

## Selbstorganisation des Wissenschaftssystems würde Datenschutz vereinfachen und Re-Analysen befördern: gesetzliches Forschungsdaten-Geheimnis könnte die Selbstorganisation unterstützen

Wagner, Gert G.

Veröffentlichungsversion / Published Version  
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:  
GESIS - Leibniz-Institut für Sozialwissenschaften

### Empfohlene Zitierung / Suggested Citation:

Wagner, G. G. (2000). Selbstorganisation des Wissenschaftssystems würde Datenschutz vereinfachen und Re-Analysen befördern: gesetzliches Forschungsdaten-Geheimnis könnte die Selbstorganisation unterstützen. *ZUMA Nachrichten*, 24(47), 75-88. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-208075>

### Nutzungsbedingungen:

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

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

### Terms of use:

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

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

## **SELBSTORGANISATION DES WISSENSCHAFTSSYSTEMS WÜRD E DATENSCHUTZ VEREINFACHEN UND RE-ANALYSEN BEFÖRDERN**

GESETZLICHES FORSCHUNGSDATEN-GEHEIMNIS KÖNNTE DIE  
SELBSTORGANISATION UNTERSTÜTZEN

GERT G. WAGNER

In den Sozial- und Wirtschaftswissenschaften besteht ohne die Möglichkeit einer Re-Analyse von statistischen Ergebnissen (gleichermaßen amtlichen wie nicht-amtlichen) die Gefahr von nicht entdeckten Irrtümern. Mit anderen Worten: Re-Analysen sind in der Wissenschaft die „Berufungsinstanz“, ohne die es keine funktionierende Scientific Community geben kann. Der Schutz vor fehlerhaften wissenschaftlichen Ergebnissen mit Hilfe von Re-Analysen ist ein „öffentliches Interesse“, das in der Datenschutzdiskussion und insbesondere bei der Auslegung von Datenschutzregelungen bislang zu wenig beachtet wurde. Das Wissenschaftssystem und der Gesetzgeber sind gleichermaßen aufgefordert, Re-Analysen zu ermöglichen ohne den Datenschutz zu verletzen. Als Instrumente werden Selbstbindungen im Wissenschaftssystem und die Schaffung eines gesetzlichen „Forschungsdaten-Geheimnisses“ diskutiert.

The re-analysis of statistical data is an effective means of protecting the public from hitherto undiscovered errors in empirical research. In this sense, re-analysis is crucial, for both official data and non-official data. However, discussions about data protection legislation do not usually take this kind of protection into consideration. Proper data protection rules should make it possible to conduct independent re-analysis of protected micro data. The paper discusses the possibility of self-binding in the scientific community towards this goal as well as the possibility of creating new legislation which would assign scientific data a special legal status with regard to data protection (*Forschungsdaten-Geheimnis*).

## 1. Problemaufriss\*

Viele wissenschaftliche Ergebnisse, welche in den Sozial- und Wirtschaftswissenschaften sehr oft statistisch-empirische Ergebnisse sind, haben - meist indirekt, manchmal auch direkt - Einfluss auf private und öffentliche Entscheidungen. Zum Beispiel richten weltweit viele Menschen ihre alltäglichen Ernährungsgewohnheiten an medizinisch-statistischen Ergebnissen aus und Staaten begründen eine Vielzahl von Entscheidungen mit wissenschaftlichen Ergebnissen, so zum Beispiel in Deutschland die Möglichkeit einer Senkung des Rentenniveaus, weil es kaum noch einkommensarme Rentner gibt. In Deutschland sollen sogar erstmals gravierende politische Entscheidungen über die Gestaltung des Arbeitsmarktes für „Niedriglohempfänger“ von der statistischen Auswertung von Experimenten abhängig gemacht werden.<sup>1</sup>

Angesichts der Bedeutung von Wissenschaft für alle Menschen wird in Abschnitt 2 auf Basis von wissenschaftstheoretischen Überlegungen gezeigt, dass neben den Grundrechten „Informationelle Selbstbestimmung“ und „Forschungsfreiheit“, die traditionell im Hinblick auf Datenschutz gegeneinander abgewogen werden, noch ein drittes Rechtsgut berücksichtigt werden sollte: Der Schutz der Öffentlichkeit vor methodisch schlechten und falschen statistischen Ergebnissen.

Nach einer Diskussion des Status quo der Datenschutzpraxis in Deutschland (Abschnitt 3) wird in Abschnitt 4 ein Weg zur Verbesserung des Verhältnisses von Datenschutz und Forschung aufgezeigt, an dessen Anfang Selbstverpflichtungen („berufsständige Regelungen“) durch das Wissenschaftssystem stehen (einschließlich der Schaffung „geschützter Bereiche“) und an dessen Ende ein gesetzliches „Forschungsdaten-Geheimnis“ stehen könnte.<sup>2</sup>

---

\* Der Autor dankt zwei anonymen Gutachtern der „ZUMA-Nachrichten“ für wertvolle Hinweise. Für die hier vertretenen Ansichten und eventuell vorhandenen Fehler können natürlich weder die Gutachter noch ZUMA verantwortlich gemacht werden. Für weitere Literaturhinweise siehe Gert G. Wagner, Wissenschaft schützt die Öffentlichkeit vor schlechten statistischen Ergebnissen, in: Datenschutz und Datensicherheit (DuD), 23. Jg., Heft 7, 1999.

<sup>1</sup> Vgl die Pressemitteilung des Bundesministeriums für Arbeit und Sozialordnung „Kabinett beschließt ‚Sonderprogramm zur Erprobung von Modellansätzen zur Förderung der Beschäftigung von Geringqualifizierten und Langzeitarbeitslosen‘“, vom 28. Juni 2000.

<sup>2</sup> Nicht zuletzt auch aufgrund der provozierenden DFG-Denkschrift (1996) sind Wissenschaftler und Datenschützer in den letzten Jahren besser miteinander ins Gespräch gekommen als dies in den achtziger Jahren der Fall war, in denen konstruktive Vorschläge, die vorlagen, nicht konsequent diskutiert wurden. Diese Gespräche haben sich in jüngster Zeit als sehr fruchtbar

Das Forschungsdaten-Geheimnis und berufsständische Detailregelungen könnten - wenn das Wissenschaftssystem die Politik davon überzeugt - unschwer bei der durch die EU-Datenschutzrichtlinie notwendig gewordenen Novellierung des Bundesdatenschutzgesetzes (BDSG) berücksichtigt werden. Freilich muss das Wissenschaftssystem sich entsprechend organisieren und die Politik von der Notwendigkeit eines Forschungsdaten-Geheimnisses überzeugen, wenn wesentliche Fortschritte für den Datenschutz im Bereich der Wissenschaft erreicht werden sollen.

## 2. Bedeutung von Re-Analysen für die Wissenschaft

Moderne Wissenschaft („kritischer Rationalismus“) definiert sich über das Gebot der „intersubjektiven Nachprüfbarkeit“, d. h. im Hinblick auf empirische Forschung durch Re-Analysen von Ergebnissen. Diese sind z. B. bei Experimentaldaten oder statistischen Daten wichtig im Hinblick auf

- Methodenabhängigkeit von Daten und Ergebnissen,
- Fehler bei der Datengewinnung und -auswertung sowie
- Betrug.

Bei statistischen Daten und Auswertungen spielt die Methodenabhängigkeit der Ergebnisse eine große Rolle. Statistische Daten werden von den gewählten Erhebungsmethoden beeinflusst (vgl. zum „Adäquationsproblem“ z. B. Grohmann 1985) und statistische Analysen von den benutzten Auswertungsmethoden (vgl. z. B. Fitzenberger/Speckesser 2000).

Fehler bei Auswertungen sind wahrscheinlich (vgl. z. B. Williamson/Jones 1983)<sup>3</sup> und selbst Betrug kann man nicht ausschließen, da auch Wissenschaftler und Statistiker eigennützige Ziele verfolgen (vgl. dazu Finetti/Himmelrath 1999).

Die Gesellschaft kann sich vor einseitigen Ergebnissen (aufgrund einer bestimmten Methodenwahl bei der Datengenerierung und/oder -auswertung), Irrtümern und Fälschungen bei statistischen Auswertungen nur schützen, wenn die benutzten Daten, d. h. die Tabellen und anderen Ergebnissen zugrundeliegenden Mikro-Daten, jedem Wissenschaftler für Re-Analysen zur Verfügung stehen. Ohne derartige Re-

---

erwiesen und sollten intensiv fortgesetzt werden. Vgl. insbesondere Hamm/Möller (1999) und die dort dokumentierten Diskussionsbeiträge sowie die Aufsätze zum Thema „Datenschutz und Forschung“ im 23. Jg, Heft 7, 1999 der Zeitschrift „Datenschutz und Datensicherheit (DuD)“.

<sup>3</sup> Beispielsweise die Zeitschrift „Journal of Political Economy“ hat deswegen eine eigene Sektion für Re-Analysen eingerichtet.

Analysen gibt es keine Chance für den Gesetzgeber, der ohnehin in einer schlechteren Informationslage ist als Wissenschaftler, die Gesellschaft vor Fälschungen zu schützen.

Theoretisch optimal wäre, wenn ein Re-Analyst nicht nur die erhobenen Daten bekommen würde, sondern auch die Möglichkeit hätte zu prüfen, wie die Datengenerierung erfolgte, also im Falle der Sozial- und Wirtschaftswissenschaften, ob Angaben in einer Erhebung bzw. in einer prozessproduzierten Datei auch dem entsprechen, was Auskunftgeber angegeben haben.

### **3. Status Quo in Deutschland**

#### **3.1 Abgrenzung und Begriffe**

Dieser Aufsatz beschäftigt sich nur mit einer Dimension des Datenschutzes in der Wissenschaft, nämlich der Anonymität von Erhebungseinheiten. Auf eine ebenso wichtige zweite Dimension, nämlich auf die Einwilligung von Personen zur Erhebung und Verarbeitung ihrer Daten, wird nicht näher eingegangen.<sup>4</sup>

Daten, die die hinter einem Datensatz stehende Persönlichkeit offenbaren, gelten als „personenbezogen“ und können nicht nach Belieben erhoben, gespeichert und analysiert werden, da sie der „informationellen Selbstbestimmung“ unterliegen (vgl. z. B. Bizer 1992: 25ff.). Generierung und Verarbeitung solcher Daten bedürfen der expliziten Zustimmung eines jeden Betroffenen. Werden solche Daten „anonymisiert“, ist ihre Auswertung unbedenklich.

Faktisch werden in Deutschland auch Firmendaten wie personenbezogene Daten behandelt. Dahinter steht das im Grundsatz berechnete geschäftliche Interesse, der Konkurrenz keine Daten zu offenbaren (obwohl dem zum Teil Publizitätspflichten von Unternehmen entgegenstehen).

---

<sup>4</sup> Der Autor hält es für eine ethisch begründete Selbstverständlichkeit, dass im Grundsatz eine explizite Einwilligung notwendig ist. Ausnahmen von der Regel kann es nur in wenigen - explizit und permanent zu begründenden - Bereichen medizinischer Forschung geben. Besonders schwierig wird das „informationelle Selbstbestimmungsrecht“, wenn aus - für andere Zwecke - prozessproduzierten Daten (die z. B. dem Sozialrecht unterliegen) für statistische bzw. wissenschaftliche Zwecke Daten generiert werden sollen. Zum einen stellt sich die Frage, ob dies grundsätzlich erlaubt ist und zum anderen die Frage, wer diese Daten auswerten darf, die ja nicht für statistische Zwecke und schon gar nicht für unabhängige wissenschaftliche Untersuchungen erhoben wurden.

Das Statistik-Geheimnis der Amtlichen Statistik ist älter als der moderne Datenschutz. Das Statistik-Geheimnis garantiert, dass Einzel-Daten Dritten nicht bekannt werden. Es bezieht sich - aus pragmatischen Gründen - auf personen- wie firmenbezogene Daten, da es die Auskunftsbereitschaft sicherstellen soll.

Es ist unstrittig, dass absolut anonymisierte Daten, für die es hundertprozentig ausgeschlossen werden kann, dass durch sie auf einzelne Einheiten geschlossen werden kann, nicht der Datenschutzgesetzgebung unterliegen und ohne weitere Schutzvorkehrungen ausgewertet werden dürfen.

Das BSDG (§ 3) unterscheidet nur in personenbezogene und (vollständig) anonymisierte Daten. Das Bundesstatistikgesetz (BStatG) differenziert - zu Recht - die „anonymisierten Daten“ in „absolut anonymisierte“ Daten und „faktisch anonymisierte“ Daten, die nur unter „unverhältnismäßig großen Aufwand“ de-anonymisiert werden können (vgl. Bizer 1992: 398f.). In der Terminologie der Statistik werden absolut anonymisierte Daten als „public use files“ und - mehr oder weniger stark - faktisch anonymisierte Daten als „scientific use files“ bezeichnet. Darauf wird unten noch näher eingegangen werden

### 3.2 Realität von Wissenschaft und Datenschutz

Ob ein Datensatz aus juristischer Sicht als „faktisch anonymisiert“ gilt, hängt nach § 16 Abs. 6 BStatG formal davon ab, ob „die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmbaren natürlichen Person zugeordnet werden können“.

In der Praxis kann es absolut anonymisierte Daten nicht geben. Solange auch nur ein Bit an Information in einem Datensatz enthalten ist, besteht auch eine Chance der De-Anonymisierung der Erhebungseinheit (und sei es nur per Zufall). Der Übergang zwischen faktisch anonymisierten und personenbezogenen Daten ist in der Praxis fließend, da eine bloße formale Anonymisierung (durch Weglassen von Name, Adresse) einen Datensatz im Allgemeinen noch nicht „faktisch anonymisiert“, da es zumindest für bestimmte Fälle nicht schwer ist, aufgrund der erhobenen Daten eine Erhebungseinheit zu identifizieren (z.B. Großunternehmen in einer bestimmten Region).<sup>5</sup>

---

<sup>5</sup> In den Sozial- und Wirtschaftswissenschaften schützen sich die bevölkerungsrepräsentativen Surveys gewissermaßen durch ihre Repräsentativität wirkungsvoll gegen systematischen Datenmissbrauch, denn diese Daten, die nach keinerlei Merkmalen sortiert sind, sind für Mar-

Es leuchtet unmittelbar ein, dass durch die Verknüpfung unterschiedlicher Datensätze, die den Informationsgehalt steigern, die faktische Anonymität geschwächt wird. Insbesondere in der Arbeitsökonomie und der Epidemiologie gewinnen derartige verknüpfte Daten eine immer größere Bedeutung (vgl. z.B. Westergard-Nielsen 1999).

In der Wissenschaft bezieht sich Datenschutz also nicht nur auf personenbezogene Daten, sondern es geht darum, ob und insbesondere unter welchen Bedingungen Daten als „faktisch anonymisiert“ klassifiziert werden. Je nach Position, in der sich ein Nutzer von Daten befindet, ist „unverhältnismäßig großer Aufwand“ offenbar unterschiedlich zu beurteilen; insbesondere dann, wenn man nicht nur direkte Kosten, sondern auch Opportunitätskosten betrachtet (d. h. mögliche Folgen der Tat). Für einen jugendlichen Computerhacker, der faktisch keine Sanktionen befürchten muss, ist ein Datensatz faktisch einfacher de-anonymisierbar als beispielsweise für einen Wissenschaftler, der hohe (auch soziale) Kosten hat, weil die Gefahr besteht, dass er seine Reputation und damit seine Karriere zerstört, wenn er bewusst Erhebungseinheiten de-anonymisiert.<sup>6</sup> Dazwischen liegen Gruppen wie Mitarbeiterinnen in Marketingabteilungen und Wahlkampfhelfer. Auch bei ihnen hängen negative Folgen nicht nur vom Gesetz ab, sondern von der „Professions-Kultur“ innerhalb derer sie agieren. Wie diese Kulturen außerhalb der Wissenschaft aussehen, ist z. Z. nicht bekannt.<sup>7</sup>

In der alltäglichen Auslegung der „faktischen Anonymität“ achten deswegen viele Datenproduzenten, die faktisch anonymisierte Daten besitzen, dass die Daten nur von Personen und Institutionen genutzt werden, die als Wissenschaftler bzw. wis-

---

ketingzwecke völlig uninteressant, weil die Befragten über das gesamte Gebiet eines Landes verstreut und vollkommen inhomogen sind.

**6** Hier soll gar nicht behauptet werden, dass Wissenschaftler ethisch besonders hochstehende Persönlichkeiten sind, die aufgrund einer „Sonderethik“ der Wissenschaft Datenschutz *eo ipso* garantieren. Im Gegenteil: Wissenschaftler seien auch für Betrug anfällig, wie die zunehmende Zahl von Fälschungsskandalen in den Naturwissenschaften und insbesondere den Biowissenschaften zeigt (vgl. Finetti/Himmelrath 1999). Aber in Bezug auf die Anonymität von Erhebungseinheiten kann man sich trotz der Anfechtungen, denen Wissenschaftler ebenso wie andere Menschen ausgesetzt sind, auf das Eigeninteresse von Wissenschaftlern zur Einhaltung des Datenschutzes verlassen, denn Wissenschaftler haben keine Vorteile davon, wenn sie Datenschutz nicht streng handhaben. Wissenschaftler haben kein Interesse an der De-Anonymisierung von Einzelfällen, da die De-Anonymisierung nicht publizierbar ist. Wissenschaftler würden durch ein derartiges illegales Handeln ihre Karriere gefährden, ohne dass dem Risiko ein entsprechender Nutzen gegenübersteht.

**7** Freilich gibt es Hinweise auf Missbrauch des „Statistik-Geheimnisses“ durch Firmen, die Direkt-Marketing betreiben (DFG 1999).

senschaftliche Einrichtungen viel zu verlieren haben, wenn sie den Datenschutz nicht ernst nehmen würden. So verfahren z.B. das Statistische Bundesamt (vgl. z. B. Bizer 1992: 398ff.) und das Deutsche Institut für Wirtschaftsforschung (Seufert/Wagner 1998). Dies ist ein Verfahren, das keine Behinderung der wissenschaftlichen Forschung darstellt.

Freilich werden Re-Analysen statistischer Untersuchungen - nahezu gleichermaßen bei amtlichen wie nicht-amtlichen Daten - durch die „Zweckbindung“, die das BDSG für schützenswerte Daten verlangt, behindert, weil im Grundsatz (personenbezogene) Daten nicht „auf Vorrat“ gespeichert werden dürfen. Auch faktisch anonymisierte amtliche Daten dürfen nur von kleinen Arbeitsgruppen zu einem eindeutig bestimmten Forschungszweck genutzt werden und müssen danach vernichtet werden (§ 16, Abs. 8 BStatG). Dies widerspricht jedoch einem Grundprinzip empirischer Forschung: Wissenschaftliche Ergebnisse müssen der Überprüfung (Verifizierung, Falsifizierung) durch andere Forscher zur Verfügung stehen. So akzeptieren führende internationale wissenschaftliche Zeitschriften Beiträge nur dann, wenn die verwendeten Daten anderen Forschern zugänglich sind.

Viele Analysen basieren auf „Sozialdaten“, die nur für Zwecke der Sozial-Verwaltung benutzt werden dürfen. Ihre Auswertung ist z. B. möglich, wenn ein Forschungsprojekt als im Interesse eines Sozialversicherungsträgers definiert wird. Dies ist unbefriedigend, da die statistische Auswertung von Daten durch die Sozial-Behörden ja kein reiner Verwaltungsakt ist, sondern auf wissenschaftlichen Grundlagen beruht und deswegen - im Interesse der Sozialversicherten und der Öffentlichkeit - wissenschaftlich nachprüfbar sein muss, was nur durch unabhängige Re-Analysen möglich ist. Wird dieses öffentliche Interesse nicht grundsätzlich unterstellt, d. h. jede Re-Analyse, die gewünscht wird, von den entsprechenden Sozialbehörden auch gestattet, dann besteht die Gefahr selektiver Forschung.

Auf der anderen Seite ist es nahezu skurril, dass die Deutsche Forschungsgemeinschaft (DFG) bislang keine eindeutigen Regeln entwickelt hat, die den Zugang zu mit Hilfe von DFG-Geldern erzeugten Daten für Re-Analysen regeln. Die DFG-Bewilligungsbescheide enthalten dazu standardmäßig keinerlei Auflagen (und es hängt von einzelnen Gutachtern ab, ob Förderungsempfängern die Datenweitergabe auferlegt wird).<sup>8</sup>

---

<sup>8</sup> Re-Analysen bei kommerzieller und nicht-veröffentlichter Forschung sind für die Öffentlichkeit kein Problem. Wird kommerzielle Forschung publiziert, um dadurch die Öffentlichkeit zu informieren (bzw. zu beeinflussen), sollte der Gesetzgeber ein Recht auf Re-Analysen erzwingen..



Die bestehende Datenschutzpraxis erlaubt durchaus wissenschaftsfreundliche Lösungen. Zum Beispiel wurde für das Zusammenführen von personenbezogenen Mikrodaten in Deutschland ein Treuhänder-Modell entwickelt, bei dem ein Notar Daten zusammenführt und schließlich nur ein faktisch anonymisierter Datensatz für die statistische Analyse genutzt wird (vgl. Metschke 1999). Eine andere Möglichkeit besteht in der Einrichtung von Forschungs-Datenzentren, in denen – in einem staatlich geschützten und kontrollierten Bereich – Analysen mit personenbezogenen bzw. nur schwach anonymisierten Daten möglich wären. Die „Schalterstelle“ des IAB ist im Prinzip ein solches Zentrum (vgl. Kölling 2000: 296f.).

Das Nachprüfen der Datengenerierung ist datenschutzrechtlich möglich, wenn Befragte auf die Möglichkeit von Nachfragen aufmerksam gemacht würden und sie dieser Möglichkeit zustimmen (dann wäre die Untersuchung mit der Ersterhebung noch nicht abgeschlossen und die Adressen dürften gespeichert werden).<sup>9</sup>

## 4. Verbesserungsmöglichkeiten in der Praxis

### 4.1 Selbstorganisation der Wissenschaft

Hinsichtlich der Unzufriedenheit des Wissenschaftssystems mit der Praxis des Datenschutzes sind zwei Grundsatzanmerkungen notwendig:

- In der Vergangenheit haben die Wissenschaftsorganisationen sicherlich oft versäumt, rechtzeitig Einfluss auf Gesetzgebungsverfahren zu nehmen, so wie dies durch Lobbyisten in Parlamentarischen Demokratien erwartet wird (vgl. Wagner 1999b).
- Insbesondere die Sozial- und Wirtschaftswissenschaften haben bislang keine ernstzunehmenden Selbstbindungsregeln zur Einhaltung des Datenschutzes entwickelt, wie sie z. B. bei den Medizinern und der Amtlichen Statistik gegeben sind.

Da das Wissenschaftssystem die gesetzliche Gestaltung der Datenschutzgesetzgebung nicht systematisch begleitet, ist es auch nicht verwunderlich, dass vom Gesetz-

---

<sup>9</sup> Dies wird allerdings allein aus praktischen-finanziellen Gründen eine Ausnahme bleiben. In der alltäglichen Praxis ist es bereits hilfreich, wenn neben den erhobenen Daten selbst auch „Feldinformationen“ dem auswertbaren Datensatz auf Mikroebene beigegeben werden; z. B. die Zahl der Kontaktversuche, Merkmale von Interviewern usw. Editierte bzw. „imputierte“ Informationen (die also nicht von Befragten stammen, sondern mit Hilfe von Hypothesen „erzeugt“ werden), sollten als solche kenntlich sein. Dann ist bereits ein hohes Maß an Nachprüfbarkeit der Erhebung gegeben.

geber bei den Anhörungen für neue gesetzliche Grundlagen, die relevanten wissenschaftliche Organisationen nicht systematisch einbezogen werden. Dies sollte geändert werden: wenn man einerseits den Datenschutz im Detail gesetzlich regelt, sollte es auch minutiöse Anhörungsvorschriften für die entsprechenden Gesetzgebungsverfahren geben. Neben der Gesetzgebungsebene könnten in der alltäglichen Verwaltungspraxis „gemeinsame“ Entscheidungsgremien, in denen z.B. „Ethikkommissionen“ und Datenschützer zusammenarbeiten, nützlich sein.

Die „Privilegierung“ der Amtlichen Statistik beim Umgang mit sensiblen Daten beruht u.a. auf einem „statistikspezifischen“ Berufsethos, das durch den Beamtenstatus, der besonders wirksame Sanktionen erlaubt, gestärkt wird. Deswegen ist die Frage sinnvoll, ob ein solches System nicht auch für die Wissenschaft im allgemeinen möglich wäre.

Ein weltweit akzeptierter, bindender, fachspezifischer „Code of Ethics“ bzw. „Code of Good Practice“ der die Spezialbelange einzelner Disziplinen berücksichtigt, wäre im Hinblick auf die Einschätzung der Gefahr des Missbrauchs personenbezogener bzw. faktisch anonymisierter Daten wahrscheinlich sehr hilfreich. Vgl. z. B. den Ethik-Kodex der Deutschen Gesellschaft für Soziologie und des Berufsverbandes Deutscher Soziologen (1992: Abschnitte I.B.7 und 8). Dieser Kodex ist freilich zu wenig spezifisch. Außerdem leidet es darunter, dass es nicht in einen gesetzlichen Rahmen eingebunden ist, der hart zu sanktionieren erlaubt. Kaase et al. (1980: 293) ist nach wie vor zuzustimmen, dass berufsständische Regeln gesetzliche Regelungen nicht ersetzen können. Sie können freilich gesetzliche Regelungen im Spannungsfeld zwischen Wissenschaftsfreiheit und Datenschutz praktikabel machen und sind deswegen unverzichtbar.

Zu einem wirksamen *Code of Ethics* müsste auch die Selbstverpflichtung der Forschungsförderungseinrichtungen gehören, dass Zuwendungsempfänger verpflichtet werden, ihre Daten nach angemessener Zeit für Re-Analysen freizugeben. Die fachspezifische Definition der „Angemessenheit“ der Exklusiv-Nutzung muss Bestandteil dieser Selbstverpflichtung sein.

#### **4.2 Ein „Forschungsdaten-Geheimnis“ als Königsweg?**

Zur Förderung und Stärkung von Selbstverpflichtungen des Wissenschaftssystems sollte geprüft werden, ob ein „Forschungsdaten-Geheimnis“ etabliert werden

könnte.<sup>10</sup> Zum einen sollte es sicherstellen, dass Wissenschaftlern ein besonderes Vertrauen beim Verarbeiten von sensitiven (d.h. faktisch anonymisierten und personenbezogenen) Daten entgegengebracht werden kann. Zum zweiten sollte es Daten, die Forschungszwecke erhoben bzw. aufbereitet wurden, davor schützen, dass der Staat zum Zwecke von Ermittlungen (z.B. in Straf- und Steuerfällen) darauf (personenbezogen) zugreift.<sup>11</sup>

Dies könnte auch bedeuten, dass die datenschutzrechtliche Trennung von Forschung in „Amtliche Statistik“ einerseits und „freie wissenschaftliche Forschung“ andererseits aufgehoben würde. Dies würde gleichermaßen eine Stärkung der Wissenschaft wie der Amtlichen Statistik bedeuten. Die Wissenschaft würde als ein besonderer Berufsstand anerkannt und der wissenschaftliche Charakter der Amtlichen Statistik würde sichtbar werden.

Ein „Forschungsdaten-Geheimnis“ könnte dadurch operationalisiert werden, dass man mit einem Diplom auch eine „Zulassung“ für die Verarbeitung sensitiver Daten (im Sinne eines novellierten BDSG) erwirbt. Verbunden werden müsste dies mit der Möglichkeit, dass einem zugelassenen Wissenschaftler auch zugetraut wird, das Personal, das für ihn arbeitet, ebenso zu verpflichten, wie man dies Ärzten und Rechtsanwälten seit langem zubilligt (vgl. § 203 Abs. 3 Strafgesetzbuch).

Wer sich nicht an die Regeln des Forschungsdaten-Geheimnisses hält, der verliert seine „Lizenz zum Analysieren“ und wird darüber hinaus bestraft. Derartige Lizenzen könnten im Rahmen einer berufsständischen Selbstverwaltung vergeben werden.

Ein zentraler Baustein des Forschungsdaten-Geheimnisses wäre eine Pflicht zum Erlernen von Datenschutzmaßnahmen und -möglichkeiten im Studium. D. h. also, dass Datenschutz in die Ausbildung aller einschlägigen wissenschaftlichen Disziplinen aufzunehmen wäre bzw. in die Statistikausbildung, die zum Standard einiger Disziplinen gehört. Datenschützer sollten an der praktischen Lehre mitwirken.

---

<sup>10</sup> Vgl. auch Bizer 1992: 229ff.) sowie die Punkte 26 und 27 des Eckwerte-Papiers der SPD-Bundestagsfraktion „Modernes Datenschutzrecht für die (globale) Wissens- und Informationsgesellschaft“. Hektographiertes Manuskript, Bonn 1999. Das Bundes-Archivgesetz könnte in Teilbereichen der „Speicherung auf Vorrat“ als Vorbild dienen.

<sup>11</sup> Dies wäre also ein Beschlagnahmungsverbot und ein Zeugnisverweigerungsrecht gegenüber der Polizei, Justiz und den Nachrichtendiensten sowie den Behörden, die die prozessproduzierten Daten, die sie für Forschungszwecke anonymisiert haben, „zurückholen“ wollen, um Forschungsergebnisse, z. B. über Missbrauch, für ihre Verwaltungszwecke nutzen zu können.

Ein Forschungsdaten-Geheimnis führt unvermeidbar zu der Frage, wie wissenschaftliche Forschung definiert wird. Grundsätzlich kann man diese Frage mit Hilfe von zwei Methoden beantworten: Zum einen „qualifikationsbezogen“, d.h. Forschung ist das, was von einem diplomierten Forscher durchgeführt wird; zum zweiten „institutionell“, d.h. Forschung ist das, was in unabhängigen Forschungseinrichtungen stattfindet (d.h. in Einrichtungen, die qua Gesetz der wissenschaftlichen Forschung dienen).

Definiert man Forschung institutionell, besteht die Gefahr, dass Datenschutz als Kontroll- und Steuerungsinstrument für staatlich gewünschte Forschung missbraucht wird; definiert man Forschung qualifikationsbezogen, besteht die Gefahr, dass die grundgesetzlich verbrieft „Freiheit der Forschung“ verletzt wird, weil nur noch lizenzierte Forscher, nicht aber Jedermann forschen kann. Im Grundgesetz heißt es aber: „Kunst und Wissenschaft, Forschung und Lehre sind frei“ (Art. 5, Absatz 3 GG).

Bei der Auslegung des Gebotes der Wissenschaftsfreiheit wird freilich - so z. B. vom Bundesverfassungsgericht - auf einen „ernsthaften planmäßigen Versuch zur Ermittlung von Wahrheit“ Bezug genommen (nach Bizer 1992: 45). D. h., dass nicht jeder Dilettant sich als Wissenschaftler definieren kann, sondern dass die „Ernsthaftigkeit und Planmäßigkeit“ von Dritten geprüft werden können. Insofern ist die „Wissenschaftsfreiheit“ bereits heutzutage eingeschränkt. Und diejenigen, die mit neuen Methoden arbeiten, wären nicht grundsätzlich ausgeschlossen, sie müssten lediglich ein Diplom besitzen, das auf Basis der Standard-Methoden erworben wurde.

Eine optimale Lösung ist offensichtlich nicht einfach zu finden. In einem demokratischen Rechtsstaat sollten Lösungen vermieden werden, die auf Kontrolle und Zensur hinauslaufen können. Definiert man wissenschaftliche Forschung qualifikationsbezogen, so kann die Freiheit der Forschung am ehesten sicher gestellt bleiben.

## 5. Zusammenfassung und Ausblick

Bei der datenschutzrechtlichen Beurteilung von Forschungsvorhaben werden bislang nur die Grundrechte auf „Forschungsfreiheit“ und „informationelle Selbstbestimmung“ gegeneinander abgewogen. In diesem Aufsatz wird argumentiert, dass auch der „Schutz der Öffentlichkeit von Fälschungen und fehlerhaften Ergebnissen“ mit in die Abwägung einbezogen werden sollte.

Es ist wissenschaftstheoretisch eindeutig, dass sich die Güte empirischer Ergebnisse, die auf (schutzwürdigen) Daten beruhen, nur mit Hilfe von Re-Analysen ermitteln

lässt. Deswegen ist die „Weitergabe“ von Forschungsdaten von Datenproduzenten an andere Wissenschaftler ein wissenschaftlich zentrales Gebot. Dies gilt auch für Daten der amtlichen Statistik, da diese auf wissenschaftlichen Methoden beruht (sowie für kommerzielle Forschung, insoweit sie veröffentlicht wird, um die Öffentlichkeit zu informieren bzw. zu beeinflussen).

Wissenschaftler (einschließlich Amtlicher Statistiker) und Datenschützer sollten deswegen allgemein gültige und klare Regelungen für die Verarbeitung und Re-Analyse personenbezogener und anonymisierter Daten erarbeiten. Dabei sollte insbesondere anerkannt werden, dass es absolut anonymisierte personen- bzw. unternehmensbezogene Daten in der Praxis nicht gibt, d.h., die klare Trennung in personenbezogene und nicht-personenbezogene Daten, die das BDSG kennt, ist unrealistisch. Das Wissenschaftssystem sollte - in Anlehnung an bewährte Regelungen innerhalb der amtlichen Statistik - einen verbindlichen Verhaltenscodex entwickeln, der auch obligatorisch im Studium gelehrt und geprüft werden müsste. Dann könnte auch die datenschutzrechtliche Trennung in „Amtliche Statistik“ und „freie Forschung“ aufgegeben werden.

Als - ein mit einer Selbstbindung der Wissenschaft beginnender - Königsweg für die Sicherstellung der Anonymität von Forschungsdaten wird ein Forschungsdaten-Geheimnis zur Diskussion gestellt,<sup>12</sup> das keinen datenschutzrechtlichen Freibrief für Forscher darstellen würde, sondern hohe Anforderungen an die Ethik und die Praxis der Wissenschaften festschreiben würde.

Das Forschungsdaten-Geheimnis und berufsständische Selbstbindungen könnten - wenn das Wissenschaftssystem die Politik davon überzeugt - unschwer bei der zweiten Stufe der durch die EU-Datenschutzrichtlinie notwendig gewordenen Novellierung des BDSG berücksichtigt werden.

### **Korrespondenzadresse**

*Univ.-Prof. Dr. Gert G. Wagner  
DIW Berlin  
Königin Luise Straße 5  
14191 Berlin  
E-Mail: gwagner@diw.de*

---

<sup>12</sup> Daneben sind Treuhänder-Modelle zur Zusammenführung sensibler Daten und besonders geschützte Forschungsdaten-Zentren zur Auswertung sensibler Daten sinnvolle Seitenpfade.

## Literatur

Bizer, J., 1992: Forschungsfreiheit und Informationelle Selbstbestimmung, Baden-Baden.

Deutsche Forschungsgemeinschaft (DFG) 1996: Forschungsfreiheit - Ein Plädoyer für bessere Rahmenbedingungen der Forschung in Deutschland. Weinheim: VCH.

Deutsche Forschungsgemeinschaft (DFG) 1999: Qualitätskriterien der Umfrageforschung, Weinheim: VCH.

Deutsche Gesellschaft für Soziologie und Berufsverband Deutscher Soziologen 1992: Ethik-Kodex (abgedruckt in: Soziologie (DGS Nachrichten), Heft 4, 1998, S. 79-85).

Finetti, M./Himmelrath, A., 1999: Der Sündenfall - Betrug und Fälschung in der deutschen Wissenschaft. Stuttgart: dtv.

Fitzenberger, B./Speckesser, St., 2000: Zur wissenschaftlichen Evaluation der Aktiven Arbeitsmarktpolitik in Deutschland - Ein Überblick. ZEW Discussion Paper Nr. 00-06, Mannheim: mimeo.

Grohmann, H., 1985: Vom theoretischen Konstrukt bis zum statistischen Begriff – Das Adäquationsproblem. Allgemeines Statistisches Archiv, Band 73: 1-15.

Hamm, R./Möller, K.-P., (Hg.) 1999: Datenschutz und Forschung. Baden-Baden: Nomos.

Kaase, M./Krupp, H.-J./Pflanz, M./Scheuch, E. K./Spiros, S., 1980: Datenzugang und Datenschutz - Konsequenzen für die Forschung. Königstein/Taunus: Athenäum.

Kölling, A., 2000: The IAB-Establishment Panel. Schmollers Jahrbuch, Band 120: 291-300.

Metschke, R., 1994: Datenschutz in Wissenschaft und Forschung. Broschüre des Berliner Datenschutzbeauftragten, Berlin: mimeo.

Metschke, R., 1999: Erfahrungen mit der Verarbeitung sensitiver Daten durch Treuhänder. S. 127-132 in: Statistisches Bundesamt (Hg.), Kooperation zwischen Wissenschaft und Amtlicher Statistik. Wiesbaden: Metzler Poeschel.

Schnell, R., 1997: Non-Response in Bevölkerungsumfragen - Ausmaß, Entwicklung und Ursachen. Opladen: Leske+Budrich.

Seufert, W./Wagner, G. G., 1998: Problems of Dissemination and Examples of „Public Use Micro-Data“ in Germany. S. 58-64 in: American Statistical Association 1997 Proceedings for the Government and Social Statistics Sections.

Simitis, S., 1999: Diskussionsbeitrag. S. 29-31 in: R. Hamm/Möller, K.-P. (Hg.) 1999, *Datenschutz und Forschung*. Baden-Baden: Nomos.

Vetter, R., 1999: Datenschutz und Forschungsfreiheit - Widerspruch oder Weg zur wechselseitigen Grundrechtsrealisierung. S. 21-41 in: R. Hamm/Möller, K.-P. (Hg.) 1999, *Datenschutz und Forschung*. Baden-Baden: Nomos.

Wagner, G. G., 1999a: Ziele und Unabhängigkeit der Wissenschaft sind Instrumente eines effektiven Datenschutzes. S. 14-20 in: R. Hamm/Möller, K.-P. (Hg.) 1999, *Datenschutz und Forschung*. Baden-Baden: Nomos.

Wagner, G. G., 1999b: Self-regulation Within the Research Community as a Means for Shaping the Statistical Infrastructure. S. 59-61 in: Statistisches Bundesamt (Hg.), *Kooperation zwischen Wissenschaft und Amtlicher Statistik*. Wiesbaden: Metzler Poeschel.

Westergard-Nielsen, N., 1999: Linking employer-employee data – the Danish experience. S. 124-126, in: Statistisches Bundesamt (Hg.), *Kooperation zwischen Wissenschaft und Amtlicher Statistik*. Wiesbaden: Metzler Poeschel.

Williamson, S./Jones, W., 1983: Computing the Impact of Social Security Using the Life Cycle Consumption Function. *American Economic Review*, Band 73: 1036-1052.