

Die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder

Zühlke, Sylvia; Zwick, Markus; Scharnhorst, Sebastian; Wende, Thomas

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Zühlke, S., Zwick, M., Scharnhorst, S., & Wende, T. (2005). Die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder. *ZA-Information / Zentralarchiv für Empirische Sozialforschung*, 56, 168-182. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-198599>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder

von **Sylvia Zühlke**, **Markus Zwick**, **Sebastian Scharnhorst** und **Thomas Wende**¹

Zusammenfassung

In den letzten Jahren wurde 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. Eine der zentralen institutionellen Empfehlungen dieser Kommission bestand in der möglichst raschen Einrichtung von Forschungsdatenzentren bei den öffentlichen 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 als gemeinsame Einrichtung aller Statistischen Landesämter mit 16 regionalen Standorten eingerichtet. Mit diesen Forschungsdatenzentren intensiviert die deutsche amtliche Statistik ihre bisherigen Bemühungen, Mikrodaten der amtlichen Statistik für wissenschaftliche Analysen zugänglich zu machen. Wesentliche Aufgabe der Forschungsdatenzentren ist es, die Dateninfrastruktur zu verbessern sowie den Zugang der Wissenschaft zu den Mikrodaten der amtlichen Statistik durch die Einrichtung mehrerer Zugangswege zu erleichtern.

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

Dieser Beitrag ist eine aktualisierte Fassung des erstmals in: *Wirtschaft und Statistik* 10/2003, S. 906 - 911, Statistisches Bundesamt, Wiesbaden veröffentlichten Artikels.

Abstract

In Germany, an intensive discussion has been going on over the last few years on granting the scientific community access to microdata of official statistics. Commissioned by the Federal Ministry of Education and Science, the Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI - Commission to improve the informational infrastructure by co-operation of the scientific community and official statistics) developed several proposals on how to improve the interaction between the scientific community and official statistics (cf. Kommission zur Verbesserung der informationellen Infrastruktur 2000). One of the central institutional recommendations of that Commission was that research data centres be set up as soon as possible at the location of data producers. That recommendation has been taken up by official statistics. In 2001, the Federal Statistical Office established such a research data centre. Another research data centre of the statistical offices of the Länder was set up in March 2002 as a joint facility of all statistical offices of the Länder with 16 regional locations. By establishing the research data centres, German official statistics has been intensifying its efforts to make official statistical micro-data accessible for scientific analyses. The major goal of the research data centres of the Federal Statistical Office and the statistical offices of the Länder is to facilitate access to microdata of official statistics for the scientific community by establishing various ways of data use. A major prerequisite for achieving that goal is a fundamental improvement of the data infra-structure by setting up a system (centralised in terms of subject-matter) of data maintenance for selected statistics and by establishing a metadata information system.

1 Einleitung

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 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 auch, 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.² Eine der zentralen institutionellen Empfehlungen dieser Kommission bestand in der möglichst raschen Einrichtung von Forschungsdatenzentren bei den öffentlichen 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 als gemeinsame Einrichtung aller Statistischen Landesämter mit 16 regionalen Standorten eingerichtet. Mit diesen Forschungsdatenzentren intensiviert die deutsche amtliche Statistik ihre bisherigen Bemühungen, Mikrodaten der amtlichen Statistik für wissenschaftliche Analysen zugänglich zu machen.

Ziel dieses Aufsatzes ist es, die neuen Formen und Wege der Nutzung amtlicher Mikrodaten, die sich aus der Einrichtung der Forschungsdatenzentren ergeben haben, vorzustellen. 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.

2 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 Gesetzes über die Statistik für Bundeszwecke (Bundesstatistikgesetz – BStatG) beeinflusst, das 1953 verabschiedet und in den Jahren 1980 und 1987 novelliert wurde.

Bei der Entstehung des Bundesstatistikgesetzes wurde die Übermittlung von Mikrodaten an die Wissenschaft kaum diskutiert, so dass in der Gesetzesfassung von 1953

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

hierzu noch keine explizite Regelung enthalten war. Aufgrund der fehlenden Möglichkeiten, umfangreiche Mikrodaten zu verarbeiten, war die Nachfrage nach Mikrodaten in den fünfziger, sechziger und frühen siebziger Jahren des letzten Jahrhunderts relativ gering. Sofern die Übermittlung von Einzelangaben nicht explizit in anderen Gesetzen geregelt war, wurden nur für wenige Projekte formal anonymisierte Mikrodaten³ 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 wurde.⁴ Für dieses 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 bereitgestellt.

Als die Nachfrage nach Mikrodaten mit der Fortentwicklung der Informationstechnik schnell anstieg, wurden die allgemeinen Regelungen des im Jahr 1977 verabschiedeten Bundesdatenschutzgesetzes über die Datenübermittlung mit der Novellierung des BStatG im Jahr 1980 für den speziellen Bereich der Statistik konkretisiert und präzisiert. Der Kreis möglicher Empfänger von Einzelangaben wurde dabei an die Bestimmungen des Strafgesetzbuches über die Strafbarkeit bei Verletzung von Privatgeheimnissen angepasst. Mit der Einführung der so genannten Weiterleitungsvorschrift des § 11 Abs. 5 BStatG 1980 wurde die Absicht verfolgt, den Zugang zu den amtlichen Mikrodaten allgemein zu regeln. Diese Vorschrift sah die Weitergabe von Mikrodaten in absolut anonymisierter Form an die Nutzer vor. Bei Anwendung dieser Möglichkeit innerhalb konkreter Projekte zeigte sich allerdings, dass die Anforderungen an ein solchermaßen absolut anonymisiertes Datenmaterial so restriktiv waren, dass es in der Folgezeit nur noch zu wenigen Datennutzungen kam. Aufgrund der gesetzlichen Vorgaben des BStatG konnten für nachfolgende Projekte wie „Vergleichende Analysen der Sozialstruktur mit Massendaten“ (VASMA) oder für den Sonderforschungsbereich 3 „Mikroanalytische Grundlagen der Gesellschaftspolitik“ der Deutschen Forschungsgemeinschaft nur noch

3 Formal anonymisiert bedeutet, dass aus dem Originaldatenmaterial direkte Identifikatoren wie etwa Name, Anschrift oder Sozialversicherungsnummer gelöscht wurden, es aber darüber hinaus unverändert ist.

4 Siehe: **Krupp, H.-J.**: Sozialpolitisches Entscheidungs- und Indikatoren-system für die Bundesrepublik Deutschland, Allgemeines Statistisches Archiv, Band 57, 1973, S. 380 ff.

absolut anonymisierte Mikrodaten zu relativ hohen Kosten zur Verfügung gestellt werden.⁵

Sowohl die informationelle Selbstbestimmung als auch die Wissenschaftsfreiheit sind als Grundrechte im Grundgesetz verankert. Damit ist der Gesetzgeber aufgerufen, für einen angemessenen Ausgleich dieser Grundrechte Sorge zu tragen. Bei der Novellierung des Bundesstatistikgesetzes im Jahr 1987, mit der das Recht der amtlichen Statistik an die Anforderungen des Volkszählungsurteils von 1983⁶ angepasst werden sollte, wurde dem Rechnung getragen, als nunmehr der Wissenschaft auch Daten übermittelt werden konnten, die eine Deanonymisierung zwar nicht mit Sicherheit ausschließen, aber Betroffenen nur zugeordnet werden können, wenn der Datenempfänger einen unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft erbringen musste. Die Einführung des Begriffs der faktischen Anonymität für die Weitergabe von Mikrodaten an die Wissenschaft erlaubte es nun, im Rahmen des so genannten 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 des letzten Jahrhunderts erste standardisierte faktisch anonymisierte Mikrodatensätze für den Bereich der Haushalts- und Personenerhebungen zur Verfügung zu stellen.⁷

Während hierdurch und durch weitere Aktivitäten außerhalb der amtlichen Statistik⁸ 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 zunächst keine vergleichbaren Fortschritte erzielt werden. Der fehlende Datenzugang, insbesondere zu den Unternehmens- und Betriebsdaten, wurde seitens

5 Zu den Projekten siehe **Hauser, R.** (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.

6 BVerfGE 65, 1.

7 Siehe hierzu insbesondere **Müller, W.; Blien, U.; Knoche, P.; Wirth, H.** u. a.: Die faktische Anonymität von Mikrodaten, Band 19 der Schriftenreihe „Forum der Bundesstatistik“, Statistisches Bundesamt (Hrsg.), Wiesbaden 1991.

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

der Wissenschaft in dem Memorandum „Erfolgsbedingungen empirischer Wirtschaftsforschung und empirisch gestützter wirtschafts- und sozialpolitischer Beratung“⁹ aufgegriffen. 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“¹⁰ im Jahr 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 Statistik. 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. Derzeit werden die Empfehlungen der Kommission zur Verbesserung der informationellen Infrastruktur vom Rat für Sozial- und Wirtschaftsdaten umgesetzt. Eine Reihe von Forschungsdatenzentren und Servicezentren konnten ihre Arbeit bereits aufnehmen.¹¹

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 nutzungsgerechtes 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 für ihr Projekt einen Förderantrag

9 Siehe *Hauser, R.; Wagner, G.; Zimmermann, K.-F.*: Erfolgsbedingungen empirischer Wirtschaftsforschung und empirisch gestützter wirtschafts- und sozialpolitischer Beratung: Ein Memorandum, Allgemeines Statistisches Archiv, Band 82, 1998, S. 369 ff.

10 Die Ergebnisse des Symposiums sind dokumentiert in *Müller, W.; Schimpl-Neimanns, B.; Krupp, H.-J.; Wiegert, R.* u. a.: Kooperation zwischen Wissenschaft und amtlicher Statistik – Praxis und Perspektiven, Band 34 der Schriftenreihe „Forum der Bundesstatistik“, Statistisches Bundesamt (Hrsg.), Wiesbaden 1999.

11 Für einen Überblick siehe *Lüttinger, P.; Schimpl-Neimanns, B.; Wirth, H.* und *Papastefanou, G.*: Mikrodaten (German Microdata Lab): Das Servicezentrum für amtliche Mikrodaten bei ZUMA, ZUMA Nachrichten Nr. 52, 2003, S. 153 ff.

beim Bundesministerium für Bildung und Forschung gestellt. Mit Hilfe dieser Mittel soll die Infrastruktur für ein umfangreiches Dienstleistungs- und Datenangebot geschaffen werden.

3 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 Erreichung 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 Statistiken dezentral in den Statistischen Landesämtern durchgeführt, so dass dort über 90 % 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, bauen die Statistischen Landesämter eine fachlich zentralisierte Datenhaltung auf. Hierdurch wird es möglich, 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, die Aufbereitung und die Qualität der Daten. Hierfür 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.

4 Datennutzungswege

Um der Wissenschaft den Zugang zum gesamten Informationspotenzial der amtlichen Statistik zu öffnen, haben die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder unterschiedliche Zugangswege zur Nutzung ihrer Mikrodaten eingerichtet. 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 Re-identifikation 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.

4.1 Absolut anonymisierte Mikrodatensätze

Absolut anonymisierte Daten werden durch Aggregation oder durch die Entfernung einzelner Merkmale so weit verändert, dass eine Identifizierung der Auskunftgebenden nach menschlichem Ermessen unmöglich gemacht wird. Die amtliche Statistik bietet absolut anonymisierte Mikrodaten in Form so genannter 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. Die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder arbeiten verstärkt an einer Ausweitung dieses Angebotes. Eine weitere Hauptzielrichtung der Public Use Files liegt im Bereich der Hochschullehre. Die Forschungsdatenzentren entwickeln derzeit so genannte Campus-Files, die an Hochschulen zu Lehrzwecken eingesetzt werden können. Diese Datensätze sollen Studentinnen und Studenten schon frühzeitig die Möglichkeit bieten, die Besonderheiten der Analyse von amtlichen Mikrodaten kennen zu lernen. Erste Campusfiles liegen für den Mikrozensus, die Sozialhilfestatistik und die Kostenstrukturerhebung vor.

4.2 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 anonym 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 können.¹² Nach Maßgabe des Bundesstatistikgesetzes können faktisch anonymisierte Daten allerdings nur wissenschaftlichen Einrichtungen und nur zur Durchführung wissenschaftlicher Vorhaben zugänglich gemacht werden.

¹² Diese Regelung geht zurück auf den § 16 Abs. 6 BStatG.

Die Hauptzielrichtung der faktischen Anonymisierung besteht 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 kommen.¹³

Faktische Anonymität resultiert allerdings 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 Informationseinbußen 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. Da jedoch allein durch die Herausgabe dieser Daten ein höheres Deanonymisierungsrisiko besteht als bei der Nutzung in einem statistischen Amt, ist die Datenanonymisierung relativ stark 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, der Zeitbudgeterhebung sowie der Lohn- und Einkommensteuerstatistik bereits ein breites Datenspektrum als Scientific Use Files an. Im Rahmen des Projektes „Faktische Anonymisierung wirtschaftsstatistischer Einzeldaten“ sind darüber hinaus erste Scientific Use Files für wirtschaftsstatistische Einzeldaten entwickelt worden, und zwar für die Einzelhandelsstatistik, die Umsatzsteuerstatistik sowie die Kostenstrukturerhebung.¹⁴

13 Ein Überblick über Anonymisierungsmethoden ist zu finden in: **Köhler, S.**: Anonymisierung von Mikrodaten in der Bundesrepublik und ihre Nutzung – Ein Überblick in Band 31 der Schriftenreihe Forum der Bundesstatistik, Statistisches Bundesamt (Hrsg.), 1999, S. 133 ff.

14 Siehe hierzu **Sturm, R.**: Wirtschaftsstatistische Einzeldaten für die Wissenschaft. In: WiSta 2/2002, S. 101 ff., **Lenz, R.**; **Vorgrimler, Dr.** und **Rosemann, M.**: Ein Scientific-Use-File der Kostenstrukturerhebung im Verarbeitenden Gewerbe in WiSta 2/2005, S. 91ff., **Scheffler, M.**: Ein Scientific-Use-File der Einzelhandelsstatistik 1999 in WiSta 3/2005, S.197ff., **Vorgrimler, Dr.**;

Die Forschungsdatenzentren des Bundes und der Länder sind bestrebt, das Angebot an Scientific Use Files sukzessive zu erweitern. So laufen zurzeit Projekte zur Anonymisierung der Gehalts- und Lohnstrukturerhebung sowie der Diagnosestatistik.

4.3 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 gesamten Ergebnisse einer Statistik anonymisiert werden, sondern lediglich die daraus 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 so genannten Gastwissenschaftlerarbeitsplätzen ausgewertet werden. Da die Mikrodaten dabei in den Räumlichkeiten der amtlichen Statistik verbleiben und kaum mit Zusatzinformationen kombiniert werden können, hat diese Nutzungsform einen weiteren wesentlichen Vorteil: Faktische Anonymität wird hier bereits bei wesentlich geringeren Veränderungen im Datenmaterial erreicht als bei der Erstellung von Scientific Use Files für die Off-Site-Nutzung und folglich verbleiben mehr Informationen im Datenmaterial.

Um die Attraktivität dieser so genannten On-Site-Nutzung von Mikrodaten zusätzlich zu erhöhen und die regionale Erreichbarkeit dieser Nutzungsform zu gewährleisten, haben die Forschungsdatenzentren Gastwissenschaftlerarbeitsplätze in allen 19 regionalen Standorten eingerichtet. An diesen Arbeitsplätzen können die Daten mit gängigen statistischen Analyseprogrammen (SAS, SPSS, teilweise STATA) ausgewertet werden.

4.4 Nutzung amtlicher Mikrodaten durch kontrollierte Datenfernverarbeitung

Die Datennutzung via Datenfernverarbeitung ist eine relativ neue Entwicklung, deren Bedeutung in der Zukunft zunehmen wird.¹⁵ Mit diesem Verfahren ist es der Wissenschaft möglich, das Informationspotenzial von 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 zurzeit 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 Auswertungsprogramme auf die Struktur der Originaldaten abzustimmen. Diese Datenstrukturfiles geben die Datenstruktur des originären Datensatzes wieder, ohne inhaltliche Informationen zu transportieren. Das Material ist also in der Merkmalsstruktur, Anzahl der Satzstellen und Datensatzlänge identisch mit dem Originalmaterial. Durch ein technisches Verfahren sind die Daten aber so verfälscht, dass nur noch synthetische inhaltsleere Datensätze zur Verfügung stehen. Zurzeit stehen solche Datenstrukturfiles für die Mikrodaten des Mikrozensus, der Gehalts- und Lohnstrukturerhebung 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 werden diese Schritte noch manuell durchgeführt. 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.

15 Erste Erfahrungen wurden hier im Bereich der Steuerstatistiken gesammelt. Siehe hierzu *Zwick, M.*: Steuerstatistische Einzeldaten und ihre Auswertungsmöglichkeiten für die Wissenschaft. Allgemeines Statistisches Archiv, Band 83, 1999, S. 248 ff.

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 Online-Datenverarbeitung.¹⁶ Die LIS/LES-Datenbank erlaubt den direkten Datenzugriff auf die Mikrodaten der Luxembourg Income Study/Luxembourg Employment Study über das Internet.¹⁷ 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 vom statistischen Amt für Forschungszwecke eingerichtet und gepflegt wird.¹⁸ 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 den Wissenschaftlern 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.

Mit den Gastwissenschaftlerarbeitsplätzen und der Datenfernverarbeitung stehen der Wissenschaft, neben den Scientific Use Files, zwei weitere Möglichkeiten für den Zugang zu Mikrodaten der amtlichen Statistik offen. Diese können auch miteinander kombiniert werden. Unabhängig davon, welche Möglichkeit des Datenzugangs gewählt wird, unterliegt die Datenbereitstellung, abgesehen von den Public Use Files, 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.

16 Darüber hinaus bietet das Programm τ -ARGUS die Möglichkeit, Ergebnisse, die in Form von Tabellen vorliegen, maschinell geheim zu halten. Das Programm wurde vom Statistischen Zentralamt 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.

17 Siehe hierzu ausführlich *Smeeding, T. M.; Jesuit, D. K* und *Alkemade, P.*: The LIS/LES Project Databank: Introduction and Overview. In: Schmollers Jahrbuch, Zeitschrift für Wirtschafts- und Sozialwissenschaften, 122. Jg., Heft 3, 2002, S. 497 ff

18 Das Modell wird in dem Bericht Access to Microdata beschrieben, der vom Statistischen Amt Schwedens 2003 herausgegeben wurde.

5 Sonderaufbereitungen

Neben den dargestellten Datennutzungswegen besteht für alle Nutzergruppen weiterhin die Möglichkeit, gegen Entgelt 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 formal anonymisierte Datenmaterial mit Hilfe des entwickelten Programms in den statistischen Ämtern ausgewertet. Die Ergebnisse werden nach einer Geheimhaltungsprüfung an die Nutzer übermittelt. Die Nutzer haben hierbei keinerlei direkten Kontakt mit den Mikrodaten.

6 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 durchgeführt.¹⁹ 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 700 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

¹⁹ Siehe hierzu ausführlich *Zühlke, S.* und *Hetke, U.*: Datenbedarf und Datenzugang: Ergebnisse der ersten Nutzerbefragung des Forschungsdatenzentrums der Statistischen Landesämter. In: Allgemeines Statistisches Archiv, Band 87, 2003, S. 321 ff. (erscheint demnächst).

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 absolut 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 verbessert die regionale Erreichbarkeit des Dienstleistungsangebots der amtlichen Statistik deutlich, so dass der Zugang zu den Mikrodaten wesentlich erleichtert wird.

7 Ausblick

Durch die beschriebene Weiterentwicklung der Dateninfrastruktur und die Einrichtung unterschiedlicher Datennutzungswege wird die Mikrodatenbasis für wissenschaftliche Analysezwecke entscheidend verbessert. 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 Rat für Sozial- und Wirtschaftsdaten unter dem Begriff des „Forschungsdatengeheimnisses“ geführt. Der Vorschlag des Rates 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 Beschlagnahmeverbot den unbefugten Zugriff Dritter auf die bei den Wissenschaftlern gespeicherten Mikrodaten 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 ist.²⁰ Dadurch ist die Heranziehung verschiedener nationaler Mikrodaten für die Wissenschaft mit einem recht hohen Aufwand verbunden. Erste Ansätze, den Datenzugang zu den Erhebungen der Europäischen Union zu vereinheitlichen, wurden mit den Verordnungen Nr. 322/97 und darauf aufbauend 831/2002 für Gemeinschaftsstatistiken unternommen.²¹ Mit diesen Verordnungen 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 Verordnungen wird jedoch zurzeit dadurch erschwert, dass die dort vorgesehenen Regelungen teilweise noch 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 weiterentwickeln werden. Die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder werden sich an diesem Prozess durch eigene Ausgestaltungsvorschläge weiterhin intensiv beteiligen.

20 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.micro2122.scb.se>.

21 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 (Amtsbl. der EG Nr. L 133 vom 18. Mai 2002, S. 7 ff.).