu beschäftigen (vgl. Pullinger/Wagner 2010).

Jahren war der Zugang der Wissenschaft zu Mikrodaten der amtlichen Statistik – obwohl öffentlich finanziert erhoben – schwierig. Es gab große Bedenken, die Daten könnten deanonymisiert werden. Die Kosten für den Datenzugang waren prohibitiv⁶ und die verfügbaren Daten häufig unzureichend dokumentiert.

Der Zugang der Forschung zu den Daten der amtlichen Statistik und den in Verwaltungsprozessen entstehenden Daten für sozialwissenschaftliche Analysen hat sich in Deutschland in den letzten Jahren enorm verbessert. Und dies in einem Maße, dass man davon sprechen kann, dass sich Deutschland im europäischen Vergleich von einem Entwicklungsland zu einem innovativen Ideengeber (vgl. Bender u. a. 2009) im Bereich des Datenzugangs entwickelt hat, und – so muss man heute hinzufügen – auch zum Umsetzer dieser Ideen. Als ein Meilenstein in der Entwicklung dieser Forschungsdateninfrastruktur in den Sozial- und Wirtschaftswissenschaften kann das im Jahr 2001 von der *Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik* (KVI) vorgelegte Gutachten gesehen werden (KVI 2001). Die KVI formulierte klare Empfehlungen zur Verbesserung der Dateninfrastruktur für die Wissenschaft. Angeregt durch das KVI-Gutachten kam es zur Gründung der ersten Forschungsdatenzentren und Datenservicezentren. Diese sollen der Wissenschaft den Zugang zu anonymisierten Mikrodaten ermöglichen und das Arbeiten mit diesen Daten erleichtern.⁷ Bereits im Herbst 2001 wurde das erste Forschungsdatenzentrum, beim Statistischen Bundesamt, gegründet. In den Jahren 2002 bis 2004 wurden drei weitere Forschungsdatenzentren (bei den Statistischen Landesämtern, der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung, und der gesetzlichen Rentenversicherung) sowie zwei Datenservicezentren (im früheren Zentrum für Umfragen, Methoden und Analysen [ZUMA] im heutigen GESIS Servicezentrum für Mikrodaten sowie beim Institut zur Zukunft der Arbeit) eingerichtet.

Die sozialwissenschaftliche Forschungsinfrastruktur hat durch diese ersten institutionellen Maßnahmen, die aus Mitteln der Forschungsförderung (mit)finanziert wurden, einen großen Entwicklungsschub erfahren. Die Einrichtung der

6 Sieht man einmal von den Sonderkonditionen für die Bereitstellung einiger Scientific Use Files ab, die durch ein vom BMBF gefördertes Pilotprojekt ermöglicht wurden (vgl. KVI 2001: 84).

7 Ein Jahr nach Vorlage des KVI-Gutachtens hat sich auch der Wissenschaftsrat, das Beratungsgremium von Bund und Ländern zu Fragen der Entwicklung der Wissenschaft, der Forschung und der Hochschulen, ebenfalls für die Einrichtung eines Rates für Sozial- und Wirtschaftsdaten und die Gründung von Forschungsdaten- und Datenservicezentren ausgesprochen (Wissenschaftsrat 2002: 68ff.).

ersten vier Forschungsdatenzentren und der beiden Datenservicezentren hat für die Wissenschaft den Zugang zu Mikrodaten der amtlichen Statistik und anderer wichtiger Datenproduzenten unter voller Wahrung des Datenschutzes geöffnet. Inzwischen gibt es 15 vom RatSWD akkreditierte FDZ und DSZ, mit steigender Tendenz (nach wie vor mit teilweiser Förderung durch das BMBF). Ein Blick auf die Homepages der Datenzentren zeigt die große Zahl der Datenbestände, die inzwischen für die Wissenschaft verfügbar sind.⁸

In Abhängigkeit von der Sensibilität der Daten können unterschiedliche Datenzugangswege genutzt werden: Faktisch anonymisierte Mikrodaten werden als *Scientific Use Files*, absolut anonymisierte Mikrodaten als *Public Use Files* zum Beispiel für die Lehre bereitgestellt. Bei sensibleren Daten besteht in den Forschungsdatenzentren die Möglichkeit des kontrollierten Fernrechnens oder der Nutzung von Gastwissenschaftler-Arbeitsplätzen. Forscherinnen und Forscher müssen heute nur vergleichsweise geringe, zum Teil keine Gebühren für die Bereitstellung von *Scientific* und *Public Use Files* entrichten.⁹ Grundsätzlich haben es sich die Datenzentren zum Ziel gesetzt, der Wissenschaft analysefreundliche Datenprodukte mit umfassenden Datendokumentationen zur Verfügung zu stellen (Bender u. a. 2009).

Zu den zentralen Empfehlungen der KVI gehörte ebenfalls die Schaffung organisatorischer Rahmenbedingungen zur Steuerung der Forschungsdateninfrastruktur. Hierfür wurde im Jahr 2004 der *Rat für Sozial- und Wirtschaftsdaten* (RatSWD¹⁰) vom BMBF im Namen der Bundesregierung eingesetzt. Der RatSWD hat die Aufgabe, den Prozess des Ausbaus und der Verbesserung der Forschungsdateninfrastruktur zu begleiten und die Zusammenarbeit von Wissenschaft und amtlicher Statistik zu fördern.

Der erleichterte Zugang zu Daten aus der amtlichen Statistik und der staatlichen Verwaltung ist Teil einer breiten Entwicklung, die insgesamt zur Verbesserung der Datenlage für die Forschung geführt hat. Auch im Bereich der wissenschaftsgetragenen Produktion von Daten gab es überaus positive Ent-

8 Siehe hierzu auch den Bericht über die Arbeit des RatSWD während seiner ersten Berufungsperiode (vgl. RatSWD 2007: A-29ff.).

9 Damit wurde der Empfehlung der KVI Rechnung getragen, den Datenproduzenten die Kosten für die Anonymisierung der Daten pauschal über Mittel der Forschungsförderung zu erstatten und die Forscherinnen und Forscher nur mit den marginalen Kosten der Abwicklung der Datenlieferung zu belasten (vgl. KVI 2001: 263f.).

10 Siehe www.ratswd.de. Der RatSWD berät die Bundesregierung und die Regierungen der Länder bei der Verbesserung und Weiterentwicklung der Forschungsinfrastruktur im Bereich der empirischen Sozial-, Wirtschafts- und Verhaltenswissenschaften.

wicklungen: neben der Verbesserung der Qualität dieser Daten durch den Fortschritt in der Surveymethodologie und den Erhebungsverfahren, sind im Laufe der Zeit immer reichhaltiger werdende, meist längsschnittliche Datensammlungen – wie die Erhebungen der Deutschen Lebensverlaufsstudie am Max-Planck-Institut für Bildungsforschung oder das *Sozio-oekonomische Panel* (SOEP) – entstanden, die sich im besonderen Maße zur Analyse von Lebensverläufen (und vieler weiterer Fragestellungen) eignen.

Insbesondere europäisch und/oder international vergleichende Datensätze wie der European Social Survey (ESS) oder das International Social Justice Project (ISJP) werden zunehmend auch als Instrumente zur Dauerbeobachtung des sozialen und politischen Wandels genutzt. Auf nationaler Ebene ist die Entwicklungstendenz zum Längsschnitt (vgl. Kaase 2001: 104) nach wie vor zu beobachten, mit pairfam¹¹ und dem *Bildungspanel*¹² (NEPS) werden in der Zukunft weitere Längsschnittsdaten zur Verfügung stehen.

Heute sehen sich die großen wissenschaftsgetragenen Datenproduktionen oft konzeptionell als „Infrastruktureinrichtungen“. Sie schaffen ein Gemeingut und es ist von vorneherein die umfangreiche Dokumentation und Verfügbarkeit der Daten für die gesamte Wissenschaftscommunity, sowie ein Service um die Daten herum, mitgedacht.

Die Analyse dieser Daten durch die gesamte nationale und internationale Wissenschaftscommunity, inzwischen auch fächerübergreifend, kann also streng genommen nicht mehr als Sekundäranalyse bezeichnet werden. Durch geeignete Kommunikations- und Einbindungsmechanismen sind auch externe Datennutzer, das heißt nicht dem datenproduzierenden Wissenschaftsteam direkt verbundenen Forscher, oft bereits an der Konzeption und an der Fortentwicklung der Datenerhebung beteiligt. Ein solches Arrangement geht weit über das hinaus, was man traditionell als Sekundäranalyse bezeichnen würde, und ist eine effektive Weiterentwicklung des alten *Sozialwissenschaften-Bus-Systems*, bei dem Forschende die Möglichkeit hatten, einzelne Fragen zu implementieren.

11 pairfam steht für Panel Analysis of Intimate Relationships and Family Dynamics und ist eine repräsentative Längsschnittstudie zur Erforschung familialer Lebensformen in der Bundesrepublik Deutschland:
<http://www.pairfam.uni-bremen.de/>

12 Das Nationale Bildungspanel liefert Daten zur Erforschung des Bildungserwerbs und seine Folgen für individuelle Lebensverläufe. <http://www.uni-bamberg.de/neps/>

Durch den Erfolg und die Modellhaftigkeit der ersten Datenzentren der amtlichen Statistik und den Anbietern der prozessproduzierten Daten hat es in den letzten Jahren eine dynamische Entwicklung gegeben: Seit Gründung des RatSWD, hat sich die Anzahl der Forschungsdatenzentren verdreifacht. Das Erfolgsmodell Forschungsdatenzentrum praktizieren nunmehr auch wissenschaftsgetragene Datenproduktionen wie der ALLBUS, der Deutsche Alterssurvey oder SHARE (Survey of Health, Ageing and Retirement in Europe) sowie das Sozio-oekonomische Panel (SOEP). Zu erwarten ist, dass die Anzahl der Forschungsdaten- wie Datenservicezentren auch in Zukunft größer werden wird, vor allem durch große wissenschaftsgetragene Datenerhebungen wie zum Beispiel dem *Bildungspanel* (NEPS) und „Panel Analysis of Intimate Relationships and Family Dynamics“ (pairfam).

Die vom RatSWD akkreditierten Forschungsdaten- und Datenservicezentren arbeiten nach definierten Kriterien des RatSWD für die Forschungsdateninfrastruktur in Deutschland. Diese Kriterien dienen der Qualitätssicherung und regulieren unter anderem, dass es keinen Exklusivzugang der jeweiligen Datenproduzenten zu den Daten gibt. Die Einrichtung und der operative Betrieb der Forschungsdaten- und Datenservicezentren werden durch die positive Entwicklung der Nutzerzahlen und die steigende Anzahl von Veröffentlichungen, die auf FDZ-Daten basieren, bestätigt (vgl. Heining 2009; Himmelreicher/Philipps 2009).

3. Technischer Fortschritt und Entwicklung der Dateninfrastruktur

Zunächst ist festzuhalten, dass sowohl empirische Analysen als auch die Gründung von Forschungsdatenzentren ohne die erreichten Fortschritte in der Technik der elektronischen Datenverarbeitung nicht denkbar wären. Diese Entwicklung wird sich fortsetzen. Insbesondere sind weiterhin steigende Geschwindigkeit der Rechner, zunehmend leichtere und billigere Speicherung großer Datenmengen und entsprechende Zugriffsmöglichkeiten zu erwarten.

Die Möglichkeiten der Internetnutzung, im Hinblick auf die Beschaffung von Metadaten (Daten über Daten) wie auch im Rahmen von kontrollierter Datenfernverarbeitung, werden bereits praktiziert und weiter intensiviert werden (vgl. Stegmann 2010). In Zukunft werden nicht mehr Forscher reisen – es sei denn,

detaillierte Fragestellungen und/oder Datenschutz erfordern Kommunikation vor Ort –, sondern Auswertungsprogramme werden auf den Weg geschickt (Rehfeld 2009).

Im Rahmen der EDV-technischen Neuerungen ist die Anwendung rechenintensiver statistischer Verfahren möglich geworden. Hierzu zählen insbesondere multidimensionale komplexe Auswertungsverfahren sowie Verfahren des statistischen Datenmatchings. Solche Verfahren machen es möglich, Auswertungspotenziale verschiedener Datensätze über sogenannte statistische Zwillinge miteinander zu kombinieren (Himmelreicher/Schröder 2010).

Darüber hinaus gibt es technische Innovationen im Bereich webbasierter statistischer Analyseprogramme, die es Datennutzern auch ohne weitergehende Kenntnisse relativ einfach erlauben, im Onlinezugriff aus im Hintergrund abgelegten Mikrodaten Tabellen eigenständig zu kreieren und z.B. auch Regressionsanalysen durchzuführen. Prominentes Beispiel und einer der Vorreiter für einen zeitlich schnellen Datenzugriff für alle wissenschaftlichen Datennutzer ist hier der European Social Survey (ESS). Auf einem allgemein zugänglichen sogenannten „Nesstar-Server“ steht ein webbasiertes „Online browsing and analysis“-Programm zur Verfügung, mit dem standardisierte Metadaten (siehe unten) und die Mikrodaten durchsucht und analysiert werden können (vgl. <http://nesstar.ess.nsd.uib.no/webview/>). Auch das GESIS-Institut für Sozialwissenschaften bietet im Rahmen seines Datenarchivs für einige nationale und internationale Studien diesen komfortablen Datenzugang an (vgl. bspw.: <http://gesis-simon.de>).

Schließlich ist darauf hinzuweisen, dass immer mehr Datenproduzenten im weiteren Umfeld der amtlichen Statistik Online-Recherchemöglichkeiten in ihren eigenen, zum Teil äußerst umfangreichen Datenbanken anbieten. Besonders positiv hervorzuheben ist inzwischen Eurostat; insbesondere deshalb, weil der Zugriff kostenfrei möglich ist. In dieser Hinsicht weniger positiv, aber dennoch sehr lohnenswert erweisen sich die verschiedenen Informationsquellen der OECD. Es ist sicherlich abzusehen, dass weitere Datenproduzenten diesem Trend folgen werden.

Die Herausforderungen, denen sich alle Innovationen in der Weiterentwicklung des Datenzugriffs durch den technischen Fortschritt stellen müssen, dürften dreifach sein: (1) Sicherung des Datenschutzes in einer Welt der

wachsenden Mengen an „Zusatzinformationen“, die im Web leicht zu finden sind und das „Enttarnen“ von faktisch anonymisierten Daten insbesondere über Firmen erleichtern können (Zühlke u. a. 2004), (2) Schutz der „Leitungen“ für das kontrollierte Fernrechnen und der Datenfernverarbeitung (Stegmann 2010) und (3) Sicherung einer nutzerfreundlichen Archivierung sowie einer Langzeitarchivierung (Pullinger/Wagner 2010).

4. Zusammenfassung und Ausblick

Vergleicht man die heutige Situation mit der Datenlage und den Datenzugangsmöglichkeiten für die Sozial-, Wirtschafts- und Verhaltenswissenschaften von vor zehn Jahren, so stellt man fest, dass es große Fortschritte beim Ausbau und der Weiterentwicklung einer Forschungsdateninfrastruktur gegeben hat.

Heute sind mehr Daten für die empirische Forschung zugänglich. Zudem gibt es auch mehr Disziplinen überschreitende, multidisziplinäre Datensätze als früher. Dadurch hat sich der Datenkranz erweitert: So ermöglicht die Erfassung von psychologischen Variablen und intergenerationalen Beziehungen die Beantwortung von Fragestellungen, die vor zehn Jahren noch nicht untersucht werden konnten. Es existieren heute auch mehr Datensätze mit Längsschnittdesign und es werden vermehrt international vergleichbare Daten produziert, mit denen unter anderem auch die Auswirkungen unterschiedlicher Wohlfahrtsregimes untersucht werden können.

Vor diesem Hintergrund sind Forscher mit einer zunehmenden Unübersichtlichkeit einer komplexen „Forschungsinfrastruktur“ konfrontiert. Das Finden der „richtigen“ Datensätze und empirisches Arbeiten auf höchstem Niveau könnte in Zukunft mehr und mehr Spezialwissen voraussetzen. Einerseits sollten deshalb Forschende ein möglichst transparentes Datenangebot vorfinden, andererseits sollten die Zugangswege zu den Daten und die Informationen über deren Potenzial und Auswertungsmöglichkeiten einfach und übersichtlich sein. Die Zusammenarbeit der Forschungsdatenzentren im RatSWD, in dem sowohl die Datenanbieter der amtlichen Statistik und der prozessproduzierten Daten als auch die wissenschaftlichen Datenanbieter in engem Kontakt stehen, bietet Möglichkeiten Synergien zu nutzen und sich auf Standards und gemeinsame Lösungen zu verständigen, um die Nutzung der Daten einfacher zu gestalten.

Gleichzeitig werden Serviceleistungen für die Nutzer angeboten. Im Bereich Datendokumentation kommt es mehr denn je darauf an, die vorhandene Vielfalt der Daten in einer transparenten und vergleichbaren Art zu beschreiben. Hier bieten sich Lösungen wie die Standards der Data Documentation Initiative (DDI) und die sogenannten Digital Object Identifier (DOI) an. Weiterhin muss zukünftig mehr darauf geachtet werden, dass Daten aufbereitende und Service leistende WissenschaftlerInnen die entsprechende Würdigung und Reputation erlangen, die sie verdienen (Wagner 2010). Neuere Entwicklungen in Richtung eindeutiger Identifikationsnummern nicht nur für Datensätze, sondern auch für WissenschaftlerInnen können dazu beitragen, Datenbereitstellung als die eigenständige wissenschaftliche Leistung anzuerkennen, die sie ist.

Ausdrücklich soll aber auch erwähnt werden, dass die empirisch arbeitende akademische Zunft es als ihren originären Auftrag ansehen muss, für eine exzellente Statistikausbildung an den deutschen Universitäten zu sorgen. Die Einführung der Bachelorstudiengänge hat hier Lücken entstehen lassen. Neben der Vermittlung von grundlegendem Statistikwissen kommt es zunehmend auch darauf an, die so erlernten Fähigkeiten in konkrete theoriegeleitete empirische Forschung mit „echten“ Daten münden zu lassen. Hier ist eine enge Zusammenarbeit zwischen den Akteuren in universitärer Methoden- wie Statistikausbildung und jenen der Datenproduzenten beziehungsweise Datenanbietern notwendig.

Mit seinem großangelegtem Projekt „Developing the Research Data Infrastructure for the Social and Behavioral Sciences in Germany and Beyond: Progress since 2001, Current Situation, and Future Demands“ hat der RatSWD eine umfangreiche Bestandsaufnahme der Entwicklungen der vergangenen Jahre angestoßen. Vor allem aber wird ein Impuls gesetzt, über zukünftig wünschenswerte Weiterentwicklungen der Forschungsinfrastruktur für die Sozial-, Wirtschafts- und Verhaltenswissenschaften nachzudenken. In der aus dem Projekt resultierenden Publikation kann man auf über 1200 Seiten die Ergebnisse dieses Nachdenkens nachlesen (vgl.: German Data Forum 2010).

Literatur:

- Adorno, Theodor W. (1970), Über Statik und Dynamik als soziologische Kategorien, in: ders., *Aufsätze zur Gesellschaftstheorie und Methodologie*, Frankfurt am Main, S. 63–85.
- Backes, Gertrud M./Clemens, Wolfgang (2008), Lebensphase Alter. Eine Einführung in die sozialwissenschaftliche Altersforschung, 3. überarbeitete und erweiterte Auflage, Weinheim/München.
- Bender, Stefan/Himmelreicher, Ralf K./Zühlke, Sylvia/Zwick, Markus (2009), Improvement of Access to Data from the Official Statistics, RatSWD Working Paper Nr. 118.
- Bütefische, Thomas (2004), Datenwege und praktischer Datenzugang, DRV-Schriften, Bd. 55, S. 20–23.
- Fachinger, Uwe/Himmelreicher, Ralf K./Rehfeld, Uwe G. (2010), Mikrodaten zur Erforschung der Alterssicherung im 21. Jahrhundert, DRV, Jg. 65, H. 2, S. 173–185.
- German Data Forum (RatSWD) (Hg.) (2010): Building on Progress. Expanding the Research Infrastructure for the Social, Economic, and Behavioral Sciences, Opladen.
- Habich, Roland/Himmelreicher, Ralf K./Huschka, Denis (2010), Datenquellen zur Analyse der Lebensverläufe und Lebensbedingungen in Ost- und Westdeutschland, in: Krause, Peter/Ostner Ilona (Hg.), Leben in Ost- und Westdeutschland. Eine Sozialwissenschaftliche Bilanz der deutschen Einheit 1990–2010, Frankfurt am Main/New York, S. 779–798
- Heining, Jörg (2009), The Research Data Centre of the German Federal Employment Agency: Data Supply and Demand between 2004 and 2009, RatSWD Working Paper Nr. 129.
- Himmelreicher, Ralf K. (2001), Soziodemographie, Erwerbsarbeit, Einkommen und Vermögen von westdeutschen Haushalten. Eine Längsschnitt-Kohortenanalyse auf Datenbasis des SOEP (1984–1997), Berlin.
- Himmelreicher, Ralf K./Schröder, Carsten (2010), Vorüberlegungen zur statistischen Verknüpfung von Querschnitts-Surveydaten mit prozessproduzierten Längsschnittdaten: EVS und VSkt, DRV, Jg. 65, H. 2, S. 208–216.
- Himmelreicher, Ralf K./Philipps, Veronika (2009), Fünf Jahre Forschungsdatenzentrum der Rentenversicherung (FDZ-RV): Datenangebot und -nachfrage zwischen 2004 und 2008, DRV, Jg. 65, H. 2, S. 132–147.
- Kaase, Max (2001), Datenzugang und Datenschutz, in: Bernhard Schäfers/Wolfgang Zapf (Hg.), Handwörterbuch zur Gesellschaft Deutschlands, Bonn, S. 103–115.
- Kohli, Martin (1985), Die Institutionalisierung des Lebenslaufes. Historische Befunde und theoretische Argumente, Kölner Zeitschrift für Soziologie und Sozialpsychologie, Jg. 37, H. 1, S. 1–229.
- Krause, Peter/Ostner, Ilona (2010), Leben in Ost- und Westdeutschland. Eine sozialwissenschaftliche Bilanz der deutschen Einheit 1990–2010, Frankfurt am Main/New York.
- KVI [Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik] (Hg.) (2001), Wege zu einer besseren informationellen Infrastruktur. Gutachten der vom Bundesministerium für Bildung und Forschung eingesetzten Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik, Baden-Baden.
- Lowe, Adolph (1926), Wie ist Konjunkturtheorie überhaupt möglich?, Weltwirtschaftliches Archiv, Bd. 24, S. 165–197.
- Pullinger, John/Wagner, Gert G. (2010), On the Respective Roles of National Libraries, National Archives and Research Data Centers in the Preservation of and Access to Research Data, RatSWD Working Paper Nr. 153.
- Rehfeld, Uwe G. (2009), Das Forschungsdatenzentrum der Rentenversicherung in stetiger Entwicklung, DRV-Schriften, Bd. 55, S. 17–26.
- Rehfeld, Uwe G. (2001), Expertise „Prozessproduzierte Daten zur Alterssicherung in Deutschland“, in: KVI (Hg.), Wege zu einer besseren informationellen Infrastruktur. Gutachten der vom Bundesministerium für Bildung und Forschung eingesetzten Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik, Baden-Baden, CD-Rom im Buch.
- Rolf-Engel, Gabriele (2010), Entwicklung der Dateninfrastruktur für die Wissenschaft, DRV, Jg. 65, H. 2, S. 186–196.
- Schmähl, Winfried (1981), Lebenseinkommens- und Längsschnittanalysen – Methodische und empirische Fragen sowie ihre verteilungs- und sozialpolitische Bedeutung, in: Philipp Herder-Dorneich (Hg.), Dynamische Theorie der Sozialpolitik, Berlin, S. 225–330.

- Sozialgesetzbuch, Viertes Buch (SGB IV) (2009), Gemeinsame Vorschriften für die Sozialversicherung, Text und Erläuterungen, 18. Auflage, herausgegeben von der Deutschen Rentenversicherung, Berlin.
- Stegmann, Michael (2010), Das aktuelle Datenangebot und Neuentwicklungen im FDZ-RV, DRV-Schriften, Bd. 55, S.27–35.
- Wagner, Gert G. (2010): Forschungsdaten fallen nicht vom Himmel, in: Forschung und Lehre, Heft 9.
- Wissenschaftsrat (2002): Empfehlungen zur Stärkung wirtschaftswissenschaftlicher Forschung an den Hochschulen, Drs. 5455 – 02, Saarbrücken, 15. November 2002 (<http://www.wissenschaftsrat.de/texte/5455-02-1.pdf>).
- Zühlke, Sylvia/Zwick, Markus/Scharnhorst, Sebastian/Wende, Thomas (2004), Die Forschungsdatenzentren der statistischen Ämter des Bundes und der Länder, DRV-Schriften, Bd. 55, S. 66–76.