

Zehn Jahre Forschungsdatenzentrum der Statistischen Ämter der Länder

Ein Blick auf Vergangenheit, Gegenwart und Zukunft der Forschungsdateninfrastruktur der amtlichen Statistik in Deutschland

Dipl.-Soz.Wiss. Patrick Rothe

Lange Zeit hinkte die amtliche Statistik in Deutschland bei der Verfügbarmachung von Einzeldaten für Wissenschaft und Forschung anderen Industriestaaten innerhalb und außerhalb Europas deutlich hinterher. Mit der Einrichtung der Forschungsdatenzentren der Statistischen Ämter der Länder und des Statistischen Bundesamtes hat sich diese Situation in den vergangenen zehn Jahren jedoch grundlegend geändert. Wissenschaftler verfügen heute über eine Reihe von Möglichkeiten des Zugangs zu einer reichen Fülle hochwertiger Daten aus dem Bereich der amtlichen Statistik, wie sie vor nicht allzu langer Zeit noch undenkbar schienen. Der vorliegende Artikel soll dazu beitragen, die stattgefundenen Entwicklungen nachzuzeichnen, die gegenwärtige Situation darzustellen und einen kurzen Ausblick in die Zukunft des Forschungsdatenzentrums der Statistischen Ämter der Länder zu wagen.

Das KVI-Gutachten als Startschuss für die Forschungsdatenzentren

Im Oktober 2001 und im April 2002 wurde mit der Einrichtung des Forschungsdatenzentrums des Statistischen Bundesamtes und wenig später des Forschungsdatenzentrums der Statistischen Ämter der Länder der Grundstein für eine moderne, an den Bedürfnissen wissenschaftlicher Nutzerinnen und Nutzer orientierte Dateninfrastruktur in Deutschland gelegt. Vorausgegangen war die Veröffentlichung des Berichts „Wege zu einer besseren informationellen Infrastruktur“ im März 2001 durch die Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI), die sich seit 1999 im Auftrag des Bundesministeriums für Bildung und Forschung (BMBF) mit der Zugangssituation zu Einzeldaten¹ für wissenschaftliche Zwecke auseinandergesetzt hatte. In dem Gutachten waren Deutschland deutliche Defizite beim Zugang zu statistischen Einzeldaten gegenüber anderen führenden Industrienationen attestiert worden. Die Gutachterkommission appellierte daher an die Verantwortlichen aus Politik und Verwaltung, diesem Missstand entgegenzuwirken und auch der Wissenschaft in Deutschland, ein hochwertiges Mikrodatenangebot bereitzustellen (KVI 2001) – der Appell verhallte nicht ungehört: Zehn Jahre später hat sich die Situation gegenüber damals grundlegend verän-

dert (Bender et al. 2008; Habich et al. 2010; Richter 2011; Rolf et al. 2008; Rolf-Engel 2010); die Forschungsdatenzentren der amtlichen Statistik sind zu einer festen Größe geworden, die aus der heutigen Wissenschaftslandschaft nicht mehr wegzudenken ist. Gemeinsam stellen sie der Wissenschaft ein breit gefächertes Dienstleistungsangebot zur Verfügung. Im Folgenden sollen jedoch anlässlich seines diesjährigen zehnjährigen Bestehens primär die Entwicklung und das Angebot des Forschungsdatenzentrums der Statistischen Ämter der Länder im Mittelpunkt dieses Beitrags stehen.

Eine kurze Chronologie des Forschungsdatenzentrums der Statistischen Ämter der Länder

Das Forschungsdatenzentrum des Statistischen Bundesamtes und das Forschungsdatenzentrum der Statistischen Ämter der Länder gehörten zu den ersten Einrichtungen, die vom ebenfalls infolge des KVI-Gutachtens ins Leben gerufenen Rat für Sozial- und Wirtschaftsdaten (RatSWD) formal als Forschungsdatenzentren (FDZ) akkreditiert wurden. Seitdem sind weitere Datenproduzenten aus Verwaltung und Wissenschaft diesem Beispiel gefolgt und ebenfalls dazu übergegangen, interessierten Wissenschaftlern Zugang zu den von ihnen erhobenen Einzeldaten zu gewähren.² Alle dieser offiziell anerkannten Einrichtungen erfüllen eine Reihe von

1 Es handelt sich bei Einzeldaten um Angaben zu einzelnen Einheiten einer Erhebung, beispielsweise Personen, Haushalten oder Unternehmen. Als Synonym für die Bezeichnung Einzeldaten findet oft auch der Begriff Mikrodaten Verwendung.

2 Eine laufend aktualisierte Auflistung aller vom RatSWD akkreditierten Forschungsdatenzentren (FDZ) sowie der thematisch verwandten Datenservicezentren (DSZ) findet sich unter <http://www.ratswd.de/dat/fdz.php>.

Mindestanforderungen, die vom RatSWD in einem verbindlichen Kriterienkatalog zusammengestellt wurden (RatSWD 2010).

Mit der großangelegten Bereitstellung statistischer Einzeldaten betrat die deutsche amtliche Statistik Neuland: In der Vergangenheit war es für Wissenschaftler nur einzelfallbezogen und verbunden mit großen Anstrengungen und Einschränkungen möglich, bei der Durchführung von Forschungsvorhaben auf amtliche Einzeldaten zurückzugreifen. Erst die veränderten rechtlichen Rahmenbedingungen durch die Überarbeitung des Bundesstatistikgesetzes 1987 und die daran anschließende Erprobung wirksamer, praxistauglicher Anonymisierungsverfahren (vgl. z.B. Müller et al. 1991) ebneten den Weg für einen allmählichen Paradigmenwechsel, der schließlich in der Gründung der Forschungsdatenzentren mündete.

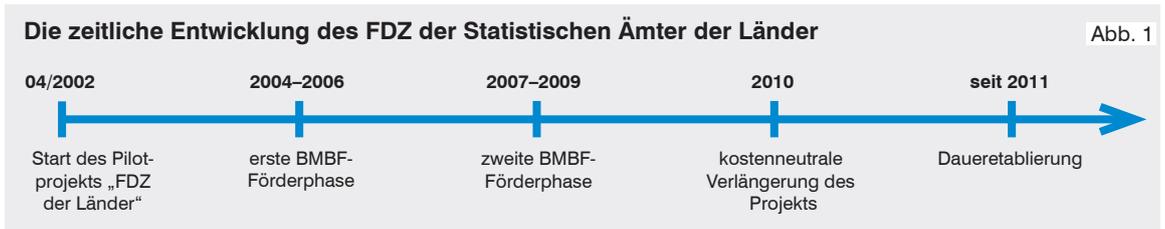
Mit deren Einrichtung im Rahmen eines Pilotprojekts (vgl. Abbildung 1) konnten die Ergebnisse der bisherigen Bemühungen zusammengeführt und der wissenschaftlichen Gemeinschaft ein zentraler Ansprechpartner für die Nutzung amtlicher Statistiken auf Einzeldatenbasis bereitgestellt werden. Ermöglicht wurde dies durch die finanzielle Förderung durch das BMBF. Während der ersten Förderphase von Januar 2004 bis Juni 2007 standen dabei Einrichtung und Betrieb einer fachlich zentralisierten Datenhaltung, die Entwicklung und Pflege eines Metadateninformationssystem, die Einrichtung von Gastwissenschaftlerarbeitsplätzen in allen beteiligten Ämtern sowie die Ermöglichung Kontrollierter Datenfernverarbeitungen an allen Standorten im Mittelpunkt des Projekts.

Neben dem Aufbau der notwendigen Strukturen und deren Anpassung an die datenschutzrechtlichen Rahmenvorgaben, galt es dabei zugleich auch der anfänglichen Skepsis in den Fachabteilungen entgegenzuwirken sowie die neue Einrichtung und deren Angebote und Leistungen unter den potentiellen Nutzern an Hochschulen und Forschungsinstituten bekannt zu machen. Auch war – und ist – es nicht immer einfach, den Entscheidungsträgern außerhalb von Wissenschaft und amtlicher Statistik den gesamtgesellschaftlichen Nutzen, der aus der Arbeit

der Forschungsdatenzentren resultiert, zu verdeutlichen. Das Hauptproblem liegt dabei darin, dass der Nutzen, der mit der Verfügbarkeit der bereitgestellten Mikrodaten der amtlichen Statistik einhergeht, nicht unmittelbar in monetären Beträgen quantifizierbar und individuell zuordenbar ist – ein Problem, das die Forschungsdatenzentren mit einer Vielzahl anderer Bestandteile der öffentlichen Infrastruktur gemeinsam haben.

Im Jahr 2006 wurde das Forschungsdatenzentrum der Statistischen Ämter der Länder anhand der Kriterien der Leibniz-Gemeinschaft erfolgreich evaluiert, wobei sich die beteiligten Gutachter nachdrücklich für eine Weiterförderung des Projekts durch das BMBF aussprachen. Bereits nach vier Jahren Projektlaufzeit bezeichneten diese das Forschungsdatenzentrum als „zentralen und für die Arbeit von zahlreichen Wissenschaftlerinnen und Wissenschaftlern unverzichtbaren Bestandteil der informationellen Infrastruktur in Deutschland“ (Ergebnisse der Evaluierung des FDZ der Statistischen Ämter der Länder 2006: 6). Aufgrund dieses überaus positiven Votums konnten die begonnenen Arbeiten auch in der zweiten Förderphase weiter vorangetrieben und auf eine Verstetigung der neu entstandenen Strukturen hingearbeitet werden. Für das Jahr 2010 erfolgte eine kostenneutrale Verlängerung der Projektförderung bis einschließlich August, um in Abstimmung mit den jeweiligen Dienstaufsichtsbehörden den dauerhaften Fortbestand des Forschungsdatenzentrums der Statistischen Ämter der Länder vorzubereiten. Seit September 2010 wird das Forschungsdatenzentrum finanziell gänzlich von den Statistischen Landesämtern getragen.

Mit dem Jahresbeginn 2011 trat das Forschungsdatenzentrum schließlich in die Phase der Daueretablierung ein, nachdem sich die Dienstaufsichten aller Statistischen Landesämter nach dem Auslaufen der Projektförderung durch das BMBF auf eine dauerhafte Fortführung des Forschungsdatenzentrums verständigt hatten. Vorangegangen war dieser Entscheidung die nachdrückliche Aufforderung durch den RatSWD (2011) sowie den Wissenschaftsrat (2011), das Pilotprojekt in einen dauerhaften Bestandteil der amtlichen Statistik umzuwandeln. Dennoch ist auch heute noch nicht in allen beteiligten



Ländern eine dauerhafte finanzielle Absicherung des Fortbestands der regionalen Standorte des Forschungsdatenzentrums gewährleistet.

Das Leistungsangebot des Forschungsdatenzentrums der Statistischen Ämter der Länder

Mit der Verfügbarmachung von Einzeldaten für wissenschaftliche Zwecke hat die amtliche Statistik in Deutschland einen neuen Weg der Veröffentlichung statistischen Datenmaterials abseits der traditionellen Darstellung aggregierter Daten in Form von Tabellen und Berichten eröffnet. Empirisch arbeitende Wissenschaftler können seitdem unter Wahrung der statistischen Geheimhaltung detaillierte, regional tief gegliederte Auswertungen durchführen und dabei komplexe quantitative Verfahren zur Anwendung bringen, die auf die Nutzbarkeit von Einzeldaten angewiesen sind. Die Möglichkeiten dieser Verfahren gehen dabei weit über das zuvor bei der Nutzung von aggregierten Daten Mögliche hinaus und sind hierdurch in der Lage, eine Vielzahl neuer Erkenntnisse zu erschließen. Die Nutzung amtlicher Statistiken als Grundlage wissenschaftlicher Analysen bietet dabei eine Reihe von Vorteilen gegenüber der Nutzung von Sekundärdaten aus anderen Quellen oder auch Daten aus eigenen Primärerhebungen. Insbesondere die Tatsache, dass es sich bei einer Vielzahl der Statistiken um Vollerhebungen handelt oder diese im Vergleich zu den allermeisten wissenschaftlichen Erhebungen auf sehr großen Stichproben beruhen, macht einen besonderen Vorteil der Nutzung amtlicher Statistikdaten aus. So kann beispielsweise die Non-Response-Problematik, die in der wissenschaftlichen Umfrageforschung mittlerweile ein immer größeres Problem darstellt, beim Rückgriff auf amtliche Statistikdaten weitgehend vernachlässigt werden (Diekmann 2010: 414). Darüber hinaus bieten die hohen Fallzahlen offizieller Statistiken eine überaus solide Grundlage für die Gewinnung aussagekräftiger statistischer Ergebnisse. Hinzu kommt das breite thematische Spek-

trum, das von den Erhebungen der amtlichen Statistik abgedeckt wird. Demgegenüber stehen die prinzipiellen Nachteile, die mit der Arbeit mit Sekundärdaten jeglicher Art einhergehen, allen voran die vom Forschenden nicht an individuelle Fragestellungen anpassbare Operationalisierung der Erhebungsmerkmale (Diekmann 2010: 414f.; Schnell et al. 2005: 238ff.).

Mehr als 2000 Forschungsvorhaben wurden seit ihrer Gründung mit den von den beiden eng miteinander kooperierenden Forschungsdatenzentren der amtlichen Statistik bereitgestellten Daten durchgeführt. Die Mehrzahl der Projekte entstammt dabei dem Bereich der Wirtschafts- und Sozialwissenschaften; aber auch Wissenschaftler aus dem medizinischen, technischen und naturwissenschaftlichen Bereich zählen zu den regelmäßigen Nutzern der Forschungsdatenzentren. Die Bandbreite der Themen und Fragestellungen, für deren Untersuchung Mikrodaten der amtlichen Statistik herangezogen werden, fällt dabei immens aus: Die Erforschung ökonomischer und sozialer Ungleichheiten, des demographischen Wandels oder verschiedenster Aspekte von Migration und Integration ist hierbei ebenso zu nennen wie die Forschung zur Energieversorgung von morgen, gesundheitswissenschaftliche Forschung oder die Untersuchung volks- und betriebswirtschaftlicher Fragestellungen. Die Einzeldaten aus dem Bestand der beiden Forschungsdatenzentren stellen zudem die vermutlich bedeutendste Ressource für den Bereich der empirischen, evidenzbasierten Politikberatung in Deutschland dar.

Das Datenangebot der Forschungsdatenzentren der Statistischen Ämter der Länder und des Bundes umfasst derzeit 107 Statistiken mit insgesamt rund 1 140 nutzbaren Datenmaterialien (Stand: Juni 2012).³ 99 dieser Statistiken werden vom Forschungsdatenzentrum der Statistischen Ämter der

³ Eine Übersicht über die momentan für die wissenschaftliche Nutzung verfügbaren Statistikdaten findet sich unter <http://www.forschungsdatenzentrum.de/datenangebot.asp>.

Länder fachlich betreut. Wenig verwunderlich ist dabei, dass die Daten aus dem Mikrozensus besonders häufig von Nutzern aus Wissenschaft und Forschung nachgefragt werden, lassen sich mit deren Hilfe doch die unterschiedlichsten sozial- und wirtschaftswissenschaftlichen Fragestellungen detailliert untersuchen. Ebenfalls sehr beliebt bei den Nutzern sind die Lohn- und Einkommensteuerstatistik sowie die Einkommens- und Verbrauchsstichprobe. Ein besonders großes Analysepotential bergen zudem diejenigen Statistiken, die in langen Zeitreihen, als Längsschnittdatensätze in Panel-Form oder in Form kombinierter, ursprünglich aus unterschiedlichen Querschnittserhebungen stammender Datenbestände angeboten werden (vgl. z.B. Wagner, J. 2010).

Allein mit der Aufbereitung und Bereitstellung der Daten in den in der Wissenschaft gebräuchlichen Formaten ist es jedoch nicht getan: Um ein sinnvolles Arbeiten mit den Daten zu ermöglichen, sind ebenso eine umfassende Dokumentation und die Erstellung aussagekräftiger Metadaten notwendig. In vielen Fällen wird erst durch die Bereitstellung dieser zusätzlichen Materialien eine effiziente Nutzung des in den bereitgestellten Daten schlummernden Analysepotentials möglich. Mit dieser Aufgabe und der eingehenden fachlichen Betreuung der Nutzer von der ersten Anfrage über die Antragsstellung bis zum erfolgreichen Abschluss der Datennutzung, leisten die Mitarbeiter der Forschungsdatenzentren einen wichtigen Beitrag an der Schnittstelle zwischen amtlicher Statistik und Wissenschaft. Dass es sich hierbei jedoch keineswegs um eine Einbahnstraße handelt, zeigen die zahlreichen Rückmeldungen der datennutzenden Wissenschaftler, die dabei helfen, mögliche qualitative Schwachstellen in den vorhandenen Daten aufzuspüren, und dadurch zu einer verbesserten Plausibilisierung und Fehlerbereinigung – und damit letztlich zu einer erhöhten Datenqualität in der amtlichen Statistik – beitragen. Dies gilt insbesondere für diejenigen Merkmale einer Statistik, die innerhalb des Standardveröffentlichungsprogramms lediglich eine untergeordnete Rolle spielen, bei einer wissenschaftlichen Analyse jedoch von zentralem Interesse für die Beantwortung der verfolgten Fragestellung sein können.

In den Standorten des Forschungsdatenzentrums wird zudem fundierte inhaltliche Arbeit geleistet, die weit über bloße Datenaufbereitung und -bereitstellung sowie Nutzerberatung hinausgeht. Im Rahmen von eigenen Forschungsprojekten werden Ansätze der Datenverknüpfung zur Erschließung neuer Auswertungspotentiale, neuartige Formen der Anonymisierung und Geheimhaltung sowie technische Möglichkeiten der Umsetzung innovativer Datenzugangswege erprobt. Auch Kooperationen mit den Forschungsdatenzentren anderer Datenproduzenten, beispielsweise des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) der Arbeitsagentur in Nürnberg (Bender/Heining 2011), zählen zu diesen Projekten. Zudem erweitern die Angebote des Forschungsdatenzentrums auch die Möglichkeiten zur Durchführung eigener inhaltlicher Forschung in den Fachabteilungen der einzelnen Landesämter. Außerdem stellt das Forschungsdatenzentrum seine fachlich zentralisierte Datenhaltung im Rahmen von Sonderauswertungen für den länderübergreifenden Zugriff durch die Auskunftsdienste bereit.

Eine der maßgeblichen Herausforderungen für die Forschungsdatenzentren der amtlichen Statistik liegt dabei im stetigen Wandel der Anforderungen hinsichtlich Verfügbarkeit, Detailgrad und Qualität, die die Wissenschaft an die angebotenen Forschungsdaten stellt. Genauso wie sich Forschungsmethoden und Themenschwerpunkte über die Zeit hinweg ändern, müssen sich dementsprechend auch die Angebote und Leistungen der Forschungsdatenzentren über die Zeit hinweg an diese Veränderungen anpassen (Zwick 2006).

Im Spannungsfeld zwischen Datenschutz und Wissenschaftsfreiheit

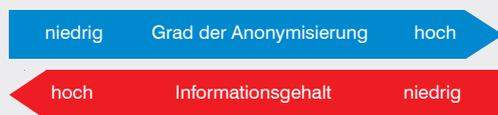
Die rechtliche Grundlage für das dargestellte nutzerorientierte Angebot der Forschungsdatenzentren stellt §16 Abs. 6 BStatG dar, in dem die rechtlichen Rahmenbedingungen für die Weitergabe statistischer Einzeldaten an Einrichtungen mit dem Ziel unabhängiger wissenschaftlicher Forschung geregelt sind. So dürfen diese Angaben unter der Maßgabe an Angehörige von Hochschulen und anderen unabhängigen Forschungseinrichtungen übermittelt werden, dass diese den Anforderungen faktischer Anonymität genügen. Dies ist dann gegeben, „wenn

die Einzelangaben nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft zugeordnet werden können“. Unter den juristischen Begriff der Übermittlung fällt dabei nicht nur die Weitergabe von Daten, beispielsweise in gespeicherter Form auf Datenträgern, sondern bereits das reine Betrachten der Daten an einem Bildschirm.

Der Wahrung des Statistikgeheimnisses kommt bei der Freigabe von amtlichen Statistikdaten für die wissenschaftliche Nutzung höchste Priorität zu, stellt es doch die essentielle Grundlage für das Vertrauen dar, das die Befragten und Erhebungspflichtigen der amtlichen Statistik entgegenbringen. Eine Veröffentlichung von zuordenbaren Einzelangaben darf weder beabsichtigt noch unbeabsichtigt erfolgen. Dies wird durch umfangreiche, auf die jeweiligen Inhalte der Statistik abgestimmte Anonymisierungsmaßnahmen – je nach Nutzungsweg in Verbindung mit einer abschließenden Geheimhaltungsprüfung – sichergestellt. Dabei gilt es, das in den Daten beinhaltete Informationspotential so wenig wie möglich zu beeinträchtigen und zugleich die mögliche Aufdeckung von Einzelangaben zu verhindern. Dass vor diesem Hintergrund in manchen Fällen nicht alles umgesetzt werden kann, was aus wissenschaftlicher Perspektive wünschenswert wäre, ist dabei der Preis, der für die Freigabe umfangreicher Einzelangaben aus zahlreichen amtlichen Statistiken in Kauf genommen werden muss.

Der Zusammenhang zwischen Anonymisierung und Informationsgehalt

Abb. 2



Aufgrund der weitgehenden Unbestimmtheit des Rechtsbegriffs der faktischen Anonymität handelt es sich um einen permanenten, einzelfallbezogenen Abwägungsprozess zwischen zwei verfassungsgemäß garantierten Grundrechten, der Informationellen Selbstbestimmung (Art. 2 Abs. 1 GG) auf der einen und der Wissenschaftsfreiheit (Art. 5 Abs. 3 GG) auf der anderen Seite (Schaar 2009). So stellt faktische Anonymisierung in letzter Instanz immer

einen Kompromiss zwischen dem informativen Gehalt eines Datenbestands und dem Schutz der einzelnen darin enthaltenen Einheiten vor einer potentiellen Re-Identifizierung dar (vgl. Abbildung 2). Ausgehend von der Annahme, es mit einem rational agierenden Datenangreifer zu tun zu haben, kann jedoch unter Berücksichtigung externer Rahmenbedingungen durch angepasste Anonymisierungsmaßnahmen sichergestellt werden, dass die anfallenden Anstrengungen, Kosten und Risiken beim Versuch einer Re-Identifizierung einzelner Einheiten den dadurch erzielbaren Gewinn bei weitem übersteigen. Diese beeinfluss- und kontrollierbaren Rahmenbedingungen sind es auch, die es ermöglichen, externen Datennutzern innerhalb der statistischen Ämter an Gastwissenschaftlerarbeitsplätzen Zugang zu weniger stark anonymisierten Daten zu gewährleisten. Durch die entsprechende Ausgestaltung des dortigen Arbeitsumfelds werden die Möglichkeiten zur Vornahme einer Re-Identifizierung, beispielsweise durch den Abgleich mit externem Datenmaterial, minimiert. Hinzu kommen rechtliche Elemente, wie die Verpflichtung der Nutzer auf die Einhaltung der statistischen Geheimhaltung, deren Missachtung strafrechtliche Folgen nach sich ziehen kann, sowie die schriftliche Vereinbarung von empfindlichen Vertragsstrafen im Falle missbräuchlicher Datennutzungen.

Möglichkeiten zur Nutzung von Mikrodaten über das Forschungsdatenzentrum

Derzeit werden vom Forschungsdatenzentrum vier unterschiedliche Datenzugangswege (vgl. Abbildung 3) angeboten. Welche Daten in welcher Form und mit welcher Detailtiefe verfügbar sind, orientiert sich dabei an der Art des genutzten Zugangswegs. Prinzipiell zu unterscheiden sind dabei die sogenannte Off-Site-Nutzung und die sogenannte On-Site-Nutzung. Zur ersten Kategorie zählt die Bereitstellung der Daten als Scientific-Use-File und als Public-Use-File. Diese können den Datennutzern in Form von Datenträgern zur externen Nutzung überlassen werden. Zur zweiten Gruppe gehören hingegen die Datennutzung am Gastwissenschaftlerarbeitsplatz sowie die „Kontrollierte Datenfernverarbeitung“. Die beiden letzteren haben gemeinsam, dass die vertraulichen Daten die Räumlichkeiten der amtlichen Statistik nicht verlassen.

Die unterschiedlichen Datenzugangswege des Forschungsdatenzentrums im Vergleich

Abb. 3

	Off-Site-Nutzung		On-Site-Nutzung	
	Public-Use-File	Scientific-Use-File	Gastwissenschaftler	kontrollierte Datenfernverarbeitung
Anonymität	absolut	faktisch	faktisch	formal
Auswertungspotential	gering	mittel	hoch	sehr hoch
direkter Datenkontakt	ja	ja	ja	nein
Nutzerkreis	Allgemeinheit	unabhängige Wissenschaft	unabhängige Wissenschaft	unabhängige Wissenschaft
Ort der Nutzung	außerhalb von Räumlichkeiten der amtlichen Statistik		innerhalb von Räumlichkeiten der amtlichen Statistik	

Bei den Scientific-Use-Files (SUF) handelt es sich in der Regel um standardisierte, faktisch anonymisierte Datenbestände, die für eine Nutzung außerhalb der statistischen Ämter freigegeben werden können. Das Analysepotential ist gegenüber den Originaldaten zumeist deutlich eingeschränkt; oftmals handelt es sich lediglich um eine Stichprobe, die aus den Ursprungsdaten gezogen wurde. Vertraglich fixierte Rahmenbedingungen, unter denen die Nutzung zu erfolgen hat, dienen dazu, die Datensicherheit zu erhöhen und datenschutzrechtlichen Belangen Nachdruck zu verleihen. Trotz des eingeschränkten Informationsgehalts sind die Scientific-Use-Files für viele Zwecke wissenschaftlicher Forschung ausreichend; zudem dienen sie oftmals zur Vorbereitung oder Ergänzung einer Nutzung derselben Statistik über einen anderen Zugangsweg. Der große Vorteil aus Sicht der Wissenschaft ist dabei zweifellos die Möglichkeit, die Daten ohne Einschränkungen und abschließende Ergebnisprüfung am eigenen Arbeitsplatz nutzen zu können. Allerdings eignen sich aufgrund der unterschiedlichen datenschutzrelevanten Sensibilität nicht alle Statistiken gleichermaßen für die Erstellung eines Scientific-Use-Files.

Eng verwandt mit den Scientific-Use-Files sind die Public-Use-Files (PUF). Diese enthalten jedoch keine faktisch anonymisierten, sondern absolut anonymisierte Daten; eine Re-Identifikation einzelner in den Daten enthaltener Einheiten – Personen, Unternehmen, Betriebe, etc. – kann dabei komplett

ausgeschlossen werden. Das hat zur Folge, dass bei der Erstellung dieser Datenbestände deutlich stärkere Anonymisierungsmaßnahmen zum Einsatz kommen, womit ein weitaus größerer Verlust an in den Daten enthaltener Information verbunden ist. Oft werden auf diesem Weg auch nur vergleichsweise alte Daten bereitgestellt, da der zeitliche Abstand zu den aktuellen Erhebungsjahren hinsichtlich der Wirkung mit einer zusätzlichen Anonymisierungsmaßnahme vergleichbar ist. Der große Vorteil von Public-Use-Files ist, dass diese ohne Einschränkung allen Interessierten auch außerhalb von Wissenschaft und Forschung verfügbar gemacht werden können. Eine besondere Unterart der Public-Use-Files stellen die CAMPUS-Files dar. Bei diesen handelt es sich um speziell für die Lehre an Hochschulen konzipierte Datensätze, die es Studierenden ermöglichen, das Gelernte bei der Arbeit mit realen Daten umzusetzen, und diese an den Umgang mit Daten der amtlichen Statistik heranzuführen (Zwick 2008). Die aktuell 16 CAMPUS-Files (Stand: Juni 2012) sind auf der Internetseite des Forschungsdatenzentrums kostenlos als Downloads verfügbar.⁴

Bei der Datennutzung am Gastwissenschaftlerarbeitsplatz (GWAP) hingegen werden den Datennutzern die gewünschten Statistikdaten an einem speziellen Arbeitsplatz in den statistischen Ämtern bereitgestellt. Die Unterbringung dieses Arbeitsplatzes in den Räumlichkeiten der amtlichen Statistik ermöglicht zusätzliche organisatorische und tech-

⁴ Zu finden sind die CAMPUS-Files unter der Internetadresse <http://www.forschungsdatenzentrum.de/campus-file.asp>.

nische Vorkehrungen und Maßnahmen, mit denen der Schutz der Daten sichergestellt wird. So ist der Analyse-PC, an dem den Nutzern die Daten sowie die Programme SPSS, STATA und SAS zur Auswertung bereitgestellt werden, technisch nach außen hin abgeschottet; der Netzwerkzugriff ist limitiert, die Schnittstellen des PCs sind deaktiviert. Ein unerlaubtes Zuspälschicken externer Daten oder das Kopieren oder Drucken der genutzten Statistikdaten wird hierdurch unterbunden. Im Gegenzug ist es möglich, den Nutzern detailliertere und weniger stark anonymisierte Daten für ihre Analysen zur Verfügung zu stellen, als dies bei der Bereitstellung eines Scientific-Use-Files der Fall ist. Die von den Gastwissenschaftlern erzeugten Ergebnisse werden von den betreuenden FDZ-Mitarbeitern auf die Einhaltung der statistischen Geheimhaltung geprüft – gegebenenfalls werden entsprechende Zellsperren vorgenommen – und im Anschluss daran in absolut anonymisierter Form an die Datennutzer übermittelt. Gastwissenschaftlerarbeitsplätze können derzeit in allen Statistischen Landesämtern, an den Standorten des Statistischen Bundesamtes in Wiesbaden, Bonn und Berlin sowie in den FDZ-Außenstellen im Deutschen Institut für Wirtschaftsforschung (DIW) in Berlin und an der Technischen Universität Dresden genutzt werden.

Einen Sonderfall unter den angebotenen Wegen des Datenzugangs stellt die „Kontrollierte Datenfernverarbeitung“ (KDFV) dar, bei der es sich unter rechtlichen Gesichtspunkten um eine spezielle Form der Sonderauswertung handelt. Im Rahmen der „Kontrollierten Datenfernverarbeitung“ lassen die Datennutzer dem betreuenden Standort Auswertungsskripte für die Programme SPSS, STATA oder SAS zukommen, die dann von den dortigen Mitarbeitern auf die lokal vorliegenden Einzeldaten angewendet werden. Während der gesamten Nutzung haben die Datennutzer keinen direkten Zugang zu den Einzeldaten der betreffenden Statistiken, so dass über die „Kontrollierte Datenfernverarbeitung“ auf formal anonymisierte Daten⁵ mit uneingeschränktem Analysepotential zurückgegriffen werden kann. Auch hier findet zur Gewährleistung der statistischen Geheimhaltung eine abschließende Überprüfung der Auswertungsergebnisse statt, bevor diese an die Datennutzer weitergegeben werden.

Der bayerische Standort des Forschungsdatenzentrums der Statistischen Ämter der Länder

Der im Landesamt für Statistik und Datenverarbeitung in München angesiedelte bayerische Standort des Forschungsdatenzentrums der Statistischen Ämter der Länder widmet sich im Verbund der regionalen Standorte insbesondere der Aufbereitung und Bereitstellung von Statistiken aus dem Bereich Bildung und Kultur, die vielfältige Analysemöglichkeiten für die wissenschaftliche Forschung – insbesondere im Bereich Hochschule – eröffnen (Wagner, S. 2009).

Ein Gastwissenschaftlerarbeitsplatz, ausgestattet mit einem Analyse-PC mit den gängigsten Statistikpaketen SPSS, STATA, SAS und optional R sowie einem Recherche-PC, steht für die Vor-Ort-Nutzung statistischer Mikrodaten zur Verfügung. Genutzt wird dieser Arbeitsplatz primär von Studierenden und Beschäftigten wissenschaftlicher Einrichtungen aus München und Umgebung. Nahezu sämtliche für die On-Site-Nutzung vorgesehenen Daten aus dem gemeinsamen Angebot des Forschungsdatenzentrums der Statistischen Ämter der Länder und des Forschungsdatenzentrums des Statistischen Bundesamtes können interessierten Wissenschaftlern hier zur Durchführung quantitativer Auswertungen bereitgestellt werden.

Datennutzer in anderen Teilen Deutschlands, die mit Statistiken aus dem fachlichen Zuständigkeitsbereich des bayerischen Standorts arbeiten möchten, haben auf zwei Wegen die Möglichkeit hierzu: So können diese sich die betreffenden Daten vom Standort München für die Nutzung in einem der anderen Standorte der Statistischen Landesämter oder des Statistischen Bundesamtes vorbereiten lassen. Oder sie können – insbesondere dann, wenn eine intensive inhaltliche Betreuung notwendig ist, wenn sehr spezielle Auswertungen durchgeführt werden sollen oder, wenn auf nur formal anonymisiertes Datenmaterial zugegriffen werden soll – auf die „Kontrollierte Datenfernverarbeitung“ zurückgreifen, bei der sämtliche Auswertungen durch den FDZ-Ansprechpartner des Standorts München durchgeführt werden.

⁵ Bei formal anonymisierten Daten handelt es sich um Originaldatenmaterial, aus dem lediglich direkte Identifikatoren, beispielsweise Anschriften, Matrikel- oder Steuernummern, entfernt wurden. Darüber hinaus erfolgt keine Veränderung der Daten unter Anonymisierungsgesichtspunkten, was bedeutet, dass weiterhin der vollständige Informationsgehalt für Analyse Zwecke zur Verfügung steht.

Ausblick in die Zukunft

Ein Jubiläum sollte nicht nur als Anlass dafür dienen, auf bislang Erreichtes zurückzublicken, sondern es sollte dabei auch ein Blick in die nahe und nicht ganz so nahe Zukunft gewagt werden. Diese wird für das Forschungsdatenzentrum der Statistischen Ämter der Länder – so viel ist sicher – neue Herausforderungen bereithalten.

Zwei Beispiele für möglicherweise prägende Zukunftsentwicklungen sind die Bereitstellung der Einzeldaten des Zensus 2011 für die wissenschaftliche Nutzergemeinde sowie der zunehmende Trend zur Georeferenzierung von Statistikdaten. Genauso wie mit der Durchführung eines registergestützten Zensus im Bereich der Erhebungsmethodik Neuland betreten wurde, so gilt dies auch für die Bereitstellung der Zensusergebnisse in Form statistischer Einzeldaten durch die Forschungsdatenzentren. Zahlreiche methodische, technische und organisatorische Herausforderungen werden bewältigt werden müssen, bis die Zensus-Daten für die wissenschaftliche Forschung verfügbar sein werden.

Auch die zunehmende Nachfrage nach georeferenzierten Daten wird die Arbeit der Forschungsdatenzentren nachhaltig beeinflussen. Die Loslösung weg von traditionellen regionalen Gliederungen, wie sie verwaltungstechnisch vorgegeben sind, hin zur Verwendung von Geo-Koordinaten, erfordert nicht nur ein inhaltliches Umdenken bei den Anwendern, sondern ebenso den Einsatz neuer technischer Ansätze und Softwarelösungen zur Verarbeitung und Visua-

lisierung georeferenzierten Einzeldatenmaterials auf Seiten der Datenproduzenten.

Doch nicht nur die Inhalte der statistischen Daten, auch die Methoden und Instrumente zum Umgang mit diesen wandeln sich: So werden neuartige Geheimhaltungs- und Anonymisierungsverfahren, die auf Datenveränderung und nicht mehr auf Datenreduktion oder Sperrung setzen, in Zukunft eine voraussichtlich immer größere Rolle spielen. Auch technologische Entwicklungen, wie automatisierte Geheimhaltungsprogramme oder Ansätze für einen modernen Remote-Access (Brandt/Zwick 2009), mit denen Nutzer aus der Ferne auf Datenbestände der amtlichen Statistik zugreifen können, ohne dass es dabei zu Verletzungen des Datenschutzes kommt, sind keine reinen Visionen mehr, sondern könnten mittel- bis langfristig zunehmend Gestalt annehmen.

Von großer Bedeutung für den Wissenschaftsstandort Deutschland wird es in den kommenden Jahren daher zum einen sein, das bislang Erreichte dauerhaft abzusichern, und zum anderen, weitere Schritte zur fortlaufenden Modernisierung der amtlichen Dateninfrastruktur zu unternehmen, um sich auch zukünftig den sich wandelnden Erfordernissen einer international wettbewerbsfähigen Wissenschaftslandschaft anpassen zu können. Dass ein Aufstieg von einem der untersten auf einen der obersten Ränge im europäischen Vergleich (RatSWD 2011) möglich ist, haben die zurückliegenden zehn Jahre gezeigt; nun gilt es, das Erreichte zu bewahren.

Literatur

Bender, S./Himmelreicher, R./Zühlke, S./Zwick, M. (2008), Forschungsdatenzentren. Eine informationelle Infrastruktur mit erfreulichem Datenzugang. S. 147–158 in: Rolf, G./Zwick, M./Wagner, G. (Hg.), Fortschritte der informationellen Infrastruktur in Deutschland. Festschrift für Johann Hahlen zum 65. Geburtstag und Hans-Jürgen Krupp zum 75. Geburtstag. Baden-Baden: Nomos.

Bender, S. /Heining, J. (2011), The research-data-centre in research-data-centre approach: A first step towards decentralised international data sharing. FDZ-Methodenreport 7/11.

Brandt, M./Zwick, M. (2009), infinitE. Eine informationelle Infrastruktur für das E-Science-Age. Verbesserung des Mikrodatenzugangs durch „Remote Access“. *Wirtschaft und Statistik* 7: 670–675.

Bundesstatistikgesetz (BStatG) (1987). Gesetz über die Statistik für Bundeszwecke.

Diekmann, A. (2010), *Empirische Sozialforschung: Grundlagen, Methoden, Anwendungen*. Reinbeck: Rowohlt.

Ergebnisse der Evaluierung des FDZ der Statistischen Ämter der Länder (2006). Bericht der Gutachtergruppe mit Empfehlungen.

Habich, R./Himmelreicher, R./Huschka, D. (2010),

- Zur Entwicklung der Dateninfrastruktur in Deutschland. RatSWD Working Paper Series 157. Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI) (Hg.) (2001), Wege zu einer besseren informationellen Infrastruktur. Baden-Baden: Nomos.
- Müller, W./Blien, U./Knoche, P./Wirth, H. (1991), Die faktische Anonymität von Mikrodaten. Stuttgart: Metzler/Poeschel.
- RatSWD (2010), Kriterien des Rates für Sozial- und Wirtschaftsdaten (RatSWD) für die Einrichtung der Forschungsdaten-Infrastruktur.
- RatSWD (2011), Auf Erfolgen aufbauend – Zur Weiterentwicklung der Forschungsinfrastruktur für die Sozial-, Verhaltens- und Wirtschaftswissenschaften. Empfehlungen des Rates für Sozial- und Wirtschaftsdaten. Opladen: Budrich.
- Richter, A. (2011), 10 Jahre „Wege zu einer besseren informationellen Infrastruktur“. Statistik und Wirtschaft in Hessen 2/3: 27–35.
- Rolf, G./Zwick, M./Wagner, G. (2008), Fortschritte und Herausforderungen der informationellen Infrastruktur in Deutschland. RatSWD Working Papers Series 41.
- Rolf-Engel, G. (2010), The recommendations of the 2001 KVI Report and their realizations up to 2008. S. 189–214 in, RatSWD (Hg.), „Building on progress. Expanding the research infrastructure for the social, economic, and behavioral sciences“. Volume 1. Opladen/Farmington Hills: Budrich.
- Schaar, P. (2009), Data protection and statistics – A dynamic and tension-filled relationship. RatSWD Working Paper Series 82.
- Schnell, R./Hill, P./Esser, E. (2005), Methoden der empirischen Sozialforschung. München: Oldenbourg.
- Wagner, J. (2010), The research potential of new types of enterprise data based on surveys from official statistics in Germany. Schmollers Jahrbuch 130: 133–142.
- Wagner, S. (2009), Bildungsdaten der amtlichen Statistik im Forschungsdatenzentrum. Bayern in Zahlen 7: 287–300.
- Wissenschaftsrat (2011), Empfehlungen zu Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften.
- Zwick, M. (2006), Forschungsdatenzentren – Nutzen und Kosten einer informationellen Infrastruktur für Wissenschaft, Politik und Datenproduzenten. Wirtschaft und Statistik 12: 1 233–1 240.
- Zwick, M. (2008), CAMPUS-Files – Kostenlose Public Use Files für die Lehre. AStA Wirtschafts- und Sozialstatistisches Archiv, Vol.2, Nr. 1–2: 175–188.