
Der folgende Aufsatz ist ein Nachdruck des Beitrages aus „Statistisches Bundesamt: Wirtschaft und Statistik“, Heft 10/2003, der gemeinsam vom Forschungsdatenzentrum des Statistischen Bundesamtes und der Geschäftsstelle des Forschungsdatenzentrums der Statistischen Landesämter erstellt worden ist.

Dr. Sylvia Zühlke, Markus Zwick, Sebastian Scharnhorst und Thomas Wende*)

Die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder

Die Komplexität des wirtschaftlichen und sozialen Wandels sowie die Fortschritte in Wissenschaft und Informationstechnik haben den Datenbedarf moderner Gesellschaften grundlegend verändert. Die zur Analyse und Gestaltung moderner Gesellschaften erforderlichen Daten müssen insbesondere Informationen über Teilgruppen der Gesellschaft liefern sowie die Analysen des wirtschaftlichen und sozialen Wandels auf der Basis von Längsschnittdaten erlauben. Aufgrund des geänderten Informationsbedarfs ist die Veröffentlichung von Ergebnissen in Form von Tabellen heute nicht mehr ausreichend. Vielmehr entspricht es den methodischen und inhaltlichen Erfordernissen, statistische Daten entsprechend dem wissenschaftlichen Datenbedarf bereitzustellen. Dazu gehört es, Zugangsmöglichkeiten zu anonymisierten und nicht anonymisierten Mikrodaten zu schaffen, mit denen differenziertere Analysen durchgeführt werden können.

Vor diesem Hintergrund wurde in den letzten Jahren in Deutschland intensiv über den Zugang der Wissenschaft zu den Mikrodaten der amtlichen Statistik diskutiert. Durch die „Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik“ (KVI) wurden dafür im Auftrag des Bundesministeriums für Bildung und Forschung verschiedene Vorschläge zur Verbesserung der Wechselwirkungen zwischen Wissenschaft und Statistik erarbeitet (siehe Kommission zur Verbesserung der informationellen Infrastruktur 2000)¹⁾. Eine der zentralen institutionellen Empfehlungen dieser Kommission bestand in der möglichst raschen Einrichtung von Forschungsdatenzentren bei den Datenproduzenten.

Diese Empfehlung wurde von der amtlichen Statistik aufgegriffen. Im Jahr 2001 hat das Statistische Bundesamt ein solches Forschungsdatenzentrum etabliert. Ein weiteres Forschungsdatenzentrum der Statistischen Landesämter wurde im März 2002 eingerichtet, als eine gemeinsame Einrichtung aller Statistischen Landesämter mit 16 regionalen Standorten. Mit diesen Forschungsdatenzentren intensiviert die deutsche amtliche Statistik ihre bisherigen Bemühungen, die Mikrodaten der amtlichen Statistik für wissenschaftliche Analysen zugänglich zu machen.

Ziel dieses Aufsatzes ist es, die neuen Formen und Wege der Nutzung amtlicher Mikrodaten vorzustellen, die sich aus der Einrichtung der Forschungsdatenzentren ergeben haben. Hierfür folgt zunächst ein Überblick über die Entwicklung der Rahmenbedingungen für die Nutzung von Mikrodaten der amtlichen Statistik in Deutschland. Anschließend werden die Ziele und Aufgaben der Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder beschrieben sowie die unterschiedlichen Möglichkeiten der Datennutzung präsentiert, die mit den Forschungsdatenzentren angeboten werden.

*) Dr. Sylvia Zühlke und Sebastian Scharnhorst sind in der Geschäftsstelle des Forschungsdatenzentrums der Statistischen Landesämter tätig, die im Landesamt für Datenverarbeitung und Statistik Nordrhein-Westfalen angesiedelt ist. Markus Zwick und Thomas Wende sind Mitarbeiter im Forschungsdatenzentrum des Statistischen Bundesamtes.

1) Siehe Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (Hrsg.) „Wege zu einer besseren informationellen Infrastruktur“, Baden-Baden 2001.

Die Nutzung von Mikrodaten der amtlichen Statistik in Deutschland

Die Nutzung von Mikrodaten der amtlichen Statistik durch die Wissenschaft wurde in Deutschland sehr stark von der Entwicklung des Bundesstatistikgesetzes (BStatG) beeinflusst, das 1951 erstmals verabschiedet und in den Jahren 1981 und 1987 novelliert wurde.

Die Übermittlung von Mikrodaten an Dritte wurde bei Entstehung des Bundesstatistikgesetzes kaum diskutiert, so dass in der Fassung von 1951 hierzu noch keine explizite Regelung enthalten war. Aufgrund der fehlenden Möglichkeiten, umfangreiche Mikrodaten zu verarbeiten, war die Nachfrage nach Mikrodaten in den 50er, 60er und frühen 70er Jahren relativ gering. Sofern die Übermittlung von Einzelangaben nicht explizit in anderen Gesetzen geregelt war, wurden für einzelne Projekte formal anonymisierte Mikrodaten zur Verfügung gestellt. So wurden amtliche Mikrodaten seitens der Wissenschaft erstmals in dem Projekt "Sozialpolitisches Entscheidungs- und Indikatoren-system für die Bundesrepublik Deutschland (SPES)" analysiert, das 1972 bis 1978 durchgeführt wurde²⁾. Diesem Projekt stellte die amtliche Statistik formal anonymisierte Mikrodaten des Mikrozensus und der Einkommens- und Verbrauchsstichprobe zur Verfügung. Für das gleiche Projekt wurde auch eine Stichprobe der Volkszählung 1970 (VZ70) auf der Grundlage des Volkszählungsgesetzes bereit gestellt.

Als die Nachfrage nach Mikrodaten mit der Fortentwicklung der Informationstechnik schnell anstieg, wurde die Übermittlung von Mikrodaten an Dritte erstmals explizit mit der Verabschiedung des Bundesdatenschutzgesetzes im Jahr 1977 und mit der Novellierung des BStatG im Jahr 1981 geregelt. Mit der Einführung der sogenannten Weiterleitungsvorschrift des § 11 BStatG 1981 wurde die Absicht verfolgt, eine bessere Versorgung der Wissenschaft mit Mikrodaten zu ermöglichen. Diese Vorschrift sah die Weitergabe von Mikrodaten in absolut anonymisierter Form an die Nutzer vor. Mit der Anwendung dieser Möglichkeit innerhalb konkreter Projekte zeigte sich allerdings, dass

die Anforderung an ein solchermaßen absolut anonymisiertes Datenmaterial so restriktiv war, dass es in der Folgezeit nur noch in Ausnahmefällen zu Datenlieferungen kam. Aufgrund der gesetzlichen Änderung des BStatG konnten für nachfolgende Projekte wie „Vergleichende Analysen der Sozialstruktur mit Massendaten“ (VASMA) oder dem Sonderforschungsbereich 3 „Mikroanalytische Grundlagen der Gesellschaftspolitik“ der Deutschen Forschungsgemeinschaft nur noch absolut anonymisierte Mikrodaten zu relativ hohen Kosten zur Verfügung gestellt werden³⁾.

Das Volkszählungsurteil von 1983 und die damit einhergehende Diskussion zeigten, dass sowohl die informationelle Selbstbestimmung als auch die Wissenschaftsfreiheit, beide in Artikel 5 des Grundgesetzes verankert, als wichtige Grundrechte gleichrangig zu behandeln sind. Somit war der Gesetzgeber aufgerufen, für einen adäquaten Datenzugang Sorge zu tragen. Mit der weiteren Modifikation des BStatG im Jahre 1987 und der Einführung einer faktischen Anonymität für die Weitergabe von Mikrodaten an die Wissenschaft kam er diesen Anforderungen nach. Die faktische Anonymität erlaubte es nun, im Rahmen des sogenannten Wissenschaftsprivilegs, unter bestimmten Voraussetzungen, Mikrodaten an die Wissenschaft zu liefern, die ein Restrisiko der Deanonymisierung aufwiesen. In der Folgezeit konkretisierten verschiedene Projekte die Ausgestaltung faktisch anonymisierter Mikrodatensätze. Insbesondere die Ergebnisse des Projektes „Die faktische Anonymisierung von Mikrodaten“ erlaubten es, ab Mitte der neunziger Jahre erste standardisierte faktisch anonymisierte Mikrodatensätze für den Bereich der Haushalts- und Personenerhebungen zur Verfügung zu stellen⁴⁾.

Während hierdurch und durch weitere Aktivitäten außerhalb der amtlichen Statistik⁵⁾ insbesondere für sozialwissenschaftliche Fragestellungen eine neue Datenbasis geschaffen wurde, konnten im Bereich der Wirtschaftswissenschaften aufgrund der größeren Schwierigkeiten bei der Anonymisierung von Betriebs- und Unternehmensdaten vergleichbare Fortschritte zunächst nicht erzielt werden.

2) Siehe Krupp, Hans-Jürgen (1973) „Sozialpolitisches Entscheidungs- und Indikatoren-system für die Bundesrepublik Deutschland“, Allgemeines Statistisches Archiv, Band 57, S. 380-387.

3) Zu den Projekten siehe Hauser, Richard (Hrsg.): Mikroanalytische Grundlagen der Gesellschaftspolitik: Ergebnisse aus dem gleichnamigen Sonderforschungsbereich, Band 1 und 2, Berlin 1994. Die Ergebnisse des Projektes VASMA sind unter <http://www.gesis.org/Dauerbeobachtung/Mikrodaten/Daten/brd/literatur.pdf> dokumentiert.

4) Siehe hierzu insbesondere Müller, Walter/Blien, Uwe/Knoche, Peter/Wirth, Heike u.a. „Die faktische Anonymität von Mikrodaten“, Band 19 der Schriftenreihe Forum der Bundesstatistik, hrsg. vom Statistischen Bundesamt, Wiesbaden 1991.

5) Außerhalb der amtlichen Statistik haben sich insbesondere mit dem Sozio-oekonomischen Panel (SOEP) und der „Amtlichen Bevölkerungsumfrage der Sozialwissenschaften“ (ALLBUS) Erhebungen etabliert, die regelmäßig durchgeführt werden und für wissenschaftliche Analysezwecke zur Verfügung stehen.

Der fehlende Zugang insbesondere zu den Unternehmens- und Betriebsdaten wurde seitens der Wissenschaft in dem Memorandum „Erfolgsbedingungen empirischer Wirtschaftsforschung und empirisch gestützter wirtschafts- und sozialpolitischer Beratung“ aufgegriffen⁶⁾. Hierin wurde gefordert, den Zugang zu schwer anonymisierbaren Mikrodaten innerhalb der Räumlichkeiten der Datenproduzenten zu ermöglichen. Die Diskussion über den Mikrodatenzugang der Wissenschaft gewann mit dem Memorandum sowie dem Symposium „Kooperation zwischen Wissenschaft und amtlicher Statistik - Praxis und Perspektiven“⁷⁾ im Jahre 1999 eine neue Dynamik, die auch von der Politik aufgegriffen wurde.

In der Folge erarbeitete die Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI) eine Reihe von Empfehlungen zur Verbesserung der Kooperation von Wissenschaft und Statistik⁸⁾. Diese reichen von der Mitwirkung der Datennutzer bei der Aufstellung der Erhebungs- und Aufbereitungsprogramme über die Perspektiven einer modernen Aus- und Weiterbildung in der Statistik bis zu den verschiedenen Möglichkeiten für den Zugang der Wissenschaft zu den Mikrodaten der öffentlichen Datenproduzenten. Eine wesentliche institutionelle Forderung bezieht sich auf die Einrichtung von Forschungsdatenzentren bei den Datenproduzenten und auf die Einrichtung von Servicezentren. In der Zwischenzeit werden die Empfehlungen der Kommission zur Verbesserung der informationellen Infrastruktur von dem Gründungsausschuss des Rates für Sozial- und Wirtschaftsdaten umgesetzt, so dass eine Reihe von Forschungsdatenzentren und Servicezentren bereits ihre Arbeit aufnehmen konnten⁹⁾.

Hierzu zählen auch die beiden Forschungsdatenzentren der amtlichen Statistik. Zwar sind diese zwei voneinander unabhängige Einrichtungen, gleichwohl stimmen sie ihre Arbeit eng untereinander ab, um der Wissenschaft ein gemeinsames Angebot für einen verbesserten Datenzugang zu unterbreiten. Den Empfehlungen der KVI folgend, nach denen die Einrichtung einer leistungsfähigen Dateninfrastruktur eine Aufgabe der Forschungsförderung ist, haben

sowohl das Statistische Bundesamt als auch die Statistischen Landesämter einen Förderantrag beim Bundesministerium für Bildung und Forschung gestellt. Mit Hilfe dieser Mittel soll ein umfangreiches Dienstleistungs- und Datenangebot geschaffen werden.

Ziele und Aufgaben der Forschungsdatenzentren

Das wesentliche Ziel der Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder besteht darin, den Zugang der Wissenschaft zu den Mikrodaten der amtlichen Statistik durch die Einrichtung unterschiedlicher Datennutzungswege zu erleichtern. Eine wesentliche Voraussetzung für die Realisierung dieses Ziels besteht in der grundsätzlichen Verbesserung der Dateninfrastruktur durch den Aufbau einer fachlich zentralisierten Datenhaltung für ausgewählte Statistiken und durch die Einrichtung eines Metadateninformationssystems.

In Deutschland wird der überwiegende Teil der Statistikerhebungen dezentral in den Statistischen Landesämtern durchgeführt, so dass dort über 90 Prozent aller Mikrodaten der amtlichen Statistik erhoben, aufbereitet und gespeichert werden. Da sich wissenschaftliche Analysen in der Regel jedoch auf mehrere Bundesländer oder das gesamte Bundesgebiet beziehen, planen die Statistischen Landesämter den Aufbau einer fachlich zentralisierten Datenhaltung. Dadurch wird es möglich sein, die Mikrodaten der amtlichen Statistik länderübergreifend an allen regionalen Standorten der Forschungsdatenzentren zu nutzen.

Damit wissenschaftliche Nutzer die Mikrodaten der amtlichen Statistik analysieren und interpretieren können, benötigen sie zudem umfassende Informationen über die Datensätze sowie über die Erhebung, Aufbereitung und Qualität der Daten. Daher werden die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder ein internetbasiertes Metadateninformationssystem entwickeln, in dem sich die Nutzer über die Erhebungen der amtlichen Statistik informieren können.

6) Siehe Hauser, Richard/Wager, Gerd/Zimmermann, Klaus „Erfolgsbedingungen empirischer Wirtschaftsforschung und empirisch gestützter wirtschafts- und sozialpolitischer Beratung: Ein Memorandum“, Allgemeines Statistisches Archiv, Band 82, S. 369-379.

7) Die Ergebnisse des Symposiums sind dokumentiert in Müller, Walter/Schimpl-Neimanns, Bernhard/Krupp, Hans-Jürgen/Wiegert, Rolf u.a. „Kooperation zwischen Wissenschaft und amtlicher Statistik – Praxis und Perspektiven – Beiträge zum Symposium am 31. Mai / 1. Juni 1999 in Wiesbaden“, Band 34 der Schriftenreihe Forum der Bundesstatistik, hrsg. vom Statistischen Bundesamt, Wiesbaden 1999.

8) Siehe Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (Hrsg.) „Wege zu einer besseren informationellen Infrastruktur“, Baden-Baden 2001.

9) Für einen Überblick siehe Lüttinger, Paul/Schimpl-Neimanns, Bernhard/Wirth, Heike/Papastefanou, Georg: Mikrodaten (German Microdata Lab): Das Servicezentrum für amtliche Mikrodaten bei ZUMA, ZUMA Nachrichten 5, 2003, S. 153-172.

Datennutzungswege

Um der Wissenschaft den Zugang zu dem gesamten Informationspotential der amtlichen Statistik zu eröffnen, richten die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder unterschiedliche Zugangswege zur Nutzung ihrer Mikrodaten ein. Damit werden den Nutzern zusätzliche und weitergehende Möglichkeiten zur Analyse von Mikrodaten der amtlichen Statistik eröffnet, als dies bislang möglich war.

Ansatzpunkt dieser zusätzlichen Nutzungswege ist es, die Vermeidung einer Reidentifikation der Auskunftgebenden nicht mehr allein durch Veränderungen im Datenmaterial sicherzustellen, sondern auch durch die Regulierung des Datenzugriffs. Die einzelnen Nutzungswege resultieren daher aus verschiedenen Kombinationen von Datenanonymisierung und Zugriffsregulierung.

a) Absolut anonymisierte Mikrodatensätze

Absolut anonymisierte Daten werden durch Aggregation oder durch die Entfernung einzelner Merkmale soweit verändert, dass eine Identifizierung der Auskunftgebenden nach menschlichem Ermessen unmöglich gemacht wird. Die amtliche Statistik bietet absolut anonymisierte Mikrodaten in Form sogenannter Public Use Files (PUF) an. Diese können allen interessierten Personen zur Verfügung gestellt werden.

Bislang wurden solche Datensätze für die Sozialhilfestatistik sowie für die Zeitbudgeterhebung erstellt. Auch in diesem Segment werden die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder zukünftig verstärkt an einer Ausweitung ihres Angebotes arbeiten. Eine weitere Hauptzielrichtung der Public Use Files liegt im Bereich der Hochschullehre. Die Forschungsdatenzentren entwickeln derzeit sogenannte Campus-Files, die an Hochschulen zu Lehrzwecken eingesetzt werden können. Es ist geplant, einen ersten Campus-File des Mikrozensus 1998 noch im laufenden Semester zur Verfügung zu stellen. Dieser Datensatz soll Studentinnen und Studenten schon frühzeitig die Möglichkeit geben, die Besonderheiten der Analyse von Mikrodaten kennen zu lernen.

b) Faktisch anonymisierte Mikrodatensätze

Eine absolute Datenanonymisierung birgt den Nachteil, dass damit auch ein erheblicher Teil der statistischen Information verloren geht. Dagegen werden Mikrodaten als faktisch anonymisiert bezeichnet, wenn die Deanonymisierung zwar nicht gänzlich ausgeschlossen werden kann, die Angaben jedoch nur mit einem unverhältnismäßig hohen Aufwand an Zeit, Kosten und Arbeitskraft dem jeweiligen Merkmalsträger zugeordnet werden können¹⁰⁾. Die Hauptzielrichtung der faktischen Anonymisierung besteht folglich darin, durch behutsame Informationsreduktion und Informationsveränderungen die Zuordnungsmöglichkeiten von Merkmalsausprägungen zu den entsprechenden Merkmalsträgern zu verringern, dabei jedoch den statistischen Informationsgehalt zu schonen. Hierfür müssen für jede einzelne Erhebung der Aufwand und der Nutzen einer Deanonymisierung analysiert werden. Dabei können unterschiedliche Anonymisierungsverfahren zur Anwendung kommen¹¹⁾. Nach Maßgabe des Bundesstatistikgesetzes können faktisch anonymisierte Daten allerdings nur wissenschaftlichen Einrichtungen und nur zur Durchführung wissenschaftlicher Vorhaben zugänglich gemacht werden.

Faktische Anonymität resultiert also nicht allein aus dem realen Informationsgehalt der Daten, sondern auch aus den bestehenden Möglichkeiten zur Deanonymisierung. Wann ein Mikrodatensatz als faktisch anonym bezeichnet werden kann, hängt daher insbesondere davon ab, unter welchen Rahmenbedingungen die Daten verarbeitet werden. So ist von entscheidender Bedeutung, welches Zusatzwissen vorliegt und wo die Datennutzung stattfindet. Abhängig davon, ob die Mikrodaten extern oder in den Statistischen Ämtern genutzt werden, kann die faktische Anonymität mit mehr oder minder starken Informationsveränderungen erreicht werden.

Ein immer wieder deutlich vorgetragener Wunsch der Wissenschaft ist es, Mikrodaten in anonymisierter Form am eigenen Arbeitsplatz zu nutzen. Die faktische Anonymisierung der Mikrodaten ermöglicht diese Übermittlung nicht vollständig anonymisierter Mikrodaten zur externen (Off-Site-) Nutzung in wissenschaftlichen Einrichtungen.

10) Diese Regelung geht zurück auf den § 16 Abs. 6 BStatG.

11) Einen Überblick über Anonymisierungsmethoden ist zu finden in Köhler, Sabine: Anonymisierung von Mikrodaten in der Bundesrepublik und ihre Nutzung – Ein Überblick. Forum der Bundesstatistik 31, Statistisches Bundesamt, 1999, 133–144.

Da jedoch allein durch die Herausgabe dieser Daten ein höheres Deanonymisierungsrisiko besteht als bei der Nutzung in einem Statistischen Amt, ist die Datenanonymisierung relativ ausgeprägt. Die für diese Nutzungsform erzeugten Datensätze werden als Scientific Use Files (SUF) bezeichnet.

Die amtliche Statistik bietet im Bereich der personenbezogenen Daten mit dem Mikrozensus, der Einkommens- und Verbrauchsstichprobe und der Zeitbudgeterhebung bereits ein breites Datenspektrum als Scientific Use File an. Die Forschungsdatenzentren des Bundes und der Länder sind bestrebt, dieses Angebot sukzessive zu erweitern. So laufen zur Zeit Projekte zur Anonymisierung der Lohn- und Einkommensteuerstatistik und der Diagnosestatistik. Im kommenden Jahr soll darüber hinaus mit der Gehalts- und Lohnstrukturhebung erstmals auch ein sogenannter Employer-Employee-Datensatz zur standardmäßigen Anonymisierung herangezogen werden. Mit dem Projekt "Faktische Anonymisierung wirtschaftsstatistischer Einzeldaten" versucht die amtliche Statistik, gemeinsam mit den wissenschaftlichen Nutzern auch in diesem Bereich anonymisierte Standardfiles zu entwickeln¹²⁾.

c) Projektbezogene faktische Anonymisierung zur On-Site-Nutzung

Bei geringer Nachfrage nach einzelnen Statistiken und bei schwer zu anonymisierenden Mikrodaten wird es in vielen Fällen nicht sinnvoll sein, in einem aufwändigen Verfahren standardisierte Scientific Use Files zu erstellen. Hier ist eine projektbezogene Anonymisierung der Daten sinnvoller. Dies hat den praktischen Vorteil, dass dabei nicht die gesamte Statistik anonymisiert wird, sondern lediglich die aus ihr benötigten Merkmale.

Die projektbezogene Anonymisierung erzeugt ebenfalls eine faktische Datenanonymität. Diese Daten können jedoch nur in den Räumlichkeiten der Forschungsdatenzentren des Bundes und der Länder an sogenannten Gastwissenschaftlerarbeitsplätzen ausgewertet werden. Da die Mikrodaten dabei die Räumlichkeiten der amtlichen Statistik nicht verlassen und die Nutzer nicht beliebig über die Daten verfügen und sie kaum mit Zusatzinformationen kombi-

nieren können, hat diese Nutzungsform einen weiteren wesentlichen Vorteil. Da die faktische Anonymität hier bereits bei wesentlich geringeren Veränderungen im Datenmaterial erreicht wird als bei der Erstellung von Scientific Use Files zur Off-Site Nutzung, verbleiben folglich auch mehr Informationen im Datenmaterial.

Um die Attraktivität der On-Site-Nutzung von Mikrodaten zusätzlich zu erhöhen und die regionale Erreichbarkeit dieser Nutzungsform zu gewährleisten, richten die Forschungsdatenzentren derzeit neben den bereits bestehenden Gastwissenschaftlerarbeitsplätzen in Berlin, Bonn und Wiesbaden weitere Arbeitsplätze in allen Statistischen Landesämtern ein. An diesen können die Daten mit den gängigen statistischen Analyseprogrammen ausgewertet werden.

d) Nutzung amtlicher Mikrodaten durch kontrollierte Datenfernverarbeitung

Die Nutzung von geschützten Datenbeständen via Datenfernverarbeitung ist eine relativ neue Entwicklung, deren Bedeutung in der Zukunft zunehmen wird¹³⁾. Mittels dieses Verfahrens ist es der Wissenschaft möglich, das Informationspotential von nicht oder nur formal anonymisiertem Mikrodatenmaterial zu nutzen, ohne jedoch selbst direkten Zugriff auf diese Daten zu haben. Die Wissenschaftler entwickeln dabei Auswertungsprogramme (Syntax-Skripte), die dann von den Mitarbeitern der Forschungsdatenzentren an den Originaldaten angewendet werden. Diese Datendienstleistung wird zur Zeit in den Programmen SPSS, SAS und STATA angeboten. Die Datenfernverarbeitung ist im Gegensatz zu den Scientific Use Files nicht auf einen bestimmten Personenkreis beschränkt und bietet daher auch für ausländische Wissenschaftler und für den nicht-wissenschaftlichen Interessentenkreis eine Möglichkeit, Mikrodaten der amtlichen Statistik zu nutzen.

Zur praktikablen Anwendung der Datenfernverarbeitung stellen die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder Datenstrukturfiles zur Verfügung, die es den Nutzern ermöglichen, ihre Analyseprogramme auf die Struktur der Originaldaten abzustimmen. Diese Datenstrukturfiles geben die Datenstruktur des

12) Siehe hierzu Sturm, Roland: Wirtschaftsstatische Einzeldaten für die Wissenschaft. *Wirtschaft und Statistik* 2, 2002, 101–109 sowie die Ergebnisse des Workshops „Anonymisierung wirtschaftsstatischer Einzeldaten“ in Tübingen. Die Ergebnisse sind dokumentiert unter <http://www.uni-tuebingen.de/iaw/fawe-nutzertagung.html>.

13) Erste Erfahrungen wurden hier im Bereich der Steuerstatistiken gesammelt. Siehe hierzu Zwick, Markus: Steuerstatistische Einzeldaten und ihre Auswertungsmöglichkeiten für die Wissenschaft, *Allgemeines Statistisches Archiv* 83, 2/1999, S.248-253.

originären Datensatzes wieder, ohne inhaltliche Informationen zu transportieren. Das Material ist im Aufbau also in der Merkmalsstruktur, Anzahl der Satzstellen, Datensatzlänge identisch zum Originalmaterial. Durch ein technisches Verfahren sind die Daten aber so verfälscht, dass nur noch synthetische inhaltsleere Datensätze zur Verfügung stehen. Zurzeit stehen Datenstrukturfiles für die Mikrodaten des Mikrozensus und der Lohn- und Einkommensteuerstatistik zur Verfügung.

Derzeit ist die Datenfernverarbeitung ein relativ arbeitsaufwändiges Verfahren, da die Programmsyntax vorab auf Deanonymisierungsstrategien und der Datenoutput anschließend auf Geheimhaltungsfälle überprüft werden müssen. Bislang müssen diese Schritte manuell durchgeführt werden. Zwar liegen mittlerweile auch erste automatisierte Verfahren für diese Prüfungen vor, jedoch ist auch mit diesen Lösungsansätzen derzeit noch keine vollständige Automatisierung der Datenfernverarbeitung möglich.

Ein wesentliches Ziel der Forschungsdatenzentren besteht daher darin, Verfahren zu entwickeln, die eine stärkere Automatisierung der Datenfernverarbeitung erlauben. Ansatzpunkte hierfür bieten insbesondere die LIS/LES-Datenbank sowie das in Dänemark praktizierte Verfahren einer On-Line-Datenverarbeitung.¹⁴⁾ Die LIS/LES-Datenbank erlaubt den direkten Datenzugriff auf die Mikrodaten der Luxembourg Income Study/Luxembourg Employment Study über das Internet¹⁵⁾. Über die Angabe eines projektbezogenen Passworts können durch die Übermittlung von SAS-, SPSS- oder STATA-Files automatisch Datenauswertungen gestartet werden. LIS/LES ist so ausgestaltet, dass auch eine begrenzte Syntax- und Ergebnisüberprüfung erfolgt.

Das dänische Modell sieht hingegen vor, dass Wissenschaftler direkt von ihrem Arbeitsplatz aus auf einen Server zugreifen können, der von dem Statistischen Amt für Forschungszwecke eingerichtet und gepflegt wird¹⁶⁾. Die Wissenschaftler erhalten hierzu Zugriff auf ein Verzeichnis, in dem nur die Daten abgelegt sind, die sie für ihr

Forschungsvorhaben benötigen. Die Daten können von dem Wissenschaftler in ein für sie angelegtes Arbeitsverzeichnis kopiert werden; von dort werden auch die Analysen durchgeführt. Die Ergebnisse werden automatisch per E-Mail versendet. Der gesamte Prozess der Datenverarbeitung kann von dem Statistischen Amt jederzeit daraufhin überprüft werden, ob die vertraglich vereinbarten Regeln eingehalten werden.

Die hier im Einzelnen dargestellten Datennutzungswege der Forschungsdatenzentren können auch miteinander kombiniert werden. Unabhängig davon, welche Möglichkeit des Datenzugangs gewählt wird, unterliegt die Datenbereitstellung allerdings einer Zweckbindung. Dies bedeutet, dass die Mikrodaten nicht für allgemeine Forschungszwecke zur Verfügung gestellt werden können, sondern nur für ein definiertes und zeitlich begrenztes Forschungsvorhaben.

Sonderaufbereitungen

Neben den dargestellten Datennutzungswegen besteht für alle Nutzergruppen weiterhin die Möglichkeit, Sonderaufbereitungen in Auftrag zu geben. Hierbei handelt es sich um Datenauswertungen, die auf den speziellen Informationsbedarf eines bestimmten Nutzers zugeschnitten sind und nicht oder nur unzureichend mit den sonstigen Datennutzungswegen umgesetzt werden können. Im Gegensatz zur kontrollierten Datenfernverarbeitung werden die Auswertungsprogramme nicht von den Nutzern, sondern von der amtlichen Statistik erstellt. Hierzu wird der Datenbedarf im Dialog zwischen amtlicher Statistik und Nutzer soweit konkretisiert, dass ein Auswertungsprogramm erstellt werden kann. Anschließend wird das nicht oder nur formal anonymisierte Datenmaterial mit Hilfe des entwickelten Programms ausgewertet. Die Ergebnisse werden nach einer Geheimhaltungsprüfung übermittelt. Der Nutzer hat demnach keinen direkten Kontakt mit den nicht oder nur formal anonymisierten Mikrodaten.

14) Darüber hinaus bietet das Programm μ -Argus die Möglichkeit, Ergebnisse, die in Form von Tabellen vorliegen, maschinell geheimzuhalten. Das Programm wurde vom Statistischen Amt der Niederlande für Eurostat entwickelt und für die tabellenübergreifende Geheimhaltung vom Landesamt für Datenverarbeitung und Statistik Nordrhein-Westfalen erweitert. Es wird zurzeit in Deutschland ausführlich getestet.

15) Siehe hierzu ausführlich Smeeding, Timothy M./Jesuit, David K./Alkemade, Paul: The LIS/LES Project Databank: Introduction and Overview. In: Schmollers Jahrbuch, Zeitschrift für Wirtschafts- und Sozialwissenschaften, 122. Jg., 3, 2002, Seite 497-517.

16) Das Modell wird in dem Bericht „Access to Microdata“ beschrieben, der von dem Statistischen Amt Schwedens 2003 herausgegeben wurde.

Datenbedarf der Wissenschaft und Nutzungspräferenzen

Die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder streben an, ihr Dienstleistungsangebot dem wissenschaftlichen Bedarf entsprechend weiterzuentwickeln und befragen daher regelmäßig potenzielle Nutzer nach ihren Präferenzen. Um den konkreten Datenbedarf der Wissenschaft bei der Entwicklung des Dienstleistungsangebotes zu berücksichtigen, hat das Forschungsdatenzentrum der Statistischen Landesämter im Sommer 2002 eine Nutzerbefragung durchgeführt¹⁷⁾. Ziel der Umfrage war es, die potenziellen Nutzerinnen und Nutzer des künftigen Datenangebotes der Forschungsdatenzentren anzusprechen und ihren konkreten Datenbedarf zu ermitteln. Dabei konnten sich die Befragten auch zu den unterschiedlichen Nutzungsmöglichkeiten, den von ihnen eingesetzten Analyseprogrammen und ihrem Interesse an geplanten Veranstaltungen der beiden Forschungsdatenzentren äußern.

Von den siebenhundert befragten Wissenschaftlern gaben immerhin knapp 600 an, dass sie im Rahmen ihrer wissenschaftlichen Tätigkeit Mikrodaten nutzen oder zukünftig benötigen werden. Insgesamt weisen die Ergebnisse der Nutzerbefragung somit auf ein großes Interesse der Wissenschaft an der Nutzung von Mikrodaten der amtlichen Statistik hin. Da der in der Umfrage angegebene Datenbedarf fachlich sehr breit gestreut ist, werden sich die Forschungsdatenzentren nicht auf die Bereitstellung einzelner weniger Statistiken konzentrieren können, sondern müssen mittelfristig ein breites Datenangebot bereithalten.

Bezüglich der angebotenen Datennutzungswege weist die Befragung auf eine eindeutige Präferenz für die Nutzung von faktisch und total anonymisierten Daten am eigenen Arbeitsplatz hin, während die Arbeit an den Gastwissenschaftlerarbeitsplätzen und die kontrollierte Datenfernverarbeitung zum Befragungszeitpunkt kaum auf das Interesse der Wissenschaft stießen. Es wird jedoch nicht möglich sein, alle relevanten Erhebungen der amtlichen Statistik als anonymisierte Datensätze anzubieten. Die Forschungsdatenzentren planen daher, die Attraktivität der

alternativen Nutzungswege durch deren nutzungsgerechte Ausgestaltung deutlich zu steigern. Insbesondere die Einrichtung von Gastwissenschaftlerarbeitsplätzen in allen regionalen Standorten der Forschungsdatenzentren wird die regionale Erreichbarkeit des Dienstleistungsangebotes der amtlichen Statistik deutlich verbessern, so dass der Zugang zu den Mikrodaten wesentlich erleichtert wird.

Ausblick

Durch die beschriebene Weiterentwicklung der Dateninfrastruktur und die Einrichtung unterschiedlicher Datennutzungswege wird die Mikrodatenbasis für wissenschaftliche Analysezwecke entscheidend verbessert werden. Die Diskussion um den Zugang der Wissenschaft zu dem Informationspotenzial der öffentlichen Datenproduzenten ist jedoch mit der Einrichtung von Forschungsdatenzentren bei den unterschiedlichen öffentlichen Datenproduzenten nicht abgeschlossen. Zukünftige Fragen des Datenzugangs betreffen neben grundsätzlichen Regelungen insbesondere auch die Bereitstellung international vergleichbarer Mikrodaten.

Die Diskussion um eine grundsätzliche Regelung des Datenzugangs wird vom Gründungsausschuss des Rates für Sozial- und Wirtschaftsdaten unter dem Begriff des „Forschungsdatengeheimnisses“ geführt. Der Vorschlag des Gründungsausschusses sieht vor, dass Wissenschaftlerinnen und Wissenschaftler den Mitarbeiterinnen und Mitarbeitern der Datenproduzenten gleich gestellt werden und somit weitergehende Rechte zur Nutzung der Mikrodaten als bisher erhalten sollen. Gleichzeitig sollen ein Zeugnisverweigerungsrecht sowie ein Beschlagenahmeverbot den unbefugten Zugriff auf Mikrodaten, die bei den Wissenschaftlern gespeichert sind, verhindern.

Ein wesentliches Problem, Mikrodaten für die international vergleichende wissenschaftliche Forschung zu nutzen, besteht darin, dass in den einzelnen Ländern der Mikrodatenzugang sehr unterschiedlich geregelt ist¹⁸⁾. Dadurch ist die Beschaffung verschiedener nationaler Mikrodaten mit einem recht hohen Aufwand verbunden. Erste Ansätze

17) Siehe hierzu ausführlich Zühlke, Sylvia/ Hetke, Uwe: Datenbedarf und Datenzugang: Ergebnisse der ersten Nutzerbefragung des Forschungsdatenzentrums der statistischen Landesämter. In: Allgemeines Statistisches Archiv 3, S. 321-334.

18) Eine Darstellung der verschiedenen nationalen Ansätze zum Datenzugang findet sich in den Tagungsbeiträgen zum internationalen „Workshop on Microdata“ vom 21./22. August 2003 unter <http://www.mirco2122.scb.se>.

ze, den Datenzugang zu den Erhebungen der Europäischen Union zu vereinheitlichen, wurden mit den Verordnungen Nr. 322/97 und 831/2002 für Gemeinschaftsstatistiken unternommen¹⁹⁾. Mit dieser Verordnung wird die länderübergreifende Bereitstellung von Mikrodaten der Arbeitskräftestichprobe, des Europäischen Haushaltspanels, der Erhebung über die berufliche Weiterbildung und der Innovationserhebung der Gemeinschaft angestrebt. Die Umsetzung dieser Verordnung wird jedoch zurzeit dadurch erschwert, dass die dort vorgesehenen Regelungen teilweise

den nationalen Regelungen in den Ländern der Europäischen Union widersprechen.

Aufgrund der dargestellten Entwicklungen ist davon auszugehen, dass sich die Datenzugangsmöglichkeiten der Wissenschaft auch zukünftig weiter entwickeln werden. Die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder werden sich an diesen Prozess durch eigene Ausgestaltungsvorschläge weiterhin intensiv beteiligen.

19) Verordnung (EG) Nr. 831/2002 der Kommission vom 17. Mai 2002 zur Durchführung der Verordnung (EG) Nr. 322/97 des Rates über die Gemeinschaftsstatistiken - Regelung des Zugangs zu vertraulichen Daten für wissenschaftliche Zwecke