

91. DGINS-KONFERENZ

KOPENHAGEN, 26. UND 27. MAI 2005

**Bereitstellung von Statistiken
für die Forschung**

*Pedro DÍAZ MUÑOZ
Direktor für Methoden und statistische Instrumente
Eurostat, Europäische Kommission*

Bereitstellung von Statistiken für die Forschung

Pedro DÍAZ MUÑOZ
Direktor für Methoden und statistische Instrumente
Eurostat, Europäische Kommission

Die Bereitstellung statistischer Daten für die Forschung ist mit einigen Schwierigkeiten verbunden. Die wichtigsten Faktoren, durch deren Widersprüchlichkeit diese Schwierigkeiten entstehen, sind: das Risiko der Identifizierung von Einzelpersonen, der sehr große ungenutzte Informationsgehalt der Daten sowie die Wahrnehmung der Gesellschaft sowohl als Datenquelle als auch als Datennutzer. In diesem Dokument werden nach einer Analyse der oben genannten Faktoren die auf EU-Ebene getroffenen Maßnahmen zur Erleichterung des Datenzugriffs für die Forschungsgemeinde erörtert sowie ein Maßnahmenbündel zur zukünftigen Verbesserung der Situation vorgeschlagen.

1. Der Informationsgehalt statistischer Daten

1.1 *Informationsgehalt statistischer Daten und tatsächliche Veröffentlichungen: Anspruch und Wirklichkeit*

Statistische Daten enthalten viel mehr Informationen als die von den Statistikämtern in der Regel veröffentlichten Dokumente. Betrachten wir zunächst den Fall einer Stichprobenerhebung. Grundsätzlich können alle möglichen Kreuztabellierungen der erfassten kategorisierten Variablen erstellt werden. Dies bedeutet, dass die Anzahl aller möglichen aggregierten Datenfelder als Produkt aller möglichen Ausprägungen der kategorisierten Merkmale ermittelt werden kann. Die Veröffentlichung von Daten mit einer solchen Detailtiefe unterliegt zweierlei Arten von Beschränkungen. Auf der einen Seite verhindert die Größe der Stichprobenfehler die Bestimmung von Indikatoren anhand von Domänen von einem zu geringeren Umfang, da ihre Genauigkeit gering und die Werte damit irreführend wären. Auf der anderen Seite ist das Risiko der Identifikation von Einzelpersonen umso höher, je geringer die Anzahl der statistischen Einheiten in der Stichprobe ist. Eine sekundäre Nutzung der Mikrodaten ist daher mit einem zweifachen Risiko verbunden. Die Identifikation von Einzelpersonen sowie falsche Schlussfolgerungen aufgrund fehlender Genauigkeit der produzierten Informationen.

Der zweite Fall betrifft Daten, die anhand einer umfassenden Volkserhebung, beispielsweise einer Volkszählung oder eines Registers gewonnen werden. Auch hier können die Möglichkeiten, aggregierte Daten zu produzieren, sehr umfassend sein. Ihre Anzahl kann als die Anzahl der Felder in allen möglichen Kreuztabellierungen der kategorisierten Variablen definiert werden. Theoretisch

sind hier keine Probleme mit der Genauigkeit zu erwarten, auch wenn einige Antwortfehler auftreten können, die sich hauptsächlich auf die Felder auswirken, in denen nur wenige Personen erfasst wurden. Die Probleme mit der Identifizierung sind allerdings noch gravierender als im vorigen Fall, denn alle erfassten Einträge beziehen sich mit Sicherheit auf eine bestimmte Einzelperson.

Eine weitere, wichtige charakteristische Informationseigenschaft statistischer Daten stellt ihr durch die Verknüpfung verschiedener Datenquellen auf der Mikroebene unter Verwendung gemeinsamer identifizierbarer Variablen gegebenes Potential dar. In diesem besonderen Fall tritt zu den bereits erwähnten Risikofaktoren ein zusätzliches Risiko hinzu. Es handelt sich um die dem Datenabgleich zugrunde liegende Hypothese sowie die möglicherweise fehlende Kohärenz der Datenquellen. Es sei darauf hingewiesen, dass in manchen Fällen einige der für den Abgleich verwendeten Datenquellen nicht für statistische Zwecke erstellt worden sind, so dass ihre Qualität für den Zweck einer statistischen Untersuchung möglicherweise nicht ausreichend ist.

Die oben angeführten Fälle demonstrieren, dass die aus statistischen Daten extrahierbare Datenmenge die Menge der von einem Statistikamt normalerweise unter Verwendung aller Verbreitungsmittel veröffentlichten Daten weit übersteigt. Unter Einhaltung entsprechender Sicherheitsvorkehrungen zur Vermeidung der dargestellten Risiken ist eine der Aufgaben der Statistikämter, eine sekundäre Nutzung der Daten zwecks Verbesserung ihrer Verwendbarkeit zu ermöglichen.

1.2 Beispiele für die Sekundärnutzung statistischer Informationen

Glücklicherweise ist die Verwendungspalette statistischer Daten sehr breit. Der Versuch, eine umfassende Beschreibung zu liefern, ist aussichtslos. Ich werde mich hier daher auf einige wenige Beispiele beschränken.

Das Protokoll des 19. CEIES-Seminars vom 26. und 27. September 2002 in Lissabon - Innovative Lösungen für den Zugang zu Mikrodaten [1] - enthält einige Beispiele für die Verwendung statistischer Daten zu Forschungszwecken. Drei Fälle verdienen hier eine besondere Beachtung. Zunächst demonstriert Richard Blundell, wie die Verwendung individueller Längsschnittdaten zu Gesundheit, Altersstruktur und Rentenalter die Analyse potentieller Effekte des Alterns der Bevölkerung sowie die ökonomischen Auswirkungen des damit verbundenen steigenden Bedarfs an Gesundheits- und Sozialdienstleistungen erleichtern kann. Die in diesem Fall verwendeten Daten entsprechen der English Longitudinal Survey on Ageing (ELSA – Englische Längsschnitterhebung zur Altersstruktur).

Zweitens gibt Robert Erikson ein interessantes Beispiel für einen Abgleich dreier Datenquellen zur Erfassung der Verteilung durchschnittlicher Noten bei Schülern aus der Arbeiterklasse nach Abschluss der Pflichtschulzeit sowie der Möglichkeiten eines Übergangs zu einer weiter führenden Schule. Bei den verwendeten Quellen handelte es sich um Angaben zum Elternberuf aus der Volkszählung aus dem Jahre 1990, Durchschnittnoten aus einem Datensatz über Schulabgänger sowie Informationen über Sekundarstufe II aus einer Datei der Zugangsberechtigungen für diese Ausbildungsstufe.

Drittens betont Niels Westergaard-Nielsen die Vorteile einer Verwendung von verknüpften Arbeitgeber- und Arbeitnehmerdaten, mit deren Hilfe eine Analyse der Unterschiede zwischen den Unternehmen sowie der Reaktionen auf verschiedene politische Maßnahmen ermöglicht wird. Auf der Grundlage der verknüpften Daten können politisch relevante Fragen im Hinblick auf die Schaffung und Verlust von Arbeitsplätzen beantwortet werden. Darüber hinaus lassen sich anhand der Daten die Auswirkungen der politischen Maßnahmen auf die Unternehmen und ihren Arbeitskräftebedarf untersuchen. In dem Dokument wird eine überaus interessante Beschreibung der Anforderungen an die für eine Verknüpfung der Arbeitgeber- und Arbeitnehmerdaten geeigneten Datenquellen geliefert.

Darüber hinaus wird in der nachfolgenden Tabelle eine Übersicht über die Projekte der Forschungsinstitute geliefert, welche im Jahre 2004 bei Eurostat die Mikrodaten des Europäischen Haushaltspanels (ECHP) angefordert hatten.

Forschungsverträge unter Verwendung der ECHP-Daten 2004 - Hauptthemen	
<i>Untersuchungen spezieller Teilgesamtheiten</i>	<i>Untersuchungen spezieller Phänomene</i>
<ul style="list-style-type: none"> • ältere Menschen • Arme • Regionen • Langzeitarbeitslose • Verheiratete Frauen • Erwerbstätigkeit von Frauen • Geschiedene • Zeitarbeitskräfte • Personen kurz vor dem Ruhestandsalter • Jugendliche 	<ul style="list-style-type: none"> • Mobilität • Einkommensungleichheit • Übergang zwischen Beschäftigung und Arbeitslosigkeit • Steuern, Subventionen • intrafamiliärer Transfer • Einkommens- und Ausbildungsungleichheiten • Lohnänderungen • Ausbildung und Gesundheit

	<ul style="list-style-type: none"> • Teilnahme am Arbeitsmarkt und Geburtenrate • Kinderbetreuung • Diskriminierung
--	--

Ein interessantes Beispiel für die Verwendung von Mikrodaten stellt EUROMOD dar - ein Steuer-Sozialleistungsmodell auf der Grundlage von Haushalts-Mikrodaten. Dieses Modell ist dazu angelegt, die Auswirkungen von Veränderungen in der Sozial- und Steuerpolitik anhand von Maßzahlen des persönlichen Einkommens und der Haushaltswohlfahrt zu ermitteln. Das Modell wurde zur Evaluierung europäischer und nationaler politischer Maßnahmen verwendet. Weitere Informationen finden Sie unter: <http://www.econ.cam.ac.uk/dae/mu/emod.htm>.

Die oben angeführten Beispiele sollten nicht den Eindruck erwecken, die Forscher seien ausschließlich an Mikrodaten interessiert. Es gibt eine wichtige Gruppe von Forschern, deren hauptsächlichstes Interesse der Analyse der wirtschaftlichen Entwicklung sowie der Klärung von Beziehungen zwischen den Mikroeinheiten gilt. Für ihre Arbeit benötigen sie große Mengen roher historischer Zeitreihen, also Daten ohne jegliche Aufbereitung, wie beispielsweise Saison- oder handelstägliche Bereinigungen, die sich auf das zu extrahierende Grundsignal auswirken könnten.

1.3 *Missbrauch statistischer Daten*

Glücklicherweise gibt es nicht viele Beispiele für einen Missbrauch statistischer Daten. Dies kann möglicherweise darauf zurückzuführen sein, dass die durch die statistischen Einrichtungen ergriffenen Schutzmaßnahmen ausreichend sind. Ein weiterer Grund könnte darin bestehen, dass die Erfassung persönlicher Daten ausreichenden Umfangs anhand von statistischen Informationen für eine gewinnbringende Verwendung nicht einfach genug zu bewerkstelligen ist. Es kann trotzdem von Nutzen sein, die Beschreibung eines fiktiven Missbrauchs statistischer Daten zu liefern, denn ein solcher Fall könnte das Vertrauen der Gesellschaft in die statistischen Systeme erheblich schädigen, so dass alle verfügbaren Mittel eingesetzt werden sollten, um dies zu verhindern. Nachfolgend wird ein solches Szenario vorgestellt. Beachten Sie, dass in den nachfolgenden Szenarien über die Offenlegung von Informationen hinaus auch Fälle in Verbindung mit irreführender Information oder fehlerhafter Datenbehandlung dargestellt werden.

- Eine Auflistung von Personen mit einem hohen Einkommensniveau kann anhand eines Personenregisters mit wirtschaftlichen Informationen, beispielsweise Steuerregister, zusammengestellt werden. Diese Register könnten anschließend mit anderen statistischen

Datensätzen, wie Volkszählungstichproben oder Wirtschaftsrechnungen privater Haushalte verknüpft werden.

- Journalistische Angriffe gegen Prominente (auf der Grundlage des Rechtes auf Information). Beispielsweise nehmen einige Tageszeitungen in Finnland die Mühe auf sich, die Daten in Bezug auf die vermögendsten Personen des Landes zu verarbeiten (und im Internet zu veröffentlichen). Üblicherweise handelt es sich hier um die Namen der tausend Personen mit dem höchsten Einkommen sowie genaue Angaben zu ihrem Jahresverdienst sowie zu Einkünften aus Kapital- und Vermögenseinkommen. Lokale Tageszeitungen¹ haben sämtliche verfügbaren Steuerdaten aller Einwohner der Region veröffentlicht. Die gleichen Einkommensvariablen werden auch in den finnischen EU-SILC-Daten verwendet).
- Szenario „neugieriger Nachbar“: der Versuch, die Daten eines Nachbarn auf der Grundlage einiger weniger Variablen (Wohnraumtyp, Anzahl der Personen, Geschlecht, Alter, Beruf ...) in einer Mikrodatei zu ermitteln.
- Verknüpfung von Mikrodaten oder sehr genauer ökonomischer Informationen mit einem Personenregister mit dem Ziel, eine Datei zu Marketing-Zwecken zu erstellen.
- Erfassung von Verhältniszahlen von Wirtschaftsinformationen über Unternehmen mit dem Ziel, Sektoren zu bestimmen, für die eine spezifische Untersuchung einzelner Unternehmen zu Verwaltungszwecken ausgeführt werden kann.
- Falsche Interpretation statistischer Daten durch Personen, denen die Grundlagen der Statistik nicht geläufig sind (beispielsweise die Herleitung eines Kausalzusammenhangs bei einer falschen Korrelation).
- Inkonsistenzen zwischen Statistiken, die anhand von fehlerhaften veröffentlichten Mikrodaten gewonnen werden, und den offiziellen aggregierten Statistiken.
- Erstellung aggregierter Daten auf der Grundlage zu weniger Beobachtungen, wodurch die Daten wertlos werden.

¹ Diese Tageszeitungen wurden nach Maßgabe des Gesetzes zum Schutz personenbezogener Daten gerichtlich belangt!

1.4 Schutz vertraulicher Daten

Zum Thema Datenschutz und Analyse des Risikos der Datenoffenlegung gibt es eine ausführliche wissenschaftliche Literatur. Ich verweise auf die Sitzung der Arbeitsgruppe „Vertraulichkeit statistischer Daten“ vom 7. bis zum 9. April 2003 in Luxemburg [2], auf der die meisten Methoden vorgestellt wurden. Erwähnenswert ist ebenfalls das kürzlich abgeschlossene CASC-Projekt (Datenverarbeitungsaspekte der statistischen Geheimhaltung) innerhalb des Forschungs-Rahmenprogramms Nr. 5 [3]. In der unten aufgeführten Tabelle wird ein schematischer Überblick über die verschiedenen gegenwärtig verwendeten Methoden gegeben.

Methoden zum Schutz vertraulicher Mikrodatenfiles	
• Globale Neucodierung	Mehrere Merkmalsausprägungen werden zu einer zusammengefasst.
• Lokale Unterdrückung	Ein oder mehrere Wert(e) in einer unsicheren Kombination wird (werden) durch einen fehlenden Wert ersetzt.
• Top- und Bottom-Codierung	Global Recoding im Falle von Rangmerkmalen.
• Randomisierung nach der Datenerhebung	Beabsichtigte Fehlklassifikation durch die Wertänderung mindestens einer kategorisierten Variablen.
• Mikroaggregation	Ersetzung individueller Werte durch Werte, die anhand kleinerer Aggregate berechnet werden.
• Rausch-Aggregation	Hinzufügen eines willkürlichen Rauschens zu den quantitativen Daten.
• Datentausch	Austauschen einiger Variablen zwischen zwei Registern, die gemeinsame Kategorien für vordefinierte Variablen aufweisen.

2. Bereitstellung von Daten an die Forscher

2.1 Kompromisse

In dem vorliegenden Abschnitt wird das bisher Gesagte zusammengefasst. Bei der Bereitstellung von Daten an die Forscher müssen einige Kompromisse berücksichtigt werden:

(1) Das Risiko der Identifikation von Daten und Datensensibilität. Man kann zwei Typen von Identifikationsrisiken unterscheiden:

- Das oben dargestellte Szenario „Neugieriger Nachbar“: Mit einigen wenigen bekannten Charakteristika einer Meldeeinheit kann es zu dem Versuch einer Identifizierung und anschließender Offenlegung der Informationen kommen.
- Angriff auf ein Register: Die kommerziellen Registerdatenbanken sind frei zugänglich. Dadurch wird die Identifizierung einer Meldeeinheit (Name, Adresse) anhand einiger Schlüsselvariablen, beispielsweise Geschlecht, Aktivität, Region usw. ermöglicht. Der Abgleich zwischen dem Register und der Mikrodatei erlaubt möglicherweise die Identifizierung einer beträchtlichen Anzahl von Auskunftspersonen für Statistiken.

Gegenwärtig werden die Gefahren des zweiten Risikotyps als gravierender eingeschätzt. Grund hierfür ist die Tatsache, dass die aktuellen Technologien die Entwicklung paralleler Datenbanken für nicht statistische Zwecke ermöglichen.

Der Schutz vertraulicher Daten wird mit Hilfe von Rechtsvorschriften sichergestellt, die Bestimmungen über die Sicherung statistischer Daten gegen Offenlegung beinhalten. Die Rechtsvorschriften legen jedoch nur allgemeine Regeln fest. Die praktische Interpretation dieser Regeln fällt von Land zu Land unterschiedlich aus, so dass das Schutzniveau entsprechend stark variiert.

(2) Kapazitäten für die Analyse der Informationen. Wie bereits unter Punkt 1.1 weiter oben erwähnt, sind sich die statistischen Einrichtungen der Tatsache bewusst, dass die von ihnen veröffentlichten Angaben lediglich einen kleinen Ausschnitt der in den Daten enthaltenen Informationen darstellen. Es liegt in der Verantwortung der statistischen Einrichtungen, eine sekundäre Verwendung dieser umfassenden Informationen zu ermöglichen und damit speziellen Bedarf zu decken.

(3) Wahrnehmung der Privatsphäre in der Gesellschaft. Das tatsächliche Risiko der Identifizierung wird von der gesellschaftlichen Sicht der bestehenden Risiken überlagert. Die rechtlichen Bestimmungen zur Sicherung statistischer Daten gegen Offenlegung mögen ausreichend oder sogar übertrieben sein. In der Wahrnehmung der Meldeeinheiten ist das Risiko eines Datenmissbrauchs sehr hoch. Dies könnte sich nachteilig auf die Qualität der Antworten auswirken, da die Meldeeinheiten möglicherweise in der Zukunft versuchen könnten, einige Charakteristika, deren Identifizierung sie vermeiden möchten, vorzuenthalten oder zu verzerren.

(4) Gesellschaftliches Interesse an Informationen. In vielen Fällen sind detaillierte statistische Analysen in Bezug auf kleine Populationen oder Gebiete von hohem Interesse für politische

Überlegungen. Wie in vielen anderen Fällen kann es hier zu einem Widerspruch kommen. Die Menschen sind über den Schutz der Privatsphäre in Bezug auf ihre Antworten besorgt, gleichzeitig sind sie unzufrieden, wenn diese Informationen nicht vollständig ausgewertet werden und damit der gesellschaftliche Bedarf nicht erkannt oder das Ungleichgewicht in der Verteilung von Reichtum bzw. die Unzulänglichkeiten der öffentlichen Dienstleistungen nicht behoben werden können.

2.2 Zugang zu Mikrodaten für wissenschaftliche Zwecke in der Europäischen Union. Verordnung (EG) Nr. 831/2002 der Kommission

Zur Berücksichtigung des Forschungsbedarfs in der EU sind im Rahmen der grundsätzlichen Rechtsvorschriften zum Thema Datenvertraulichkeit zwei Instrumente entwickelt worden (Verordnung Nr. 1588/90 und Nr. 322/97 des Rates). Es handelt sich hierbei um den Ausschuss für die statistische Geheimhaltung mit Durchführungsbefugnissen in allen Fragen der Geheimhaltung sowie die Verordnung Nr. 831/2002 der Kommission über die Regelung des Zugangs zu vertraulichen Daten für wissenschaftliche Zwecke. Eine detaillierte Beschreibung und Analyse dieses Rechtsaktes finden Sie in dem von John King und Jean Louis Mercy auf der Sitzung der Arbeitsgruppe „Vertraulichkeit statistischer Daten“ in Luxemburg vom 7. bis zum 9. April 2003 vorgestellten Arbeitspapier. Die Verordnung hat zwar hohe Erwartungen an die Verfügbarkeit von Mikrodaten für die Forschung geweckt, ihre Umsetzung war jedoch mit einigen Schwierigkeiten verbunden, so dass die Fortschritte bei der Entwicklung relativ langsam vonstatten gingen.

Der Ausschuss für die statistische Geheimhaltung hat auf seiner Sitzung vom Dezember 2004 eine Analyse der bei der Umsetzung dieser Verordnung erzielten Fortschritte geliefert und eine Einigung über die Entwicklung von Schnellverfahren zur Verarbeitung der Forscheranfragen sowie für Zulassungsverfahren für Forschungsinstitute erzielt. Gegenwärtig können die Mikrodaten nur für zwei statistische Bereiche bereitgestellt werden. Hierbei handelt es sich um das Europäische Haushaltspanel (ECHP) und, seit kurzem, um die Arbeitskräfteerhebung (AKE). Darüber hinaus werden gegenwärtig im Rahmen der Arbeitsgruppe „Innovationserhebung der Gemeinschaft“ die Kriterien für die Bereitstellung der Mikrodaten der Erhebung erörtert. Ferner wurde zum gleichen Zweck eine Taskforce für die kommende Erhebung über die Lebens- und Arbeitsbedingungen (EU-SILC) eingesetzt. Parallel hierzu wurden die Ressourcen mit dem Ziel aufgestockt, den Rückstand bei der Antragsbearbeitung zu beheben. Die Verfahren zielen auf eine kurzfristige Verbesserung der Situation ab.

3. Die internationale Perspektive. Die UN/EU-Taskforce zur Geheimhaltung von Daten und Mikrodaten

Auf der Konferenz der europäischen Statistiker im Juni 2003 wurde die Taskforce zu Datengeheimhaltung und Mikrodaten unter dem Vorsitz von Dennis Trewin vom Australian Bureau of Statistics eingerichtet. Die angestrebten Ergebnisse sollten die Erarbeitung vereinbarter Grundsätze im Hinblick auf die Bereitstellung eines Zugangs zu den Mikrodaten sowie die Präsentation von Fallstudien über bewährte Verfahren im Einklang mit diesen Grundsätzen umfassen.

Im Mai 2004 legte die Taskforce ein Diskussionspapier [4] vor, in dem die Anliegen der nationalen Statistikämter und der Forschungsgemeinde behandelt wurden. Ferner wurden die Wege zur Auflösung der zwischen diesen beiden Perspektiven bestehenden Spannungen sowie die zu diesem Zweck verwendbaren Mittel, beispielsweise anonyme Mikrodatenfiles, Fernzugriffsmöglichkeiten sowie Datenlabors vorgestellt. Es wurden eine Reihe von Grundsätzen vorgeschlagen und einige Fragen zur Diskussion vorgelegt.

Im Oktober 2004 wurden die Stellungnahmen der Mitgliedstaaten in einem Dokument [5] zusammengefasst. Das Dokument enthält ebenfalls eine Liste der Fragen, über die weitgehend Einigkeit herrscht sowie der Themen, über die es verschiedene Auffassungen gibt.

Zum Abschluss ihres Mandats plant die Taskforce die Vorlage eines Dokuments mit den Leitlinien für bewährte Verfahren auf der Plenarsitzung des CES im Juni 2005.

4. Ausblick in die Zukunft

Es sind mehrere parallele Methoden denkbar, mit deren Hilfe statistische Daten bei einer ausreichenden Garantie der Sicherung gegen Offenlegung in stärkerem Maße nutzbar für die Forschung gemacht werden können. Im Folgenden möchte ich einen Überblick über diese Methoden geben:

(1) Harmonisierte Kriterien für das Risiko der Offenlegung. Im Allgemeinen sind die auf nationaler und europäischer Ebene geltenden Rechtsvorschriften in Bezug auf die als vertraulich geltenden Daten weitgehend harmonisiert. Die bei der Umsetzung der Rechtsvorschriften verwendeten Kriterien weisen jedoch von Land zu Land erhebliche Unterschiede auf. Die Kriterien sind in manchen Fällen historisch bedingt, in anderen wiederum verfügen sie nicht über eine solide wissenschaftliche Grundlage. In vielen Fällen ziehen sie konservative Lösungen nach sich, da die tatsächlichen Risiken nicht ausreichend berücksichtigt werden.

Diese Vielfalt der Interpretationen ist eine Konsequenz aus der Tatsache, dass es keine harmonisierte Vorgehensweise in Bezug auf das Risiko der Datenoffenlegung gibt. Für die Festlegung von Maßnahmen zur Sicherung der Daten gegen Offenlegung ist es zunächst notwendig, eine Einigung über die Sensibilität der Daten (den „Vertraulichkeitsgrad“ der Variablen in der Datei) sowie über die Möglichkeit des Abgleichs dieser Daten mit externen Quellen im Hinblick auf das Vorhandensein von Schlüssel- oder Identifizierungsvariablen festzulegen. Zweitens ist es notwendig, eine harmonisierte Methode zur Erfassung des Risikos zu erarbeiten. Für den Ausgleich zwischen den verschiedenen Arbeitsmethoden bzw. für die Auswahl einer bestimmten Vorgehensweise ist methodische Arbeit nötig.

Ganz offensichtlich ist der Bedarf an gemeinsamen Kernkriterien, die bei einer adäquaten Stufe der Harmonisierung eine gewisse Flexibilität ermöglichen, damit eine Anpassung an die spezifische Wahrnehmung der Gesellschaft in den einzelnen Ländern gewährleistet ist. Ein weiterer Vorteil besteht darin, eine international vereinbarte Grundlage zur Verfügung zu haben, auf deren Basis die nationalen Entscheidungen im Hinblick auf die Freigabe von Mikrodaten besser zu rechtfertigen sind.

(2) Zulassungsverfahren für Forscher und Forschungsprojekte. Gegenwärtig sind die Kriterien für die Zulassung von Forschungsprojekten und Forschern auf der Ebene der Europäischen Union nicht ganz transparent und natürlich nicht homogen.

Bei der Beurteilung der Forscher neigt man allzu oft dazu, in Wirklichkeit die entsprechende Forschungseinrichtung zu bewerten und die Verbindung zwischen der Einrichtung und der Einzelperson zu stark zu gewichten. Eine A-Priori-Zulassungsbeurteilung ist ein Schritt, und er erfordert große Sorgfalt. Ein vom administrativen und qualitativen Standpunkt viel mehr versprechendes Verfahren scheint jedoch eine Ex-Post-Beurteilung auf der Grundlage einer Standard-Beurteilung der Einrichtung sowie entsprechender Daten zu sein. Die Möglichkeit, eine schwarze Liste mit den Daten der Einrichtungen aufzustellen, bei denen Datenmissbrauch vorgekommen ist, würde den Druck auf die Forschungsinstitutionen erhöhen.

Bei der Beurteilung von Forschungsprojekten ist die Einbeziehung der technischen Abteilungen, die bereits mit den Forschern und deren Arbeit etwas vertraut sind, bereits entsprechende Studien zu dem Thema durchgeführt oder ein direktes Interesse an der Untersuchung haben könnten, von entscheidender Bedeutung. Bedenken Sie, dass in einigen Ländern, wie beispielsweise im Vereinigten Königreich, das Interesse des Statistikamtes an einem Forschungsprojekt das entscheidende Kriterium darstellt. Auch andere Aspekte, beispielsweise die Originalität der

Forschungsarbeit, die tatsächliche Notwendigkeit der vertraulichen Daten, Kollision der Interessen zwischen den nationalen Statistikämtern und dem Forscher können berücksichtigt werden.

Im Rahmen der Verordnung Nr. 831/2002 der Kommission wurden die Zulassungskriterien entwickelt und entsprechende Bewertungsfragebögen entworfen. Auf dieser Grundlage könnte eine Reihe von transparenten Kriterien entwickelt werden, die eine gleichberechtigte Behandlung der wissenschaftlichen Gemeinschaft in der EU gewährleisten würden.

(3) Rechtliche Bestimmungen für den Fall eines Datenmissbrauchs. Einen wichtigen Aspekt des Datenschutzes bildet das Wissen des Benutzers in Bezug auf die rechtliche Verantwortlichkeit für seine Aktionen und die rechtlichen Schritte im Falle eines Missbrauchs. Für den Fall einer internationalen Verwendung sind die rechtlichen Verantwortlichkeiten festzulegen und explizit mitzuteilen.

(4) Die Rolle der CENEX bei der Sicherung statistischer Daten gegen Offenlegung. Die vom ASP eingesetzte Taskforce zum Thema Kompetenzzentren legte den Vorschlag vor, im Laufe des Jahres 2005 ein Pilotprojekt zum Konzept der Kompetenzzentren (CENEX) vorzulegen. Grob skizziert besteht das Konzept darin, ein Team aus den Mitgliedern der nationalen statistischen Einrichtungen zu bilden, das dann ihr Fachwissen für spezifische Bereiche zur Verfügung stellen sowie Hilfsprogramme und Wissen entwickeln sollte, von denen nicht nur die Teammitglieder, sondern auch der Rest der ESS-Gemeinschaft profitieren würde. Die statistische Offenlegungskontrolle ist als eines der beiden Themen in Betracht gezogen worden, die zur Integration dieser CENEX-Pilotphase beitragen sollen. Eurostat ist gegenwärtig mit der Vorbereitung der Dokumentation für den Projektstart beschäftigt. Die zum Abschluss des planmäßig für ein Jahr angesetzten Pilotprojekts erwarteten Ergebnisse sollen eine Bestandsaufnahme des Bedarfs, Entwicklung von entsprechenden Computer-Hilfsprogrammen sowie ein Handbuch über gängige Verfahren im Bereich der Offenlegungskontrolle umfassen.

(5) Ein kürzlich vom Ausschuss für das statistische Programm angenommener Verhaltenskodex beinhaltet einige Bestimmungen über die Verwendung statistischer Daten durch die Forscher sowie über den Schutz vertraulicher Daten. Dieser Verhaltenskodex stellt einen Rahmen für die Entwicklung harmonisierter Aktivitäten in diesem Bereich bereit.

(6) Datenfernzugriff für die Forscher. Die Mitgliedstaaten sind sich weitgehend darüber einig, dass der Fernzugriff auf Daten eine viel versprechende Methode darstellt. Bislang gibt es aber nur sehr vereinzelte Anwendungsbeispiele. Es sind zwei Typen eines Datenfernzugriffs zu erwähnen. Bei der ersten Methode handelt es sich um eine Fernausführung: Der Forscher legt ein Programm

vor und die Ergebnisse werden anschließend per E-Mail übermittelt. Bei der zweiten Methode handelt es sich um einen Fernzugriff im eigentlichen Sinne: Die Forscher führen eine Analyse aus und können die Antworten direkt auf dem Bildschirm verfolgen. Auf der Sitzung der IT-Direktorengruppe vom Oktober 2004 stellte Eurostat eine Analyse von drei gegenwärtig ausgeführten Fernzugriffprojekten vor. Zwei dieser Projekte, nämlich die Projekte aus dem US National Center of Health Research sowie das Projekt Luxemburger Einkommenserhebung betreffen die Fernausführung, das dritte Beispiel von Denmark Statistics bezieht sich dagegen auf Datenfernzugriff.

(7) Öffentlich genutzte (Public Use) Dateien. Eine sehr häufig gestellte Frage bezieht sich auf das Problem, ob der Schutz einer Datei gegen Offenlegung so hoch angesetzt werden kann, dass ein öffentlicher Zugriff auf die anonymisierten Daten möglich ist. Beispiele für eine öffentliche Verwendung von Dateien in der Europäischen Union lassen sich in Großbritannien, Frankreich, Italien und Österreich finden. Öffentlich genutzte Dateien können natürlich nicht die vertraulichen Dateien ersetzen, da diese viel mehr Möglichkeiten für eine Analyse bieten. Auf der anderen Seite werden die öffentlich genutzten Dateien mit einem minimalen Kostenaufwand produziert und ihre Bereitstellung für die Nutzer kann sofort erfolgen.

In Dezember 2004 legte Eurostat dem Ausschuss für die statistische Geheimhaltung eine Strategie für die Erstellung von öffentlich genutzten Dateien sowie einen spezifischen Vorschlag in Bezug auf die Arbeitskräfteerhebung (AKE) vor. Dieser Vorschlag fand jedoch im Ausschuss nicht genügend Unterstützung.

5. Schlussfolgerungen

(1) Es besteht ein beträchtlicher Unterschied zwischen den in den Statistikdaten enthaltenen Informationen und den tatsächlich von den Statistikämtern veröffentlichten Daten. Ein Weg, diesen Unterschied auszugleichen, besteht darin, der Forschungsgemeinde Mikrodattendateien zur Verfügung zu stellen.

(2) Die Bereitstellung statistischer Daten birgt viele Risiken, die einer Lösung bedürfen, beispielsweise in Bezug auf den rechtlichen Schutz gegen Identifizierung von Einzelpersonen, die Möglichkeit eines Datenmissbrauchs sowie die Wahrnehmung der Einzelnen in Bezug auf einen manipulativen Missbrauch ihrer Daten. In Bezug auf diese Risiken besteht Handlungsbedarf.

(3) Unser Ziel muss darin bestehen, den bestehenden Unterschied auszugleichen und gleichzeitig den Risiken Rechnung zu tragen. Zu diesem Zweck können mehrere rechtliche und technische Maßnahmen in Erwägung gezogen werden. Die rechtlichen Maßnahmen beziehen sich

auf Zulassungsverfahren für Forscher und Forschungsprojekte. Die technischen Maßnahmen dagegen beziehen sich auf verschiedene Methoden im Bereich Schutz vertraulicher Daten.

(4) Die Reaktionen auf internationaler Ebene zeigen, dass trotz der bestehenden weitgehenden Übereinstimmung in Bezug auf die Bereitstellung von Mikrodaten für die Forschungsgemeinde in vielen Detailfragen weiterhin unterschiedliche Meinungen bestehen. Insbesondere bei den Kriterien für eine für Verbreitungszwecke ausreichende Schutzstufe sind die Unterschiede sehr groß.

(5) Die EU-Verordnung 831/2002 stellt einen nützlichen rechtlichen Rahmen für die Bereitstellung von Mikrodaten für die Forschungsgemeinde zur Verfügung. Nach der zunächst schwierigen Umsetzungsphase werden die durch diesen Rechtsakt festgelegten Prozesse und Fristen im Hinblick auf eine mögliche Verbesserung überprüft.

(6) Bei der Frage der optimierten Bereitstellung von Mikrodaten an die Forscher können mehrere Methoden in Betracht gezogen werden. Zunächst die Entwicklung harmonisierter Kriterien für die Anonymisierung sowie für die Zulassung von Forschern/Forschungsprojekten. Zweitens der rechtliche Rahmen zur Vermeidung eines Missbrauchs. Drittens die Anwendung eines Verhaltenskodex. Viertens die Möglichkeit eines Fernzugriffs auf die Mikrodaten. Und schließlich die Erstellung von öffentlich genutzten Dateien.

Literaturangaben:

[1] (2003), 19. CEIES-Seminar: Innovative Lösungen für einen Zugang zu Mikrodaten, Lissabon, 26. und 27. September 2002, Luxemburg: Amt für amtliche Veröffentlichungen der Europäischen Gemeinschaften

[2] (2004), Monographs of official statistics: Work session on statistical data confidentiality, Luxembourg, 7 to 9 April 2003, Luxemburg: Amt für amtliche Veröffentlichungen der Europäischen Gemeinschaften

[3] Domingo-Ferrer, J. & Torra, V. (2004), Privacy in Statistical Databases – CASC Project Final Conference, PSD 2004, Barcelona, Katalonien, Spanien, Juni 2004, Proceedings, Springer-Verlag

[4] Trewin, D. (2004), CES Task Force on Confidentiality and Microdata – Discussion Paper, 52nd Plenary Session, Paris, 8-10 June 2004, Kommission des Wirtschafts- und Sozialrates der Vereinten Nationen, Statistikkommission und Wirtschaftskommission für Europa, Konferenz Europäischer Statistiker

[5] Trewin, D. (2004), CES Task Force on Confidentiality and Microdata, First meeting of the 2004/2005 Bureau, Washington, DC, 18-19 October 2004, Statistikkommission und Wirtschaftskommission für Europa, Konferenz Europäischer Statistiker, Dokument Nr. 10