Zur Lage soziologischer Forschung

Probleme bei der Beschaffung von Daten aus amtlichen Statistiken

Einige Überlegungen aus der Sicht einer demographisch interessierten Sozialwissenschaft

Gerhard Heilig

Lehrstuhl für Bevölkerungswissenschaft, Universität Bamberg, Hornthalstr. 2, D-8600 Bamberg

Einführung

Die amtliche Statistik ist sicher eine außerordentlich wichtige Basis vieler empirisch arbeitender Sozialwissenschaften, insbesondere jener Teilgebiete, die sich mit demographischen und familiensoziologischen Fragen befassen. Zwar wird man sich ergänzend auch der Erhebungsmethoden der empirischen Sozialforschung bedienen und beispielsweise Daten aus Survey-Erhebungen benutzen. Solche Studien ersetzen jedoch keineswegs massenstatistische Erhebungen (wie Volkszählung und Mikrozensus) oder die laufende amtliche Erfassung demographischer Ereignisse (wie beim Geburten- und Sterberegister). Im Gegenteil: sozialwissenschaftliche Primärerhebungen zu sozio-demographischen Problemen erhöhen tendenziell sogar die Nachfrage nach amtlichen Eckdaten. Diese werden beispielsweise bei der Stichprobenerstellung, zur Berechnung von Gewichtungsfaktoren oder als Bezugsgrößen bei der Auswertung benötigt.

Die amtliche Statistik liefert aber der Sozialwissenschaft nicht nur Rahmendaten für deren Erhebungen, sondern erstellt auch räumlich und sachlich tief gegliederte Datensätze, die sich für tiefergehende (statistische) Analysen eignen. Karl Schwarz hat mehrfach darauf hingewiesen, daß gerade der Mikrozensus eine unschätzbare Datenbasis für die sozialwissenschaftliche Forschung darstellt, die noch längst nicht voll ausgeschöpft ist.

Man kann die These aufstellen, daß sich hier für die amtliche Statistik, neben dem reinen Zählen und Darstellen von (demographischen) Sachverhalten, ein zusätzlicher Aufgabenbereich eröffnet hat. Dabei geht es um die Bereitstellung von Datensätzen, die (vom Umfang der Merkmalskataloge her gesehen) auch als Basis für eine quantitative *Ursachenforschung* geeignet sind. Ein Beispiel dafür ist der außerordentlich differenzierte Fragen-

katalog der letzten Mikrozensus-Erhebung. Umso wichtiger ist es, daß die hier erhobenen Daten nicht nur zur laufenden Berichterstattung der Statistischen Ämter, sondern auch für die Analysen der Wissenschaft optimal zugänglich sind. Dabei gibt es allerdings in der Bundesrepublik Deutschland gewisse Probleme. Sie stammen aus zwei grundlegenden Quellen: Aus den Modalitäten der Datenweitergabe an die Wissenschaft und aus der Struktur und dem Merkmalskatalog der laufenden Statistiken. Beide Aspekte sollen nun näher beleuchtet werden.

1. Probleme bei der Weitergabe von amtlichen Daten an die Sozialwissenschaft

Im Jahre 1982 wurde auf der 29. Tagung des Statistischen Beirates das Generalthema „Interdependenzen zwischen amtlicher Statistik und empirischer Sozialwissenschaft“ behandelt. In seinem Referat über die amtliche Statistik aus der Sicht der empirisch orientierten Forschung sagte Prof. Walter Müller u. a.:

„Ich kann nicht umhin, einige Punkte klar beim Namen zu nennen, denn die gegenwärtige Situation ist weniger als suboptimal. Sie hat sich gegenüber den Möglichkeiten, die noch vor einigen Jahren bestanden, deutlich verschlechtert, und sie ist auch außerordentlich unbefriedigend, wenn wir die Gegebenheiten in der Bundesrepublik mit den Bedingungen vergleichen, die für die Forschung in anderen Ländern bestehen.“ (Müller, 1982, S. 7).

Diese harsche Kritik an der Praxis der Datenbereitstellung durch die Statistischen Ämter hat in den vergangenen Jahren nichts an ihrer Berechtigung verloren. Im Gegenteil: Durch die teilweise hysterische Diskussion um den Datenschutz sahen sich die Statistischen Ämter in der Bundesrepublik offenbar gezwungen, bei ihrer Weitergabep Praxis ausgerechnet gegenüber der Wissenschaft im allgemeinen noch restriktiver und vorsichtiger zu ver-

fahren, als dies ohnehin schon der Fall war. Vergleicht man die Situation in der Bundesrepublik mit den Gegebenheiten in anderen europäischen Ländern, so wird die Behinderung der deutschen Sozialwissenschaft bei der Analyse amtlich erhobener Datenbestände überdeutlich. Es sind vor allem drei Probleme, die bei der Datenweitergabe der Statistischen Ämter an die (universitäre) Wissenschaft auftreten:

- Probleme, die mit dem *Datenschutz* zusammenhängen bzw. mit der unangemessen restriktiven Weitergabepaxis statistischer Ämter.
- Finanzielle Schwierigkeiten, die deshalb entstehen, weil die universitäre Forschung meist nur mit äußerst *knappen Ressourcen* ausgestattet ist. Deshalb ist es oft unmöglich, Sonderauswertungen in den Ämtern vornehmen zu lassen.
- Vor allem aber Probleme, die sich aus der *Art der Datenweitergabe* ergeben. Dabei geht es darum, daß amtliche Statistiken überwiegend in Form von Standardtabellen bereitgestellt werden, während der Bedarf der empirischen Forschung sehr häufig in eine ganz andere Richtung geht.

So verständlich das Bemühen der Statistischen Ämter ist, den Schutz der von ihnen weitergereichten Daten unter allen Umständen zu gewährleisten, so problematisch ist es, wenn unter Hinweis auf die gesetzliche Verpflichtung aus dem *Datenschutz*, die Wissenschaft daran gehindert wird, die amtlicherseits erhobenen statistischen Daten so auszuwerten, wie es dem Stand der Forschung entsprechen würde. So werden beispielsweise anonymisierte Einzelfalldatensätze aus dem Mikrozensus in der Bundesrepublik entweder gar nicht oder nur nach einer außerordentlich aufwendigen Einzelfallanforderung weitergegeben. In der Schweiz oder in Österreich ist eine derartige Form der Datenweitergabe an der Tagesordnung, wobei es meines Wissens von seiten der Wissenschaft noch zu keiner Verletzung der Anonymität gekommen ist. Man kann sich des Eindruckes nicht erwehren, als würden manche Vertreter einer restriktiven Weitergabepaxis in der Bundesrepublik stillschweigend davon ausgehen, die Wissenschaft würde die ihr überlassenen Daten lediglich zum Zwecke der De-Anonymisierung auswerten. Ein wenig Vertrauensvorschuß würde hier – für beide Seiten – die Zusammenarbeit erheblich erleichtern. So mutet es manchmal schon etwas komisch an, wenn einzelne Statistische Ämter in umfangreichen Tabellen von Sonderauszählungen mit „Tipp-Ex“ (!) alle Zellenbesetzungen unter 3 unkenntlich machen – so als hätte der Forscher nichts Besseres

zu tun, als beispielsweise die drei ausländischen Frauen ausfindig zu machen, die in der Stadt „X“ wohnen, älter als 16 Jahre sind und ein Kind bekommen haben. Problematisch ist eine solche Praxis für die Forschung, weil sich aus den so zensierten Tabellen keine nachträgliche Zusammenfassung von Merkmalen auszahlen läßt.

Die Statistischen Ämter gehen davon aus, daß bei den an die Wissenschaft weitergegebenen Daten die Möglichkeit einer De-Anonymisierung *zweifelsfrei* ausgeschlossen sein muß, daß sich also beispielsweise auch nicht durch Kombination von Merkmalen Einzelfälle identifizieren lassen dürfen. Je nachdem, welches methodische Raffinement man den Benutzern unterstellt, läßt sich damit die Meßlatte für die Datenweitergabe fast beliebig heraufsetzen. Zumindest kann man damit die Überlassung von Einzelfalldaten fast immer verhindern – schon aus dem einfachen Grund, weil sich durch entsprechende Kombination von Merkmalen aus dem Datenmaterial fast immer Tabellen bilden lassen, deren Zellenbesetzung vereinzelt unter 3 sinkt. Das Sicherheitsniveau von Datenschutz soll aber, nach Aussage des ehemaligen Datenschutzbeauftragten Hans Peter Bull, dem praktischen Schutzbedürfnis der Betroffenen dienen und nicht einer theoretisch unüberwindbaren Anonymisierung. (Bull, 1982)

Deshalb wäre meines Erachtens im Hinblick auf die Datenweitergabe der Statistischen Ämter ein weniger restriktiver Begriff der „Anonymisierung“ wünschenswert und vertretbar. Beispielsweise könnte man Datensätze als anonymisiert betrachten, bei denen bestimmte, eindeutige Identifizierungsmerkmale (z. B. der Name) gelöscht sind. Auf das Kriterium der Reidentifizierbarkeit durch Merkmalskombination würde man dabei im Interesse der Forschung verzichten. Der Datenschutz ließe sich trotzdem wirkungsvoll sichern, indem man den wissenschaftlichen Benutzer besonders auf Geheimhaltung verpflichten würde. Es ist schließlich nicht ganz einzusehen, warum Hunderte von Datentypistinnen in den Statistischen Ämtern jeden Tag mit „Einzelfalldaten“ umgehen dürfen, während der (meist ohnehin beamtete und zur Verschwiegenheit in Dienstangelegenheiten verpflichtete) Wissenschaftler an einer Universität wie ein feindlicher Dechiffrierspezialist behandelt wird. Die Statistischen Ämter der Länder und des Bundes haben in den letzten Jahren Schritte unternommen, um das Problem der Anonymisierung von Einzelfalldaten zu lösen. So wurde erst kürzlich (3.–5. März, 1986) am Statistischen Bundes-

amt ein wissenschaftliches Kolloquium durchgeführt, das sich mit den „Bedingungen und Möglichkeiten der Nutzung von anonymisierten Einzelangaben (insbesondere Public-Use Samples) aus Daten der Amtlichen Statistik“ befaßte. Es ist zu hoffen, daß sich dadurch für die Sozialwissenschaft die Zugangsmöglichkeiten zu Einzelfalldatensätzen (vor allem aus dem Mikrozensus) verbessern werden.

Ein zweites generelles Problem bei der Datenweitergabe durch die Statistischen Ämter in der Bundesrepublik sind bisweilen die hohen *Kosten*, die mit Sonderauszählungen oder auch nur mit der Erstellung von Kopien intern vorliegender Tabellen verbunden sind. Zwar bemühen sich verschiedene Ämter, besonders das Statistische Bundesamt, in vorbildlicher Weise, die Kosten von Sonderauszählungen niedrig zu halten oder (bei bestimmten Voraussetzungen) ganz zu erlassen. Andererseits scheinen einzelne Statistische Ämter Kostenvoranschläge aber auch als Abschreckungsmittel gegenüber aufwendigen Auszählungswünschen der Wissenschaft zu betrachten. Auch hier könnte eine Sonderregelung für die Wissenschaft den Datenzugang wesentlich verbessern. Denkbar wäre beispielsweise, daß Datenaufbereitung und Datenweitergabe (bis zu einer bestimmten Höchstgrenze) für wissenschaftliche Benutzer grundsätzlich kostenlos ist. Eine ähnliche Regelung zugunsten der (universitären) Forschung besteht bereits bei der Benutzung von Großrechenanlagen an den regionalen Rechenzentren, wo es, wegen der außerordentlich teuren Rechenzeit, oft um erhebliche Beträge geht.

Das weitaus wichtigste Problem, mit dem sich die Wissenschaft bei der Beschaffung von Daten aus den Statistischen Ämtern konfrontiert sieht, ist jedoch die *Art der Datenweitergabe*. Das Auswertungsprogramm der amtlichen Statistik ist weitgehend auf die Bereitstellung von Standardtabellen eingestellt. Diese Verfügungsweise amtlicher Daten entspricht jedoch heute nicht mehr allen Ansprüchen der empirischen Sozial- und Bevölkerungswissenschaft. Dafür sind vor allem drei Entwicklungstendenzen in der empirischen Forschung verantwortlich:

- die zunehmende Verwendung von multivariaten Analyseverfahren,
- die Entwicklung von Mikro-Simulationsmodellen und
- die allgemeine Verfügbarkeit leistungsfähiger Datenbank- und Statistiksoftware (wie SIR, SPSS, SAS oder BMDP), die auch dem universitären

Forscher die Arbeit mit sehr umfangreichen Datensätzen erlaubt und einen sukzessiven Analyseprozeß ermöglicht.

Nach dem Stand der Wissenschaft erfordern viele Fragestellungen den Einsatz multivariater (statistischer) Analyseverfahren, da die herkömmlichen, bivariaten Auswertungen, wie sie durch die Standardtabellen nahegelegt werden, häufig Scheinergebnisse erbrachten. Schon seit langem wird deshalb mit den einfachen Verfahren der Standardisierung von Tabellen gearbeitet. Die modernen, multivariaten Verfahren sind in gewisser Hinsicht eine Weiterentwicklung der Tabellenstandardisierung. Sie benötigen jedoch fast immer Daten, die auf den Einzelfall bezogen sind. Eine Verhinderung multivariater Analysen durch die Art der Datenbereitstellung wäre gleichbedeutend mit dem Verzicht auf den heutigen Stand der Auswertungsmethodik.

Die Entwicklung von Mikro-Simulationsmodellen ist deshalb so wichtig, weil soziale und demographische Phänomene einen dynamischen Charakter haben. Es handelt sich fast immer um Abläufe, die durch viele Faktoren gleichzeitig bestimmt werden, wobei es zu komplizierten Wechselwirkungen kommen kann. Mikro-Simulationsmodelle versuchen solche sozialen Systeme oder Ablaufsequenzen mathematisch nachzubilden. Zur Schätzung der Modellgleichungen und zum Test dieser Modelle werden im allgemeinen ebenfalls Einzelfalldaten benötigt.

Schließlich muß darauf hingewiesen werden, daß die empirische Sozialforschung zunehmend Fragestellungen entwickelt, für die die Datensätze außerordentlich flexibel und manipulierbar sein müssen. Zum Beispiel ist es häufig nach Durchsicht von Tabellenauszahlungen notwendig, Merkmalsausprägungen zusammenzufassen, weil die entsprechenden Zellen zu schwach besetzt sind. Oft sind auch Umkodierungen oder die Bildung von Indizes angebracht, und schließlich muß die Möglichkeit bestehen, Skalierungsverfahren einzusetzen oder den untersuchten Merkmalsraum mit Hilfe von Cluster- oder Faktorenanalysen auf wenige, aussagekräftige Dimensionen zu reduzieren. Dieses moderne Instrumentarium der Datenaufbereitung und Datenanalyse kann heute mit Hilfe ausgereifter Statistik- und Datenbanksoftware an jedem Universitätsrechenzentrum eingesetzt werden. Viele Sozialwissenschaftler wurden, besonders in den letzten Jahren, während ihres Studiums speziell darin ausgebildet, diese Methoden der Datenaufbereitung und -analyse sachkundig einzusetzen.

zen. Dies alles ist aber nur möglich, wenn anonymisierte Einzelfalldatensätze von den Statistischen Ämtern zur Verfügung gestellt werden. So manche Sonderauswertung könnte der Wissenschaftler dann wesentlich schneller und flexibler selbst durchführen als dies heute durch die Statistischen Ämter geschehen kann.

Das Standardtabellenprogramm der Statistischen Ämter reicht heute also nicht mehr aus, um empirische Ursachenforschung mit Hilfe amtlicher Daten durchzuführen. Der Bedarf der Wissenschaft an anonymisierten Einzelfalldatensätzen ergibt sich aus methodischen und theoretischen Entwicklungen der empirischen Forschung. Eine fortdauernde Zurückweisung dieses Bedarfs würde den wissenschaftlichen Fortschritt erheblich behindern.

Andere Länder haben dies längst erkannt. In den USA profitiert die Sozialwissenschaft von den sog. „public-use files“. Das sind amtliche (oder halbamtliche) Datensätze, die für die allgemeine Benutzung freigegeben sind und auf Datenband weitergereicht werden. Zum Teil handelt es sich um anonymisierte Einzelfalldatensätze oder zumindest um räumlich und sachlich außerordentlich tief gegliederte Datenbestände, die von der Wissenschaft – aber auch von kommerzieller Seite – wesentlich effektiver aufbereitet werden können als dies einem Statistischen Amt möglich ist. Vielleicht sollte hier erwähnt werden, daß in den USA ein ganzer Industriezweig aus dem Boden geschossen ist, der beispielsweise die Volkszählungsdaten optimal (z. B. mit thematischer Kartographie oder in Form hervorragender Graphiken) aufbereitet und für bestimmte Klientengruppen auswertet. Das Statistische Amt erstellt dafür gewissermaßen nur das Rohmaterial, während Wissenschaft und kommerzielle Institute die Datenaufbereitung und Präsentation vornehmen. Sicher kann man dieses „kommerzielle Modell“ realistischerweise nicht für die Bundesrepublik Deutschland empfehlen, aber zumindest sollte die (universitäre) Forschung bei amtlichen Daten nicht ganz von der Verwendung modernen Auswertungs- und Analyseverfahren abgeschnitten werden.

In der gegenwärtigen Situation muß man leider davon ausgehen, daß ein erheblicher Teil der von den statistischen Ämtern gesammelten Daten nicht oder nur zum Teil ausgewertet wird. Die Ämter haben für tiefergehende Analysen im allgemeinen weder die zeitlichen noch die personellen Möglichkeiten, während die Wissenschaft zu etlichen Datenbeständen nur unzureichenden Zugang hat.

Nur eine intensive Nutzung der amtlichen Daten ist aber die Rechtfertigung für eine zunehmende Belastung des Bürgers durch statistische Erhebungen. Sammeln und Horten statistischer Daten sowie die bloße Aufbereitung in (manchmal höchst „unverdaulichen“) Standardtabellen kann nicht das Ideal der amtlichen Statistik sein. In ihrem eigenen Interesse muß sie sich um eine möglichst große Mehrfachnutzung ihrer Datenbestände bemühen. Ebenso hat natürlich die Wissenschaft nicht nur das Recht, einen besseren Zugang zu den Daten der statistischen Ämter zu fordern, sondern auch die Pflicht, sich über die vorhandenen, teilweise noch ungenutzten Möglichkeiten zu informieren.

2. Probleme, die sich aus der Struktur und dem Merkmalskatalog der vorhandenen Statistiken ergeben

Der zweite Aspekt, aus dem sich für die Wissenschaft Probleme bei der Datenbeschaffung ergeben können, sind die Struktur und der Merkmalskatalog der bei den Statistischen Ämtern erstellten Datenbestände.

Statistische Ämter müssen sich bei der Festlegung ihrer Erhebungsprogramme sowie bei der Aufbereitung der Statistiken an einem weiten Benutzerkreis orientieren. Es kann deshalb gar nicht ausbleiben, daß einzelne Benutzer ihre Vorstellungen darüber, welche Daten wichtig sind, nicht vollständig durchsetzen können. Um einen vernünftigen Ausgleich der manchmal divergierenden Interessen zu schaffen, wurden statistische Beiräte an den Ämtern geschaffen, in denen sich die verschiedenen Interessengruppen artikulieren können.

Nun muß man leider feststellen, daß die demographisch interessierte Sozialwissenschaft, schon allein wegen der kleinen Zahl der heute in der Bundesrepublik mit ihr Befassten, keine sehr starke Lobby für ihre Datenwünsche mobilisieren konnte. So ist es nicht verwunderlich, wenn die Bevölkerungstatistik zu etlichen wichtigen soziodemographischen Tatbeständen kaum Daten erhebt und aufbereitet. Das gilt für die Volkszählung und, in ganz besonderem Maße, für den Mikrozensus. Drei Tendenzen sind dabei vor allem zu beachten:

- Die Ausdünnung des Erhebungsprogrammes auf Kosten demographischer und sozialer Merkmale,
- die einseitige Konzentration des Fragenprogrammes beim Mikrozensus auf die Untersuchungseinheit „Haushalt“. Dies bedeutet eine

Ausblendung wichtiger „personenbezogener“ demographischer Merkmale (wie z. B. die Zahl der Kinder, die eine Frau im Laufe ihres Lebens geboren hat),

- den Verzicht auf spezielle Sondererhebungen zu drängenden demographischen Problemen (wie beispielsweise der Verzicht auf die Teilnahme am „World-Fertility-Survey“).

Die *Ausdünnung des Erhebungsprogrammes* auf Kosten demographisch relevanter Daten zeigt sich deutlich an den Zusatzerhebungen im Rahmen des Mikrozensus. Diese Verarmung im sozial- und bevölkerungswissenschaftlichen Erhebungsprogramm steht im paradoxen Gegensatz zur gesellschaftlichen Entwicklung in der Bundesrepublik. Überspitzt könnte man sagen: je mehr die Geburtenzahlen absanken, je stärker die Scheidungszifferen nach oben schnellten und die Heiratsneigung zurückging, desto weniger Informationen über die Hintergründe dieser Entwicklungen wurden von der amtlichen Statistik gesammelt und der Wissenschaft bereitgestellt. Angesichts der gravierenden Veränderungen im generativen Verhalten der Bevölkerung während der letzten 10 Jahre ist es nur schwer verständlich, warum zu diesen aktuellen Problemen keine eigenen Zusatzbefragungen im Rahmen des Mikrozensus durchgeführt wurden.

Vollends unerklärlich ist mir jedoch, warum sich im Fragenprogramm der geplanten Volkszählung 1986 nicht ein einziges jener demographischen Probleme widerspiegelt, die im letzten Jahrzehnt der Bundesrepublik den Welt-„Rekord“ als Land mit der niedrigsten Fruchtbarkeit und einem seit Jahren absoluten Bevölkerungsrückgang (aus der natürlichen Bevölkerungsbewegung) eingebracht haben. Während beispielsweise bei der Volkszählung im Mai 1970 (10% der) Frauen nach den Geburtsjahren aller lebendgeborenen ehelichen Kinder gefragt wurden, ist eine vergleichbare Frage in der geplanten Volkszählung 1986 *nicht* vorgesehen. In anderen Ländern wird bei Volkszählungen und großen amtlichen Surveys nicht nur die Zahl bisheriger Lebendgeburten, sondern oft eine Geburtenbiographie bei allen Frauen (also auch bei den ledigen) routinemäßig abgefragt. Und da verzichtet ausgerechnet jenes Land auf diese schlichte Frage, das seit Jahren die negativste Geburtenbilanz aufweist!

Natürlich ist diese unglückliche Entwicklung vor allem der öffentlichen Diskussion um den Datenschutz anzulasten. Die teilweise hysterischen Reaktionen von Teilen der Öffentlichkeit auf die Volkszählung hat es dem amtlichen Statistikern

wahrlich nicht leicht gemacht, das minimale Fragenprogramm zu „retten“, das für eine Volkszählung unabdingbar ist. Aber die Frage muß erlaubt sein, ob nicht auf einige weniger wichtige Merkmale zugunsten grundlegender demographischer Merkmale hätte verzichtet werden können. Dem mit Bevölkerungsfragen befaßten Sozialwissenschaftler jedenfalls leuchtet nicht ein, warum eine Vielzahl detaillierter Fragen zur Wohnsituation im Erhebungsbogen enthalten sein muß, während ganz grundlegende demographische Tatbestände nicht erfaßt werden. In der heutigen sozialen Situation der Bundesrepublik ist eine Frage nach der bisherigen Kinderzahl oder dem Heirats- bzw. Scheidungsdatum wohl ebenso wichtig, wie die Frage nach der Art einer eventuellen Mietpreismäßigung für die gemietete Wohnung, oder danach, wie lange man für den Weg zur Arbeit benötigt. „Ist die Wohnung (der befragten Person, d. V.) vom Eigentümer verbilligt, kostenlos überlassen oder die Miete wegen finanzieller Vorleistungen ... ermäßigt“ (Entwurf für Volkszählung 1986, Wohnungsbogen, Frage 10b).

Damit kommen wir zu einem weiteren heiklen Thema, daß ich dennoch hier in aller Deutlichkeit ansprechen möchte: Die Verarmung des sozial- und bevölkerungswissenschaftlichen Erhebungsprogrammes (z. B. beim Mikrozensus) ist keineswegs gleichbedeutend mit einer Entlastung der Befragten. Sieht man sich den Fragebogen der letzten Mikrozensuserhebung an, so wird man zugeben müssen, daß er wohl den umfangreichsten und differenziertesten Fragenkatalog seit langem enthält. Nur leider wurden nicht jene Aspekte zusätzlich berücksichtigt, die dem Demographen und Soziologen am Herzen liegen, sondern vor allem haushalts- und wirtschaftsstatistische Tatbestände. Ganz offensichtlich verstehen es die Wirtschaftswissenschaften sowie staatliche Planungs- und Verwaltungsinstanzen sehr viel besser, ihre Datenwünsche zur Geltung zu bringen, als die Bevölkerungswissenschaft und die Sozialforschung. Dies ist im übrigen eine Diagnose, die im Grundsatz bereits 1982 Prof. Walter Müller auf der schon erwähnten Tagung des Statistischen Beirates stellen mußte (s. 8).

Hier einige wenige Beispiele von Themenkreisen, für die sich Bevölkerungswissenschaftler eine stärkere Berücksichtigung im Fragenprogramm eines Mikrozensus wünschen würden:

- Die demographische Struktur und soziale Situation der nichtehelichen Lebensgemeinschaften,
- die Erfassung einer vollständigen Geburtenbio-

graphie aller Frauen (also nicht nur der verheirateten), wie sie im österreichischen Mikrozensus schon mehrfach durchgeführt wurde,

- die *gleichzeitige* Erhebung einer vollständigen Heirats-, Geburten- und Erwerbsbiographie, um auf vernünftige Weise lebenszyklische Zusammenhänge zwischen Erwerbsverhalten und generativem Verhalten analysieren zu können (damit nicht weiterhin, wie es jetzt geschieht, mit akrobatisch anmutenden Verrenkungen komplizierte Filterauszählungen vorgenommen werden müssen, um wenigstens einige grundlegende Zusammenhänge analysieren zu können).

Insbesondere der letztgenannte „Datenwunsch“ ist von großer Bedeutung. Er zeigt im übrigen, daß es gar nicht primär darum geht, möglichst viele „weiche“ Merkmale zu erfassen (wie Statistiker den Sozialwissenschaftlern gerne unterstellen), sondern nur auf vernünftige Weise grundlegende sozio-demographische Fakten (wie Heirat, Scheidung, Erwerbstätigkeit, Geburt eines Kindes) im Erhebungskonzept des Mikrozensus berücksichtigt werden sollten.

Als ein weiteres Problem der Datenbeschaffung, das sich aus den Statistiken selbst ergibt, haben wir die *einseitige Fixierung auf den „Haushalt“* als Untersuchungseinheit beim Mikrozensus kritisiert. Diese Dominanz der wirtschafts- und haushaltsstatistischen sowie der planungstechnischen Sichtweise zeigt sich ganz deutlich im Fragenkatalog. Die erhobenen Tatbestände beziehen sich fast ausschließlich auf diese Bezugseinheit bzw. die in ihr lebenden Personen. Beispielsweise ist so die Zahl der Kinder (unter 18 Jahren) erfaßt, die in diesem Haushalt leben – einerlei, ob sie leibliche Kinder eines der erwachsenen Haushaltsmitglieder sind oder nicht. Völlig irrelevant ist dabei auch, ob eines der erwachsenen Haushaltsmitglieder noch weitere (lebende oder schon gestorbene) Kinder außerhalb des Haushaltes hat. Nun, genau dies ist aber aus bevölkerungswissenschaftlicher Sicht eine höchst interessante Frage, besonders seit immer mehr Kinder nicht in jenen Familien aufwachsen, in die sie hineingeboren wurden. (Man nimmt an, daß heute in Großstädten wegen der zunehmenden Scheidungen 30 bis 40% aller heute geborenen Kinder nicht in dem Haushalt groß werden, in den sie hineingeboren wurden).

Kinder interessieren den Sozialwissenschaftler nicht primär als Wirtschaftsfaktor eines Haushaltes, sondern als „Endergebnis“ eines bestimmten sozialen Verhaltens. Ein solches „generatives“ Verhalten ist aber an *Personen* geknüpft und nicht

an *Haushalte*. Gerade die jüngsten demographischen Entwicklungen (wie die enorme Zunahme der Scheidungen, die zunehmende Verbreitung nichtehelicher Lebensgemeinschaften und sog. „unvollständiger“ Familien) in der Bundesrepublik machen deutlich, daß sich der generative Prozeß in seinen verschiedenen Phasen durchaus auf *verschiedene* Haushalte verteilen kann.

Das Ablaufgeschehen von Partnerfindung, Heirat, Elternschaft, eventuell Scheidung mit erneuter Heirat und Elternschaft usw. ist ein einziger, „ganzheitlich“ zu analysierender Vorgang, der verschiedene Haushalte durchläuft, ja zum Teil in mehreren Haushalten gleichzeitig stattfinden kann (etwa, wenn eine Ehe geschieden ist, oder ein Teil der Kinder bereits die elterliche Wohnung verlassen hat). Die einzig sinnvolle Art, die Tatbestände dieses „generativen Verhaltens“ zu erfassen, ist es deshalb, sie mit den beteiligten Personen zu verknüpfen, die ihre „generative Geschichte“ gewissermaßen mit sich tragen, auch wenn sie (beispielsweise durch einen Partnerwechsel) die Haushaltszugehörigkeit ändern.

Will der Bevölkerungswissenschaftler mit Hilfe des Mikrozensus die durchschnittliche Kinderzahl in den heutigen Ehen bestimmen, dann muß er mit einem ungenauen Hilfskonstrukt: „Kinderzahl (unter 18 Jahren) in der Familie“ vorliebnehmen. Die Berechnung einer „completed fertility rate“ für ältere Ehepaare ist aus dem Mikrozensusmaterial ganz unmöglich, weil ja die Kinder längst den Haushalt der Eltern verlassen haben. Wie einfach wäre es, wenn man im Mikrozensus jede Frau einfach danach fragen würde, wieviele Kinder sie (bisher) geboren hat.

Damit kein Mißverständnis auftritt: Es geht nicht darum, den Mikrozensus als Haushaltsstichprobe zu kritisieren oder gar das Haushaltskonzept als solches abzulehnen. Auch die Bevölkerungswissenschaft ist an „Haushalten“ und ihrer demographischen Zusammensetzung höchst interessiert. Es kann gar kein Zweifel darüber bestehen, daß der Haushalt eine wichtige analytische Bezugseinheit für viele Untersuchungen darstellt. Insofern ist es auch unbedingt erforderlich, die Stellung der Haushaltsmitglieder zueinander zu erfassen, damit (zumindest nachträglich) die Bildung von Haushaltstypen möglich ist. Es geht lediglich um die (zumindest partielle) Trennung von Erhebungs- und Untersuchungseinheit. Niemand hat etwas gegen den Mikrozensus als Haushaltsstichprobe. Nur sollten auch einige, für die Bevölkerungswissenschaft zentrale Fragestellungen erhoben werden,

die sich auf die Lebensgeschichte der Personen beziehen und über den gerade bestehenden Haushaltszusammenhang hinausweisen können. Andernfalls sind z. B. empirische Untersuchungen, die sich an den höchst interessanten biographietheoretischen Ansätzen orientieren, mit amtlichen Daten kaum durchführbar.

Ein weiteres Problem der Datenbeschaffung betrifft schließlich das völlige *Fehlen thematischer Schwerpunkterhebungen* zu aktuellen demographischen Entwicklungen, wie z. B. die Durchführung eines Fertility-Survey. Die Statistischen Ämter der Bundesrepublik hätten das „Know-How“ und die technischen Voraussetzungen für sehr große Stichprobenuntersuchungen (etwa in der Größenordnung zwischen Mikrozensus und den Stichprobengrößen der kommerziellen Umfrageforschung), mit denen regional und sachlich tief gegliederte Datensätze zu bestimmten sozialen und demographischen Themenschwerpunkten erstellt werden könnten (z. B. Partnerschafts- und Heiratsverhalten, Reproduktion, Scheidungen). In vielen Ländern der Erde gibt es (zusätzlich zum normalen statistischen Programm) solche Surveys von amtlichen Stellen, die mit ihrem ausgeweiteten Fragenprogramm und ihrer verhältnismäßig kleinen Stichprobe eine aktuelle Sozialberichterstattung erlauben.

Ich fasse zusammen: Sowohl die Weitergabemodalitäten der Statistischen Ämter als auch die amtlichen Datenbestände selbst können Probleme bei der Datenbeschaffung der Bevölkerungswissenschaft aufwerfen. Dabei spielen auch Faktoren eine Rolle, die weder die Statistischen Ämter noch die Wissenschaftler zu verantworten haben, wie die teilweise überzogenen Datenschutzforderungen aus dem politischen Raum. Im Falle der kommenden Volkszählung ist zu befürchten, daß der „schwächste“ Interessent an den Daten, nämlich die universitäre Sozialforschung, die größten Schwierigkeiten haben wird, überhaupt irgendwelche tief gegliederte Daten zu erhalten, ganz abgesehen davon, daß viele, für den demographisch interessierten Sozialwissenschaftler wesentliche Fragen ohnehin nicht im Erhebungsprogramm enthalten sein werden. Ausgerechnet die Forschung an den Universitäten ist also, so scheint es, der große „Verlierer“ der Datenschutzdebatte.

3. Der Datenbedarf aus der Sicht der (bevölkerungswissenschaftlichen) Theoriebildung – das Beispiel des „generativen Verhaltens“

In den vorangegangenen beiden Abschnitten haben wir uns mit Datenbeschaffungsproblemen aus der Blickrichtung der (amtlichen) Datenbereitstellung befaßt. Dies ist aber nur *eine* Perspektive, aus der man das Problem betrachten kann. Eigentlich sind in unserem Zusammenhang nämlich *nicht* die „Daten“ das Entscheidende, sondern die „Theorien“. Daten bekommen ihre Bedeutung erst im Lichte einer theoretischen Interpretation. Es ist ein Irrtum zu glauben, die Daten sprächen schon für sich selbst. Bereits die Auswahl der als relevant angesehenen Tatbestände, die dann als (amtliche) „Daten“ Leben gewinnen, ist ein Akt bewußter (häufig aber auch unbewußter) Selektion unter Alternativen auf dem Hintergrund einer bestimmten Deutung der Welt. Auch die Aufbereitung der Rohdaten und die Präsentation der Statistiken ist keinesfalls frei von interpretativen, d. h. letztlich auf dem Hintergrund von „Theorien“ gefällten Entscheidungen. Es ist deshalb zu fragen, ob von der amtlichen Statistik jene Daten erhoben, aufbereitet und weitergereicht werden, die (wenigstens teilweise) dem konzeptionellen Gebäude der (demographisch interessierten) Sozialwissenschaft entsprechen.

Die empirische Forschung benötigt vor allem solche Daten, mit denen sie *Hypothesen* testen kann, und weniger solche, die aufgrund eines bestimmten Erhebungszusammenhanges gerade anfallen. Nur das Testen von Hypothesen, die auf dem Hintergrund einer Theorie formuliert wurden, kann man als Ursachenforschung bezeichnen. Folglich ergibt sich der Datenbedarf (auch) aus der Art der jeweiligen Theorie.

Nehmen wir als Beispiel die Untersuchung des „generativen Verhaltens“: Generatives Verhalten ist die Gesamtheit aller sozialen Prozesse, die zu einem, für eine bestimmte Gesellschaft typischen Fruchtbarkeitsniveau führen. Dazu gehört also beispielsweise das Heirats- und Scheidungsverhalten, die Sanktionierung von Illegitimität, die Verbreitung von Kinderlosigkeit, die Terminierung der Geburten, die durchschnittliche Kinderzahl etc. Generatives Verhalten ist also nicht nur das Reproduktions„ergebnis“, sondern der ganze soziale Prozeß, der es hervorbringt. Wir sprechen deshalb auch von Reproduktionsmodus. Fragen wir nach der Bedeutung des Heiratsverhaltens für die Reproduktion, so müssen wir bedenken, daß

der Entschluß zu Partnerbindung und Elternschaft kein „punktuell“ Ereignis ist, sondern der Übergang von einer Lebensphase in eine andere. Die frühere Lebensform beeinflusst deshalb ganz erheblich das spätere (eheliche) Lebensstadium. Wer sich z. B. in einer langjährigen vorehelichen Partnerschaft während der Studentenzeit an die Ungebundenheit der Kinderlosigkeit gewöhnt hat, der wird in einer späteren Ehe kaum ein traditionelles Reproduktionsverhalten aufnehmen. Es kommt deshalb darauf an, die *Übergänge* zwischen den einzelnen Stadien im Bindungs- und Reproduktionsverhalten aufzudecken und typische Lebensläufe zu isolieren. Diese theoretische Einsicht in den prozeßhaften Charakter der Reproduktion hat nun aber ganz erhebliche Konsequenzen für die Art der Daten, die für deren Analyse benötigt werden.

Erstens müssen es die Daten erlauben, die einzelnen Phasen des generativen Verhaltens im Zusammenhang zu untersuchen; sie müssen also *Längsschnitte auf Individualniveau* ermöglichen. Dies ist nur möglich mit echten Panel-Daten oder mit retrospektiven Fragestellungen (sog. „Biographien“). Eine Geburtenbiographie z. B. erzählt die Stationen der Reproduktionsgeschichte einer Frau. Werden viele solcher „Geschichten“ (mit Hilfe multivariater Verfahren) zu Typen zusammengefaßt, dann erhält man die für eine bestimmte Gesellschaft und eine bestimmte Generation charakteristischen Ablaufmuster der Reproduktion.

Zweitens sollten weitere wesentliche „Elemente“ im Reproduktionsprozeß, wie das Heirats- und Scheidungsverhalten, oder die Erwerbsbeteiligung der Frau, auf Individualebene mit den Daten der Reproduktionsgeschichte kombinierbar sein. Es hilft wenig, wenn man aus irgendeiner anderen Statistik zum Beispiel die Erwerbsquote von Frauen in einer bestimmten Phase ihres Reproduktionsprozesses kennt. Nicht der statistische Durchschnitt der verschiedenen sozioökonomischen Hintergrundvariablen bestimmt das Ablaufmuster einer Reproduktion, sondern ganz bestimmte Kombinationen von Tatbeständen bringen charakteristische Ablauftypen des generativen Verhaltens hervor. Was man benötigt, sind also *integrierte* Datensätze, nicht solche, die nur aus verschiedenen Statistiken zusammengetragen wurden. Ein hervorragendes Beispiel für eine solcherart integrierte Erhebung ist die jüngste Studie des Berliner Max-Planck-Instituts für Bildungsforschung über „Lebensläufe und gesellschaftlichen Wandel“.

Drittens erfordert unser theoretisches Wissen über das generative Verhalten die Berücksichtigung (eines repräsentativen Querschnitts) *aller* Bundesbürger. Wir wissen heute, daß die Entscheidung für oder gegen Kinder vielfach bereits *vor* der Ehe fällt. In der Bundesrepublik heiratet man heute vielfach nur deshalb, weil man sich für eine Familiengründung bereits entschieden hat. Nimmt man also nur Ehepaare, oder gar (wie häufig geschehen) nur Ehefrauen in die Erhebungsstichprobe, dann wird gerade jene Phase des generativen Prozesses ausgeblendet, um die es geht.

4. Einige Schlußfolgerungen

Amtliche Statistik und Sozialwissenschaft sind aufeinander angewiesen. Ohne eine laufende Bereitstellung fundierter Eckdaten zu sozialen und demographischen Tatbeständen ist unsere Disziplin nicht möglich. Umgekehrt steht eine amtliche Bevölkerungs- und Sozialstatistik ohne die kritischen Analysen der Wissenschaft in der Gefahr, bloße Datenberge anzuhäufen, die wenig zur Erklärung der aktuellen sozio-demographischen Probleme beitragen. Kooperation und gegenseitige Anregung sind also notwendig, um die Qualität von Forschung und statistischer Berichterstattung ständig zu verbessern.

Dringend erforderlich ist dabei vor allem, daß ein vernünftiger Ausgleich zwischen dem Bedarf der Wissenschaft an Einzelfalldaten und der Notwendigkeit des Datenschutzes gefunden wird. Diese Frage läßt sich jedenfalls nicht mit einem bloßen Rückzug auf bestehende Rechtspositionen lösen. Vielmehr müssen gemeinsam praktikable Regelungen gefunden und dem Gesetzgeber vorgeschlagen werden, um den Datenbestand der Ämter modernen wissenschaftlichen Analyseverfahren zugänglich zu machen.

Eine möglichst intensive Nutzung ihrer Datenbestände kann nur im eigenen Interesse der statistischen Ämter sein. Sie legitimiert letztlich den großen finanziellen Aufwand bei den amtlichen Statistiken und die Belastung der Bürger durch die Erhebungen. Die geplante Volkszählung soll die Länder 305,3 Mio., den Bund 30,1 Mio. und die Gemeinden und Kreise ca. 210 Mio. DM kosten. Alles in allem wird der Steuerzahler also gut eine halbe Milliarde für die Volkszählung 1986 aufzubringen haben.

Schließlich sollte auch darüber nachgedacht werden, ob das Erhebungsprogramm wichtiger amtli-

cher Massenerhebungen (wie der Volkszählung oder des Mikrozensus) den heutigen Bedürfnissen der verschiedenen Nutzer noch gerecht wird. Dabei ist auch der Bedarf der universitären Sozialwissenschaft zu beachten und nicht nur die Forderungen von staatlicher Verwaltung und Planung. Man könnte z. B. daran denken, das Standard-Fragenprogramm des Mikrozensus und der Volkszählung

deutlich zu reduzieren und als Ausgleich dafür regelmäßig themenspezifische Sondererhebungen einführen, wie das beispielsweise im österreichischen Mikrozensus geschieht. Solche Zusatzerhebungen sollten in Zusammenarbeit mit der jeweils einschlägigen wissenschaftlichen Disziplin konzipiert werden.