

# Statistische Analysen und Studien Nordrhein-Westfalen

## Band 6

### Impressum

Herausgeber :  
Landesamt für Datenverarbeitung  
und Statistik Nordrhein-Westfalen

Redaktion:  
Bianca Klose, Hans Lohmann

Preis dieser Ausgabe: 1,30 EUR

Erscheinungsfolge: unregelmäßig

Bestellungen nehmen entgegen:

das Landesamt für Datenverarbeitung  
und Statistik NRW,  
Postfach 10 11 05,  
40002 Düsseldorf,  
Mauerstraße 51,  
40476 Düsseldorf  
Telefon: 0211 9449-2516/3516  
Telefax: 0211 442006  
Internet: <http://www.lds.nrw.de>  
E-Mail: [poststelle@lds.nrw.de](mailto:poststelle@lds.nrw.de)

sowie der Buchhandel.

Pressestelle:  
0211 9449-2521/2518

Zentraler Informationsdienst:  
0211 9449-2495/2525

© Landesamt für Datenverarbeitung  
und Statistik NRW, Düsseldorf, 2002

Für nicht gewerbliche Zwecke sind  
Vervielfältigung und unentgeltliche  
Verbreitung, auch auszugsweise, mit  
Quellenangabe gestattet. Die Verbrei-  
tung, auch auszugsweise, über elek-  
tronische Systeme/Datenträger bedarf  
der vorherigen Zustimmung. Alle üb-  
rigen Rechte bleiben vorbehalten.

Bestell-Nr. Z 08 1 2002 56

ISSN 1619-506X

### Inhalt

#### **Datenbedarf der Wissenschaft – Ein Bericht des Forschungsdatenzentrums der statistischen Landesämter über die erste Nutzerbefragung**

**3**

Dr. Sylvia Zühlke und Uwe Hetke – Geschäftsstelle des  
Forschungsdatenzentrums der statistischen Landesämter

## **Zeichenerklärung**

(nach DIN 55 301)

- 0 weniger als die Hälfte von 1 in der letzten besetzten Stelle, jedoch mehr als nichts
- nichts vorhanden (genau null)
- . Zahlenwert unbekannt oder geheim zu halten
- ... Zahlenwert lag bei Redaktionsschluss noch nicht vor
- ( ) Aussagewert eingeschränkt, da der Wert Fehler aufweisen kann
- / keine Angabe, da der Zahlenwert nicht sicher genug ist
- x Tabellenfach gesperrt, weil Aussage nicht sinnvoll
- p vorläufige Zahl
- r berichtigte Zahl

Abweichungen in den Summen erklären sich aus dem Runden der Einzelwerte.

# Datenbedarf der Wissenschaft

## Ein Bericht des Forschungsdatenzentrums der statistischen Landesämter über die erste Nutzerbefragung

Dr. Sylvia Zühlke und Uwe Hetke  
Geschäftsstelle des Forschungsdatenzentrums der statistischen Landesämter

### 1 Einleitung

Um den Zugang der Wissenschaft zu den Mikrodaten der amtlichen Statistik zu erleichtern, empfiehlt die Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI) die Einrichtung von Forschungsdatenzentren (siehe Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik 2001). Entsprechend dieser Empfehlung richten die statistischen Landesämter und das Statistische Bundesamt zurzeit jeweils ein Forschungsdatenzentrum ein. Um bei der Entwicklung des Dienstleistungsangebotes den konkreten Datenbedarf der Wissenschaft berücksichtigen zu können, hat das Forschungsdatenzentrum der statistischen Landesämter in Abstimmung mit dem Forschungsdatenzentrum des Statistischen Bundesamtes im Sommer 2002 die „Umfrage zum Datenbedarf der Wissenschaft“ durchgeführt. Sie bezieht sich auf den gesamten Bestand an Mikrodaten der Statistischen Ämter des Bundes und der Länder, der zum allergrößten Teil in den statistischen Landesämtern liegt.

Ziel der Umfrage war es, die potenziellen Nutzerinnen und Nutzer der Mikrodaten der amtlichen Statistik anzusprechen und ihren konkreten Datenbedarf zu ermitteln. Weiterhin konnten sich die Befragten zu unterschiedlichen Nutzungsmöglichkeiten, den von ihnen eingesetzten Analyseprogrammen und ihrem Interesse an geplanten Veranstaltungen der beiden Forschungsdatenzentren äußern. Die Angaben werden im Folgenden ausgewertet und sollen bei der zukünftigen Entwicklung des Dienstleistungsangebotes der Forschungs-

datenzentren berücksichtigt werden. Die Wissenschaftlerinnen und Wissenschaftler wurden für die Umfrage auf mehreren Wegen angesprochen: Zum einen wurde die Adressdatei „Vademecum“ genutzt, die freiwillige Angaben der Wissenschaftlerinnen und Wissenschaftler zu ihren Forschungsschwerpunkten und ihren Adressen enthält. Da diese Adressdatei nicht in erster Linie Personen, sondern Institutionen enthält, wurden zum anderen die größeren wissenschaftlichen Fachgesellschaften<sup>1)</sup> mit der Bitte angeschrieben, ihre Mitglieder über die Umfrage zu informieren. Weiterhin wurde die Umfrage von Wissenschaftlerinnen und Wissenschaftlern im Schneeballverfahren an interessierte Kollegen weitergeleitet<sup>2)</sup>. Auf diese Art und Weise konnte eine große Anzahl an Wissenschaftlerinnen und Wissenschaftlern erreicht werden<sup>3)</sup>.

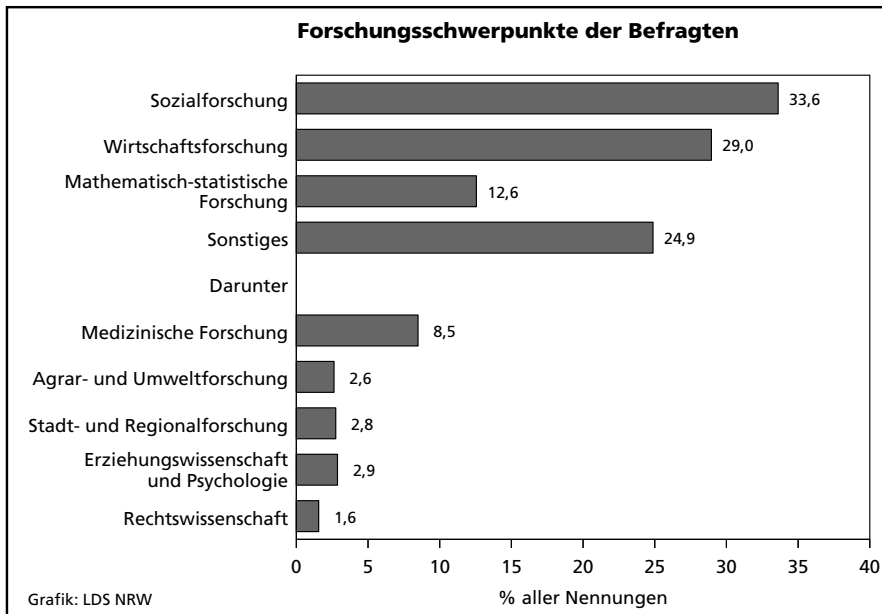
Insgesamt wurden 695 gültige Angaben von Wissenschaftlerinnen und Wissenschaftlern erfasst. Von diesen haben 584 angegeben, dass sie Mikrodaten für ihre wissenschaftliche Tätigkeit nutzen oder benötigen. Dieser Personenkreis bil-

det somit den potenziellen Nutzerkreis der Forschungsdatenzentren der statistischen Landesämter und des Statistischen Bundesamtes. Die übrigen 111 Wissenschaftlerinnen und Wissenschaftler gaben an, dass sie zurzeit keine Mikrodaten für ihre wissenschaftliche Tätigkeit benötigen. Doch auch bei diesem Personenkreis ist das Interesse an der Umfrage sowie an den Aktivitäten der Forschungsdatenzentren sehr hoch. So möchte etwa die Hälfte dieses Personenkreises über die Umfrage und die weiteren Aktivitäten der Forschungsdatenzentren informiert werden.

### 2 Forschungsgebiete und Datenbedarf

In der amtlichen Statistik werden verschiedene Erhebungen in unterschiedlichen periodischen Abständen durchgeführt. Diese einzelnen Statistiken können aus Kapazitätsgründen nicht alle sofort für Analysezwecke bereitgestellt werden. Daher wurden die Wissenschaftlerinnen und Wissenschaftler, die einen Bedarf an Mikrodaten geäußert haben, nach ihren Forschungsschwerpunkten und ihrem konkreten Datenbedarf gefragt. Die Forschungsschwerpunkte der Wissenschaftlerinnen und Wissenschaftler, die sich an der Umfrage beteiligt haben, liegen im Bereich der Sozialforschung (33,6 %), gefolgt von den Bereichen „Wirtschaftsforschung“ (28,9 %) und „Mathematisch-statistische Forschung“ (12,6 %). Die restlichen 24,8 Prozent der Nennungen entfallen auf die sonstigen Forschungsschwerpunkte, wobei hier die medizinische Forschung einen Anteil von 8,5 % aufweist. Bei den Sonstigen konnten weiterhin folgende Forschungs-

1) Angesprochen wurden die folgenden Fachgesellschaften: Deutsche Arbeitsgemeinschaft für Epidemiologie, Deutsche Mathematikervereinigung, Deutsche Mathematikervereinigung Fachgruppe Stochastik, Gesellschaft für Angewandte Mathematik und Mechanik (GAMM), Gesellschaft für Klassifikation e. V. (GfKl), Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS) e. V., Internationale Biometrische Gesellschaft Deutsche Region (IBS-DR), Stifterverband für die Deutsche Wissenschaft, Verband Deutscher Städtestatistiker (VDSt), Verein für Sozialpolitik, Deutsche Gesellschaft für Demographie (DGD) e. V., Deutscher Verband für Angewandte Geographie (DVAG) e. V., Akademie für Raumforschung und Landesplanung – ARL, Deutsche Gesellschaft für Soziologie (DGS), Deutsche Statistische Gesellschaft (DStatG). – 2) Da dieses Verfahren Doppelerfassungen nicht ausschließt, wurden die Daten des endgültigen Bestandes über die Adressangaben auf Doppelerfassungen überprüft. – 3) Aufgrund des genannten Auswahlverfahrens kann jedoch keine Rücklaufquote ermittelt werden.



schwerpunkte abgegrenzt werden: Erziehungswissenschaft und Psychologie (2,9 %), Stadt- und Regionalforschung (2,8 %), Agrar- und Umweltforschung (2,6 %) und Rechtswissenschaften (1,6 %).

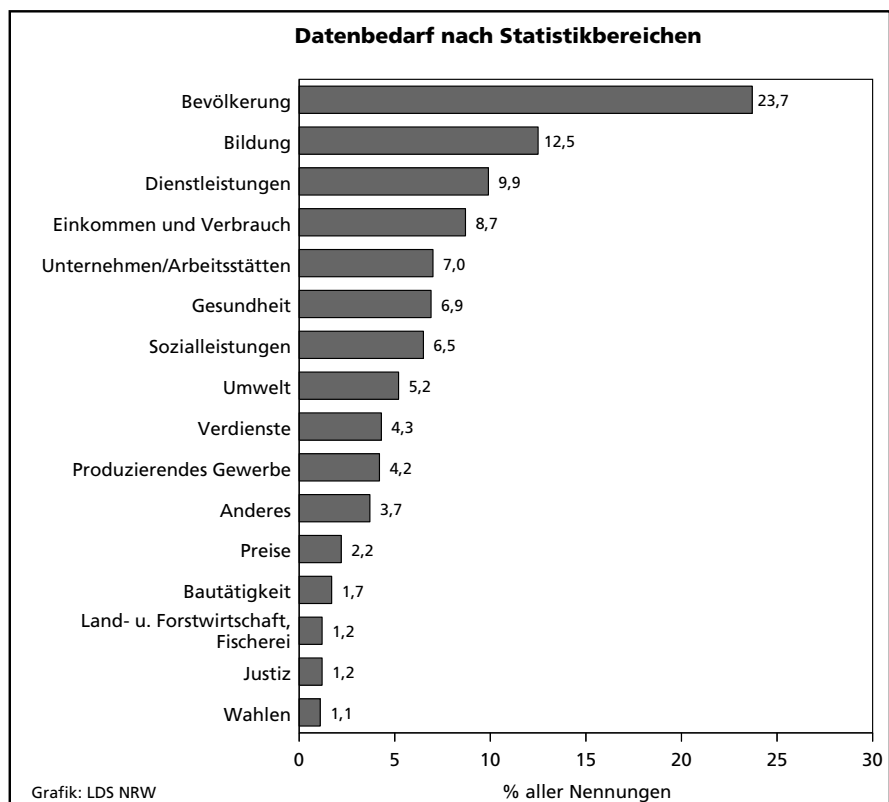
Um ihren konkreten Datenbedarf anzugeben, hatten die Wissenschaftlerinnen und Wissenschaftler die Möglichkeit, aus 60 Statistiken bis zu fünf Statistiken auszuwählen. Diese Anzahl wurde von einigen Befragten als zu gering erachtet, so dass in einigen Fällen mehr als fünf Statistiken angegeben wurden<sup>4)</sup>. Insgesamt wurden 4 106 Angaben gemacht, so dass im Durchschnitt etwa sieben Statistiken von den Befragten ausgewählt wurden.

Betrachtet man zunächst nicht die einzelnen Statistiken, sondern größere Statistikbereiche, so zeigt sich, dass der Schwerpunkt des Datenbedarfs insbesondere im Bereich „Bevölkerung“ (23,7 %) liegt. An zweiter und dritter Stelle befinden sich die Bereiche „Bildung“ und „Dienstleistungen“, die jedoch mit 12,5 % bzw. 9,9 % deutlich weniger Nennungen aufweisen. Der Schwerpunkt des Datenbedarfs liegt somit eher im Bereich der Sozialstatistiken als im Bereich der Wirtschaftsstatistiken. Dies wird auch deutlich, wenn man die weitere Rangfolge betrachtet: So entfallen auf den Bereich „Einkom-

4) Bei der Auswahl von mehr als fünf Statistiken wurden alle Angaben in die Auswertung einbezogen.

men und Verbrauch“ 8,7 % aller Nennungen, gefolgt von den Bereichen „Unternehmen und Arbeitsstätten“ (7,0 %), „Gesundheit“ (6,9 %), „Sozialleistungen“ (6,5 %) und „Umwelt“ (5,2 %).

Wechselt man auf die Ebene der Einzelstatistiken, so ist zu erkennen, dass der Datenbedarf der Wissenschaft sehr stark streut. Zwar ist ein Schwerpunkt bei den Haushaltsstichproben, insbesondere beim Mikrozensus zu erkennen. Sieht man je-



doch von dieser Schwerpunktsetzung ab, werden sehr viele unterschiedliche Statistiken genannt. Betrachtet man die zehn am häufigsten genannten Statistiken (siehe nebenstehende Tabelle), so ergibt sich die Rangfolge: Mikrozensus (7,5 %), Volkszählung (4,6 %), Europäisches Haushaltspanel (4,4 %), Löhne und Gehälter, Arbeitskosten (4,3 %), Einkommens- und Verbrauchsstichproben (4,2 %), Hochschulen (3,5 %), Natürliche Bevölkerungsbewegungen (3,0 %), Dienstleistungsstatistiken (2,9 %), Wanderungen (2,8 %) und Berufsbildung (2,6 %).

Bei der Weiterentwicklung des Dienstleistungsangebotes der Forschungsdatenzentren der statistischen Landesämter und des Statistischen Bundesamtes wird der dargestellte Datenbedarf berücksichtigt. In diesem Zusammenhang ist zu erwähnen, dass für einzelne Bereiche, wie beispielsweise den Haushaltsstichproben, bereits faktisch anonymisierte Datensätze angeboten werden (siehe hierzu näheres in Abschnitt 7). Da dieses Angebot möglicherweise nicht allen interessierten Nutzerinnen und Nutzern bekannt ist, werden das Forschungsdatenzentrum der statistischen Landesämter und

Statistiken nach Anzahl der Nennungen (Rang 1 bis 40)			
Rang	Statistik	Antworten	
		Anzahl	%
	<b>Antworten insgesamt</b>	<b>4 106</b>	<b>100</b>
1	Mikrozensus	310	7,5
2	Volkszählungen	187	4,6
3	Europäisches Haushaltspanel	181	4,4
4	Löhne und Gehälter, Arbeitskosten	175	4,3
5	Einkommens- und Verbrauchsstichproben	173	4,2
6	Hochschulen (Studenten, Personal, Habilitationen)	143	3,5
7	Natürliche Bevölkerungsbewegungen	124	3,0
8	Dienstleistungsstatistik	121	2,9
9	Wanderungen	113	2,8
10	Berufsbildung	108	2,6
11	Allgemein bildende und berufliche Schulen	107	2,6
12	Zeitbudgeterhebung	105	2,6
13	Todesursachen	104	2,5
14	Krankenhäuser (Grunddaten, Kosten, Diagnosen)	97	2,4
15	Verarbeitendes Gewerbe, Bergbau und Gewinnung von Steinen und Erden	95	2,3
16	Sozialhilfe	91	2,2
17	Preise	89	2,2
18	Unternehmensregister	87	2,1
19	Laufende Wirtschaftsrechnungen	79	1,9
20	Ausgaben und Einnahmen für Wissenschaft und Forschung	75	1,8
21	Steuern (Einkommen-, Körperschaft-, Umsatz-, Gewerbesteuer)	74	1,8
22	Kostenstrukturen	67	1,6
23	Verkehr und Nachrichtenübermittlung, Straßenverkehrsunfälle	62	1,5
24	Insolvenzen, Jahresabschlüsse	61	1,5
25	Einbürgerungen	58	1,4
26	Personal im öffentlichen Dienst	58	1,4
27	Arbeitsstätten	57	1,4
28	Kinder- und Jugendhilfe	56	1,4
29	Rehabilitationsmaßnahmen	56	1,4
30	Land- und Forstwirtschaft, Fischerei	51	1,2
31	Luftverunreinigungen	51	1,2
32	Umweltschutzausgaben und -produkte	50	1,2
33	Europäische Erhebung über betriebliche Weiterbildung	49	1,2
34	Pflege (Heime, Dienste, Pflegegeldempfänger)	47	1,1
35	Repräsentative Wahlstatistik	46	1,1
36	Energie- und Wasserversorgung	46	1,1
37	Einzelhandel	46	1,1
38	Außenhandel	46	1,1
39	Gebäude- und Wohnungsbestand	44	1,1
40	Wasserversorgung und Abwasserbeseitigung	38	0,9

### 3 Datenschutz und Möglichkeiten der Datennutzung

Die amtliche Statistik ist nach § 16 des Bundesstatistikgesetzes verpflichtet, Einzelangaben über persönliche und sachliche Verhältnisse, die von den Befragten für eine Bundesstatistik gemacht werden, geheim zu halten. Die gleiche Regelung ist in den vorhandenen Landesstatistikgesetzen enthalten. Sie dient dem Ziel,

- den Einzelnen vor der Offenlegung seiner persönlichen und sachlichen Verhältnisse zu schützen,
- das Vertrauensverhältnis zwischen den Befragten und den statistischen Ämtern zu erhalten

und

- die Zuverlässigkeit der Angaben und die Berichtswilligkeit der Befragten zu gewährleisten.

Das Bundesverfassungsgericht hat dem Statistikgeheimnis im Volkszählungsurteil eine herausragende Bedeutung beigemessen. Es betrachtet das Statistikgeheimnis nicht nur als konstitutiv für die Funktionsfähigkeit der Bundesstatistik, sondern auch im Hinblick auf den Schutz des Rechts auf informationelle Selbstbestimmung.

Besondere Bedingungen gelten für die Weitergabe von faktisch anonymisierten Einzelangaben, die durch § 16 Abs. 6 des Bundesstatistikgesetzes ermöglicht wird. Danach dürfen Einzelangaben an Hochschulen oder sonstige Einrichtungen mit der Aufgabe unabhängiger wissenschaftlicher Forschung übermittelt werden, wenn die Einzelangaben nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft zugeordnet werden können und die Empfänger zur statistischen Geheimhaltung verpflichtet sind bzw. werden.

Die Konkretisierung der Wendung „unverhältnismäßig großer Aufwand an Zeit, Kosten und Arbeitskraft“ wurde im Rahmen eines gemeinsamen Forschungsprojektes der Wissenschaft und der amtlichen Statistik vorgenommen (siehe Müller/Blien/Kno-

das Forschungsdatenzentrum des Statistischen Bundesamtes im Rahmen von Nutzerkonferenzen und E-Mail-Aktionen zukünftig systematisch über dieses Datenangebot informieren.

Bei der Weiterentwicklung des Datenangebotes sollen zudem die Empfehlungen berücksichtigt werden, die der Gründungsausschuss des Rates für Sozial- und Wirtschaftsdaten für den Aufbau des Forschungsdatenzentrums der statistischen Landesämter ausgesprochen hat. Diese Empfehlungen decken sich in weiten Teilen mit den 15 am häufigsten genannten Statistiken. Nach den Empfehlungen des Gründungsausschusses sollen für folgende Statistiken im

Forschungsdatenzentrum der statistischen Landesämter prioritär Daten bereitgestellt werden:

- Statistiken des Verarbeitenden Gewerbes (Produktionsstatistik, jährliche Unternehmenserhebung) sowie des Baugewerbes,
- Monatliche Statistiken des Einzelhandels und des Gastgewerbes, jährliche Dienstleistungsstatistik, sowie monatliche Tourismusstatistik,
- Mikrozensus als Regional- und Paneldatensatz,
- Kurzfristige Konjunkturstatistiken im Produzierenden Gewerbe Deutschlands,
- Diagnose- und Todesursachenstatistik.

che/Wirth u. a. 1991). Im Rahmen dieses Projektes wurden anhand des Mikrozensus Reidentifikationsrisiken sowie Methoden der faktischen Anonymisierung untersucht und hieraus Empfehlungen für die Erstellung faktisch anonymisierter Datensätze abgeleitet. Diese Empfehlungen wurden von der amtlichen Statistik auf mehrere haushalts- und personenbezogene Datensätze angewendet (Mikrozensus, Europäisches Haushaltspanel, Einkommens- und Verbrauchsstichprobe und Sozialhilfestatistik).

Faktisch Anonymisierte Datensätze haben für die Nutzer den Vorteil, dass sie flexibel am eigenen Arbeitsplatz ohne Geheimhaltungsprüfung durch ein statistisches Amt ausgewertet werden können. Daher ist die amtliche Statistik bestrebt, das Angebot an faktisch anonymisierten Datensätzen auszuweiten. Einer flächendeckenden Bereitstellung solcher Datensätze stehen jedoch ökonomische sowie fachliche Gründe entgegen:

- Der Datenbedarf der Wissenschaft ist sehr vielfältig, so dass der potenzielle Nutzerkreis bei einzelnen Erhebungen relativ klein ist. Da die Erstellung von faktisch anonymisierten Datensätzen in der Regel hohe Kosten verursacht, ist für die einzelnen Erhebungen jeweils abzuwägen, welche Form der Datenbereitstellung unter Kosten-Nutzen-Gesichtspunkten die günstigste ist.
- Bei der Erstellung von faktisch anonymisierten Datensätzen wird in der Regel die Regionalkennung weggelassen, da diese ein hohes Reidentifikationsrisiko birgt<sup>5)</sup>. Auch müssen Merkmale mit sehr vielen Ausprägungen in der Regel zusammengefasst werden. Daher können Analysen, die auf eine starke regionale oder fachliche Differenzierung abstellen, häufig nicht auf der Basis von faktisch anonymisierten Datensätzen durchgeführt werden.
- Für bestimmte Arten von Erhebungen ist die Erstellung von faktisch anonymisierten Datensätzen aufgrund fehlender Anonymisierungsverfahren zum jetzigen Zeitpunkt

nicht möglich. Dies betrifft insbesondere Betriebs- und Unternehmensdaten, aber auch Paneldaten. In einigen Ländern, wie zum Beispiel Kanada, werden daher grundsätzlich keine Datensätze der Wirtschaftsstatistiken erstellt, die nach außen gegeben werden (siehe Tambay/Goldmann/White 2001). In Deutschland wird die Anonymisierung von wirtschaftsstatistischen Mikrodaten zurzeit im Rahmen eines gemeinsamen Projektes der amtlichen Statistik und der Wissenschaft untersucht (siehe Sturm 2002).

- Schließlich ist zu berücksichtigen, dass durch die Weitergabe von faktisch anonymisierten Datensätzen die Reidentifikationsrisiken steigen, da bei den einzelnen wissenschaftlichen Einrichtungen in der Regel verschiedene Datensätze parallel gehalten werden. Dieses Problem hat vor 15 Jahren, als das oben genannte Projekt zur Bestimmung von Reidentifikationsrisiken durchgeführt wurde, in der Weise noch nicht existiert und sollte daher untersucht werden.

Aus den genannten Gründen werden die beiden Forschungsdatenzentren der statistischen Landesämter und des Statistischen Bundesamtes neben der Erstellung von faktisch anonymisierten Datensätzen folgende Möglichkeiten für die Datennutzung anbieten:

- In den statistischen Ämtern werden voraussichtlich bis Mitte 2004 Arbeitsplätze für Gastwissenschaftler eingerichtet, die zwei Formen des Datenzugangs ermöglichen: Zum einen können im Rahmen von Forschungsprojekten, die in Kooperation zwischen amtlicher Statistik und externen Fachleuten durchgeführt werden, nicht oder nur formal anonymisierte Mikrodaten in den jeweils zuständigen Forschungsdatenzentren genutzt werden. Zum anderen können externen Wissenschaftlerinnen und Wissenschaftlern für eigene Forschungsprojekte in den jeweiligen Forschungsdatenzentren spezielle faktisch anonymisierte Datensätze für die On-Site-Nutzung bereitgestellt werden, die zwar anonymi-

siert sind, jedoch wesentlich mehr Informationen enthalten als die zuvor beschriebenen faktisch anonymisierten Datensätze. Bei beiden Varianten ist zu prüfen, ob die statistische Geheimhaltung eingehalten wird.

- Die statistischen Ämter werden voraussichtlich ab Mitte 2005 die Möglichkeit der kontrollierten Datenfernverarbeitung anbieten, bei der Wissenschaftlerinnen und Wissenschaftler ihre Analyseprogramme an ihrem Arbeitsplatz schreiben und diese dann an ein statistisches Amt schicken, in dem dann die Programme nach Syntaxprüfung ausgeführt werden. Nachdem die Analyseergebnisse daraufhin untersucht wurden, ob die statistische Geheimhaltung eingehalten wird, werden sie an den jeweiligen Nutzer übermittelt.

#### 4 Präferenzen für unterschiedliche Nutzungswege

Die Wissenschaftlerinnen und Wissenschaftler wurden im Rahmen der Befragung aufgefordert, die genannten Nutzungsmöglichkeiten in eine Rangfolge von 1 (höchste Präferenz) bis 4 (niedrigste Präferenz) zu bringen. Um die Präferenzen für die einzelnen Nutzungswege zu analysieren, wird für die einzelnen Nutzungswege die Verteilung der Präferenzen betrachtet. Dabei fällt auf, dass die Nutzung von faktisch und total anonymisierten Datensätzen von den meisten Befragten am stärksten präferiert wird: 69,9 % der Befragten haben für diese Nutzungsart die höchste Präferenz vergeben. Demgegenüber waren dies bei der kontrollierten Datenfernverarbeitung nur 20,6 %, bei der Durchführung eigener Forschungsprojekte in einem statistischen Amt 14,1 % und bei der Durchführung gemeinsamer Forschungsprojekte mit einem statistischen Amt 3,2 %. Betrachtet man die Anteile der zweithöchsten Präferenz, so ist hier das Ergebnis nicht eindeutig: Während 37,1 % der Befragten die kontrollierte Datenfernverarbeitung gewählt haben, ist der Anteil mit 31,8 % bei der Durchführung eige-

5) Für eine zusammenfassende Übersicht über Anonymisierungsmethoden siehe Köhler 1999.

Präferenzen für die einzelnen Nutzungswege								
Ausprägung	Eigene Forschungsprojekte		Gemeinsame Forschungsprojekte		Anonymisierte Datensätze		Kontrollierte Datenfernverarbeitung	
	Anzahl	%	Anzahl	%	Anzahl	%	Anzahl	%
Präferenz 1 (höchste)	62	14,1	14	3,2	337	69,9	94	20,6
Präferenz 2	140	31,8	61	14,1	86	17,8	169	37,1
Präferenz 3	163	37,0	155	35,8	31	6,4	76	16,7
Präferenz 4 (niedrig)	75	17,0	203	46,9	28	5,8	117	25,7
<b>Gesamt</b>	<b>440</b>	<b>100</b>	<b>433</b>	<b>100</b>	<b>482</b>	<b>100</b>	<b>456</b>	<b>100</b>

ner Forschungsprojekte in einem statistischen Amt fast genauso hoch. Bei der Nutzung von faktisch und total anonymisierten Datensätzen und der Durchführung gemeinsamer Forschungsprojekte mit einem statistischen Amt wurde die zweite Präferenz mit 17,8 % und 14,1 % deutlich seltener genannt.

Betrachtet man die Anteile für die dritthöchste Präferenz, so ist dieser bei der Durchführung eines eigenen Forschungsprojektes in einem statistischen Amt (37,0 %) und der Durchführung gemeinsamer Forschungsprojekte in einem statistischen Amt (35,8 %) am höchsten. Bei der kontrollierten Datenfernverarbeitung beträgt dieser Anteil lediglich 16,7 % und bei der Nutzung von faktisch oder total anonymisierten Datensätzen 6,4 %. Bei der niedrigsten Präferenz sind die Ergebnisse hingegen wieder eindeutig: Dieser Anteil ist bei der Durchführung von gemeinsamen Forschungsprojekten in einem statistischen Amt mit 46,9 % am höchsten, gefolgt von der kontrollierten Datenfernverarbeitung (25,7 %), der Durchführung eines eigenen Forschungsprojektes in einem statistischen Amt (17,0 %) und der Nutzung von faktisch oder total anonymisierten Datensätzen (5,8 %).

Zusammenfassend lässt sich festhalten, dass die Befragten der Nutzung von faktisch und total anonymisierten Datensätzen die höchste Präferenz und der Durchführung von gemeinsamen Forschungsprojekten mit einem statistischen Amt die niedrigste Präferenz zuordnen. Die Durchführung eigener Forschungsprojekte und die kontrollierte Datenfernverarbeitung nehmen eine Mittelposition ein und lassen sich nicht in eine eindeutige Rangfolge bringen. Die

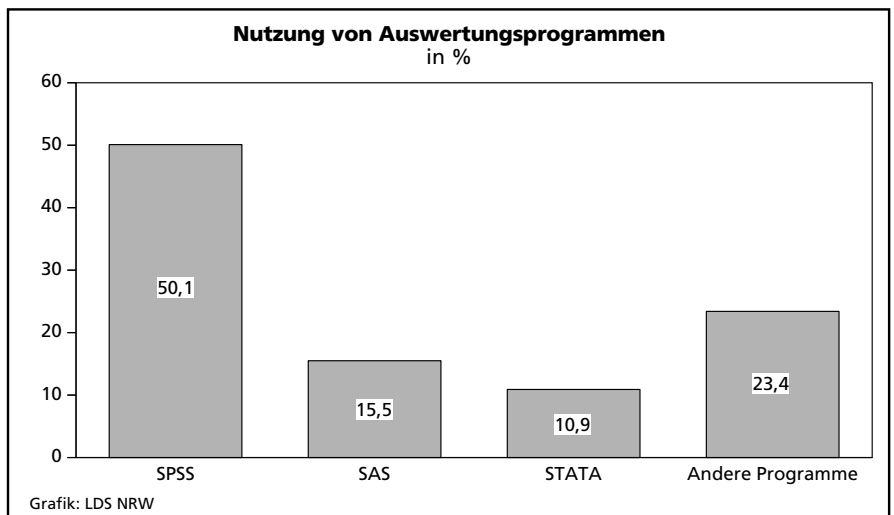
ausgeprägte Präferenz für die Nutzung von faktisch und total anonymisierten Datensätzen war zu erwarten, da dies die einzigste Nutzungsart ist, bei der die Wissenschaftlerinnen und Wissenschaftler an ihrem eigenen Arbeitsplatz direkt mit den Mikrodaten der amtlichen Statistik arbeiten können. Diese Nutzungsart kommt jedoch – wie oben ausgeführt wurde – nicht für alle Erhebungen in Frage.

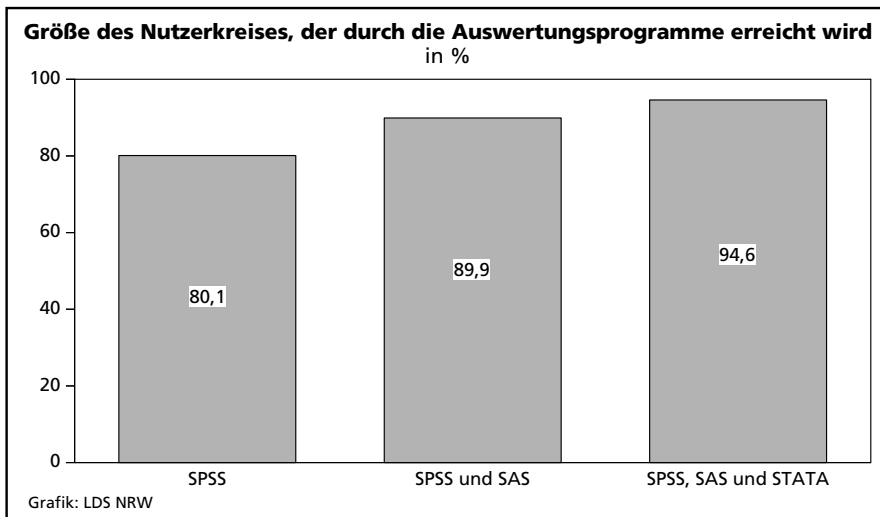
Betrachtet man die Ergebnisse zum Datenbedarf und zu den einzelnen Nutzungswegen, ist ein Zielkonflikt zwischen der Präferenz für anonymisierte Datensätze einerseits und der breiten Streuung des Datenbedarf andererseits zu erkennen. Auf welche Art und Weise dieser Zielkonflikt gelöst werden kann, soll unter anderem im Rahmen des von den statistischen Ämtern geplanten Nutzerworkshops thematisiert werden. Eine mögliche Alternative zur Nutzung von faktisch und total anonymisierten Daten bietet die kontrollierte Datenfernverarbeitung sowie die Einrichtung von Gastwissenschaftlerarbeitsplätzen. Ein Vorteil dieser beiden Nutzungswege besteht darin,

das ein fachlich breit gestreutes Datenangebot aufgebaut werden kann, das gleichzeitig fachlich und regional differenzierte Analysen erlaubt. Aus diesem Grund werden das Forschungsdatenzentrum der statistischen Landesämter und das Forschungsdatenzentrum des Statistischen Bundesamtes neben der Erstellung von faktisch und total anonymisierten Daten die kontrollierte Datenfernverarbeitung aufbauen und Gastwissenschaftlerarbeitsplätze an allen regionalen Standorten einrichten.

## 5 Analyse- und Auswertungsprogramme

In den statistischen Ämtern werden die Mikrodaten vorrangig mit Verbundprogrammen ausgewertet, die in der Regel auf SPLV bzw. STATSPEZ basieren. Die in der Wissenschaft verbreitete Analysesoftware wird bislang nur in einigen statistischen Ämtern eingesetzt. Daher muss für die Forschungsdatenzentren zum einen neue Software für die Gastwissenschaftlerarbeitsplätze und die kontrollierte Datenfernverarbeitung beschafft werden und zum anderen ist das Personal entsprechend zu schulen. Um festzulegen, welche Analysesoftware für die Forschungsdatenzentren der statistischen Landesämter und des Statistischen Bundesamtes prioritär angeschafft werden sollte, wurden die Wissenschaftlerinnen und Wissenschaftler ge-





fragt, welche Auswertungsprogramme sie nutzen. Bei dieser Frage entfiel über die Hälfte der Nennungen auf SPSS, während die anderen Programme eine deutlich geringere Bedeutung haben. So weisen SAS 15,5 % und STATA 10,9 % auf. Auf andere Programme entfallen 23,4 % aller Nennungen. Am häufigsten sind hier GAUSS, EViews, S-PLUS, ACCESS und TDA vertreten.

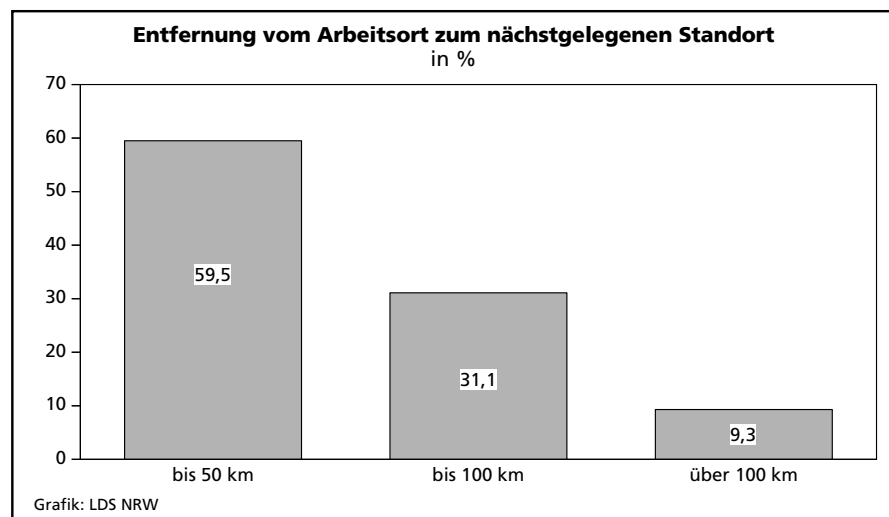
Bezüglich der Anwendung von Auswertungsprogrammen ist auch eine Auswertung auf der Befragtenebene von Interesse, da diese Aufschluss darüber gibt, wie groß der Personenkreis ist, dessen Bedarf mit einem bestimmten Auswertungsprogramm gedeckt werden könnte. Diese Auswertung weist darauf hin, dass 80,1 % aller Befragten unter anderem SPSS für Auswertungszwecke nutzen. Dies bedeutet, dass durch die Anschaffung von SPSS einem sehr großen Teil der potenziellen Nutzerinnen und Nutzer die Auswertung der Mikrodaten der amtlichen Statistik an den Gastwissenschaftlerarbeitsplätzen oder über die kontrollierte Datenfernverarbeitung ermöglicht werden würde. Dieser Anteil kann durch die Anschaffung von SAS um 9,8 Prozentpunkte auf 89,9 Prozent und durch die Anschaffung von STATA um weitere 4,7 Prozentpunkte auf 94,6 Prozent erhöht werden. Dies bedeutet, dass durch die Anschaffung der genannten Standardsoftware fast alle potenziellen Nutzerinnen und Nutzer die Mikrodaten der amtlichen Statistik mit Hilfe ihrer gewohnten

Auswertungsprogramme analysieren können. Gleichzeitig macht die Auswertung jedoch auch deutlich, dass die Anschaffung von SPSS die erste Priorität haben sollte, da mit diesem Programm die Mehrheit der potenziellen Nutzerinnen und Nutzer arbeitet.

## 6 Workshop

Die Forschungsdatenzentren der statistischen Landesämter und des Statistischen Bundesamtes planen, einen Workshop mit dem Ziel durchzuführen, über die Ergebnisse dieser Umfrage zu informieren und mit Vertretern der Wissenschaft zu diskutieren über

- Aspekte der Ausgestaltung der Forschungsdatenzentren,
- unterschiedliche Datenbedarfe und Möglichkeiten ihrer Befriedigung,
- Formen der Kooperation mit der Wissenschaft,



- technische und organisatorische Anforderungen der Wissenschaft,
- Organisation der Kontakte mit der Wissenschaft.

Im Rahmen der Umfrage wurden die Wissenschaftlerinnen und Wissenschaftler daher gefragt, wie hoch ihr Interesse an der Teilnahme an diesem Workshop sei. Es zeigte sich, dass das Interesse der Wissenschaft an diesem Informations- und Kommunikationsangebot sehr hoch ist: Über die Hälfte aller Befragten gaben hier an, dass Sie an einer Teilnahme interessiert seien. Absolut betrachtet entspricht dies einem potenziellen Teilnehmerkreis von 318 Personen.

## 7 Regionale Erreichbarkeit

Wie bereits in Abschnitt 3 erwähnt wurde, können nicht alle Mikrodaten ohne weiteres als faktisch anonymisierte Datensätze für die Wissenschaft zugänglich gemacht werden. Aus diesem Grunde richtet das Forschungsdatenzentrum des Statistischen Bundesamtes an seinem Standort und das Forschungsdatenzentrum der statistischen Landesämter an allen 16 regionalen Standorten Gastwissenschaftlerarbeitsplätze ein. Ziel dieser Maßnahme ist es, den Zugang der Wissenschaft zu den nicht oder nur formal anonymisierten Mikrodaten durch die breite regionale Streuung der Gastwissenschaftlerarbeitsplätze zu erleichtern. Für die Nutzer ist hierbei u. a. von entscheidender Bedeutung, wie leicht die regionalen Standorte



erreicht werden können, d. h. wie weit der nächstgelegene regionale Standorte vom Arbeitsort des Nutzers entfernt ist.

Betrachtet man die Entfernung vom Arbeitsort der Wissenschaftlerinnen und Wissenschaftler zum nächstgelegenen Standort der Forschungsdatenzentren, so haben knapp 59,5 % der möglichen Nutzer eine Distanz von bis zu 50 Kilometer zurückzulegen. Für 31,1 % der Wissenschaftlerinnen und Wissenschaftler käme eine Strecke von 50 km bis 100 km in Betracht und auf 9,1 % der Nutzer entfällt eine Strecke, die über 100 km liegt. Dies bedeutet, dass nur ein sehr geringer Anteil der Nutzer eine relativ weite Strecke zurücklegen müsste, um einen regionalen Standort zu erreichen. Da an den regionalen Standorten sowohl die Mikrodaten aller statistischen Landesämter als auch die Mikrodaten des Statistischen Bundesamtes genutzt werden können, wird durch die Einrichtung von Gastwissenschaftlerarbeitsplätzen an allen regionalen Standorten der Zugang zu den Mikrodaten in entscheidender Weise erleichtert.

## **8 Aktuelles Mikrodatenangebot der amtlichen Statistik und zukünftige Entwicklungen**

Das Angebot der amtlichen Statistik an Mikrodaten für wissenschaftliche Zwecke besteht zur Zeit insbesondere aus faktisch anonymisierten Datensätzen der Haushaltsstichproben. Jedoch werden momentan auch Forschungsprojekte zur Erstellung weiterer faktisch anonymisierter Datensätze, wie zum Beispiel der Lohn- und Einkommensteuerstatistik oder der wirtschaftsstatistischen Mikrodaten, durchgeführt. Folgende Datenbestände stehen bereits heute als faktisch anonymisierte Datensätze bzw. als total anonymisierte Datensätze zur Verfügung:

- Mikrozensus unterschiedlicher Jahrgänge,
- Einkommens- und Verbrauchsstichprobe unterschiedlicher Jahrgänge,

- Zeitbudgeterhebung 1991 (faktisch anonymisierter Datensatz und total anonymisierter Datensatz),
- Sozialhilfestatistik 1998,
- Europäisches Haushaltspanel 1994 bis 1996.

Darüber hinaus werden zur Zeit gemeinsam von den Statistischen Ämtern des Bundes und der Länder folgende Projekte zur Erstellung von faktisch anonymisierten Datensätzen durchgeführt oder befinden sich in Planung:

- Projekt „Faktische Anonymisierung von Einzeldaten über Unternehmen und Betriebe“: Im Rahmen dieses vom Bundesministerium für Bildung und Forschung geförderten Projektes werden gemeinsam von der amtlichen Statistik und der Wissenschaft Anonymisierungsmethoden für wirtschaftsstatistische Mikrodaten untersucht. Ziel des Projektes ist es, faktisch anonymisierte Datensätze verschiedener wirtschaftsstatistischer Erhebungen anbieten zu können<sup>6)</sup>.
- „Regionalfile des Mikrozensus“: Im faktisch anonymisierten Datensatz des Mikrozensus ist aus Datenschutzgründen die Regionalerkennung unterhalb der Ebene der Bundesländer nicht enthalten. Um auch Regionalanalysen unterhalb der Länderebene zu ermöglichen, wird zur Zeit gemeinsam von der amtlichen Statistik und der Wissenschaft ein Anonymisierungskonzept entwickelt, in dem fachliche Merkmale, insbesondere die Angaben zur beruflichen Tätigkeit, in Abhängigkeit von ihrer regionalen Verteilung zusammengefasst werden.

- Projekt „Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe“: Ziel dieses Projektes ist es, den Mikrozensus als Längsschnittdatensatz aufzubereiten und ein geeignetes Hochrechnungs- und Gewichtungungsverfahren zu entwickeln. Das wesentliche Problem, das hierbei gelöst werden muss, besteht darin, dass fort-

<sup>6)</sup> Für weitere Informationen über das Projekt siehe Sturm 2002.

gezogene Haushalte und Personen im Mikrozensus nicht wiederbefragt werden<sup>7)</sup>. Sofern die methodischen Probleme gelöst werden können, soll der Längsschnittdatensatz faktisch anonymisiert und für wissenschaftliche Analysezwecke bereitgestellt werden. Es ist geplant, die methodischen Probleme im Rahmen eines gemeinsamen Forschungsprojektes der amtlichen Statistik und der Wissenschaft zu bearbeiten.

- Weiterhin werden zurzeit Projekte zur „Faktischen Anonymisierung der Lohn- und Einkommensteuerstatistik 1998“ sowie zur „Faktischen Anonymisierung der Straßenverkehrsunfallstatistik“ durchgeführt.

Wie bereits erwähnt, eignen sich nicht alle Daten für eine faktische oder totale Anonymisierung. Daher werden in allen statistischen Landesämtern und im Statistischen Bundesamt Gastwissenschaftlerarbeitsplätze eingerichtet sowie die Infrastruktur für eine kontrollierte Datenfernverarbeitung aufgebaut<sup>8)</sup>. Beide Maßnahmen zielen darauf, nicht oder nur geringfügig anonymisierte Mikrodaten der amtlichen Statistik für wissenschaftliche Analysezwecke zugänglich zu machen. Hierbei werden das Forschungsdatenzentrum des Statistischen Bundesamtes und das Forschungsdatenzentrum der statistischen Landesämter eng kooperieren, so dass sich die Wissenschaftlerinnen und Wissenschaftler nur an eine Stelle wenden müssen, um die Mikrodaten der amtlichen Statistik von allen regionalen Standorten gleichermaßen nutzen zu können.

Die genannten Aufgaben der Forschungsdatenzentren der statistischen Landesämter und des Statistischen Bundesamtes gehen über die genuinen Aufgaben der amtlichen Statistik hinaus. Da die statistischen Ämter der Bereitstellung von Mikrodaten für wissenschaftliche Analyse-

<sup>7)</sup> Für weitere Informationen zu den Längsschnittdaten des Mikrozensus siehe Heidenreich 2002, Zühlke 2001 sowie Zühlke und Hetke 2002. – <sup>8)</sup> Im Statistischen Bundesamt kann die Lohn- und Einkommensteuerstatistik bereits auf der Basis der kontrollierten Datenfernverarbeitung genutzt werden.

zwecke eine hohe Bedeutung zu messen, haben sie Förderanträge gestellt, um die genannten Aufgaben wahrnehmen zu können. Daneben werden die statistischen Ämter diejenigen Mittel, die sie heute schon für die Bereitstellung von Informationen für die Wissenschaft aufwenden, bündeln und für die Aufgaben der Forschungsdatenzentren bereitstellen. Der Aufbau der Forschungsdatenzentren wird jedoch in entscheidender Weise davon abhängen, in welchem Umfang die Wahrnehmung der genannten Aufgaben durch Mittel aus der Wissenschaftsförderung gefördert werden kann.

## 9 Literatur

Heidenreich, Hans-Joachim (2002): Längsschnittdaten aus dem Mikrozensus: Basis für neue Analysemöglichkeiten. In: Allgemeines Statistisches Archiv 2, 213 – 231

Köhler, Sabine (1999): Anonymisierung von Mikrodaten in der Bundesrepublik und ihre Nutzung – Ein Überblick. In: Forum der Bundesstatistik, Bd. 31/1999, Statistisches Bundesamt, 133 – 144

Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (2001): Wege zu einer besseren informationellen Infrastruktur. Frankfurt am Main

Müller, Walter/Blien, Uwe/Knoche, Peter/Wirth, Heike u. a. (1991): Die faktische Anonymität von Mikrodaten. Band 19 der Schriftenreihe Forum der Bundesstatistik. Statistisches Bundesamt, Wiesbaden

Sturm, Roland (2002): Wirtschaftsstatistische Einzeldaten für die Wissenschaft. In: Wirtschaft und Statistik 2, 101 – 109

Tambay, Jean-Louis/Goldmann, Gustave/White, Pamela (2001): Providing Greater Access to Survey Data for Analysis at Statistics Canada. Proceedings of the Annual Meeting of the American Statistical Association, August 5 – 9, 2001

Zühlke, Sylvia (2001): Längsschnittdatenanalysen auf der Basis des Mikrozensus: Methodische Probleme und Lösungsansätze. In: Statistische Analysen und Studien Nordrhein-Westfalen 4, 3 – 13

Zühlke, Sylvia/Hetke, Uwe (2002): Regionale Mobilität und Arbeitsmarktprozesse: Eine Analyse auf der Basis des Sozio-oekonomischen Panels. Statistische Analysen und Studien Nordrhein-Westfalen 2

## 10 Ansprechpartner

<b>Ansprechpartner des Forschungsdatenzentrums des Statistischen Bundesamtes und des Forschungsdatenzentrums der statistischen Landesämter</b>				
Forschungsdatenzentrum/ regionaler Standort	Ansprechpartner			
	Name	Telefon	Fax	E-Mail
<b>Forschungsdatenzentrum des Statistischen Bundesamtes</b>				
Statistisches Bundesamt	Markus Zwick	0611 75 4220	0611 75 3950	markus.zwick@destatis.de
<b>Regionale Standorte des Forschungsdatenzentrums der statistischen Landesämter</b>				
Baden-Württemberg	Dr. Helmut Büringer	0711 641-2418	0711 641-2440	helmut.bueringer@stala.bwl.de
Baden-Württemberg	Ivar Cornelius	0711 641-2570	0711 6412440	ivar.cornelius@stala.bwl.de
Bayern	Dr. Willibald Lausberger	089 2119-601	089 2119-1601	willibald.lausberger@lfstad.bayern.de
Berlin	Prof. Dr. Ulrike Rockmann	030 9021-3502	030 5158-8326	u.rockmann@statistik-berlin.de
Brandenburg	Christiane Siegmund	0331 39431	0331 39487	christiane.siegmund@lds.brandenburg.de
Bremen	Jürgen Wayand	0421 361-2370	0421 496-2370	juergen.wayand@statistik.bremen.de
Hamburg	Sven Wohlfahrt	040 42831-1784	040 427964-032	sven.wohlfahrt@statistik.hamburg.de
Hessen	Dr. Frank Padberg	0611 3802-822	0611 3802-890	fpadberg@hsl.de
Hessen	Günter Grün	0611 3802-810	0611 3802-890	ggruen@hsl.de
Mecklenburg-Vorpommern	Dr. Margit Herrmann	0385 4801-686	0385 4801-123	m.herrmann@statistik-mv.de
Niedersachsen	Rainer Geumann	0511 9898-1612	0511 9898-4111	rainer.geumann@nls.niedersachsen.de
Nordrhein-Westfalen	Helmut Eppmann	0211 9449-2637	0211 9449-2302	helmut.eppmann@lds.nrw.de
Nordrhein-Westfalen	Dr. Sylvia Zühlke	0211 9449-2547	0211 9449-2302	sylvia.zuehlke@lds.nrw.de
Rheinland-Pfalz	Hans Libowitzky	02603 71-280	02603 71-247	hans.libowitzky@statistik.rlp.de
Saarland	Barbara Barth	0681 501-2804	0681 501-2970	dv@stala.saarland.de
Sachsen	Manfred Jakoby	03578 33-1910	03578 33-1999	manfred.jakoby@statistik.sachsen.de
Sachsen-Anhalt	Christoph Schäfer	0345 2318-252	0345 2318-923	schaef@stala.mi.isa-net.de
Schleswig-Holstein	Bernd Struck	0431 6895286	0431 6895498	bernd.struck@statistik-sh.de
Thüringen	Ronald Münzberg	0361 3784-111	0361 3784-697	rmuenzberg@tls.thueringen.de