Dieses Papier enthält Vorschläge und Anregungen für die Diskussion. Fühlen Sie sich ermutigt, weitere Gesichtspunkte und Fragen einzubringen!

**Inhalt**

# 1 Grundlagen für alle Themen

Grundsätzlich spechen wir auf diesem Workshop über Personendaten. Für Betriebsdaten gelten im Prinzip die gleichen Überlegungen, jedoch ist das Reidentifikationsrisiko bei Betriebsdaten i.d.R. höher als bei Personen (wg. größerem Stichprobenauswahlsatz, mehr nutzbarem Zusatzwissen, evtl. auch höherer Attraktivität für Datenangreifer).

## *1.1 Was ist faktische Anonymität?*

- Reidentifizierung ist gar nicht oder nur mit großem Aufwand an Zeit, Kosten und Arbeitskraft möglich.

  Z.B. SGB X, § 67: "(8) Anonymisieren ist das Verändern von Sozialdaten derart, dass 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."

  Die meisten Autoren sind sich darin einig, dass sich "unverhältnismäßig" an der Relation von Aufwand und Nutzen aus der Reidentifizierung bemisst. → Deanonymisierungsmotiv des potentiellen Datenangreifers

  Das Motiv des wissenschaftlichen Nachweises, dass Deanonymisierung prinzipiell möglich ist, darf hierbei nicht mit einbezogen werden, denn:

  "Die Einbeziehung des Motivs des Nachweises der prinzipiellen Machbarkeit der Deanonymisierung würde die vom Gesetzgeber mit dem § 16 Abs.6 BStatG verfolgte Intention geradezu torpedieren. Mit der Regelung der faktischen Anonymität hat der Gesetzgeber ja implizit anerkannt, dass unter bestimmten Voraussetzungen Deanonymisierungen möglich sind und er bereit ist, solche Fälle als verbleibendes Risiko hinzunehmen, wenn sie nur unter Einsatz unverhältnismäßiger Mittel zustande kommen können." (Müller et al. 1991: 157)[1]

  Das gilt in gleichem Maße für Sozialdaten, denn § 67 SGB X Abs (8) übernimmt die Formulierung des BStatG.

- Alleine das spontane Identifizieren von Fällen mit extremen Werten durch den Forscher während der Arbeit mit den Daten ist kein Bruch des Datenschutzes, denn das Wissen wurde nicht einer unautorisierten Person geoffenbart, und es handelt sich nicht um einen gezielten Deanonymisierungsversuch, so Hafner / Ritchie / Lenz (2015)[2]:

  "Note that it is not a risk that a researcher spontaneously notes the characteristics of an observation and muses on the company identity but does not follow up - there has been no disclosure to an unauthorised person, and no deliberate attempt to identify a company." (Hafner / Ritchie / Lenz 2015: 7)

---

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

2 Hafner / Ritchie / Lenz (2015): User-focused threat identification for anonymised microdata, UWE Bristol, Economics Working Paper Series 1503, http://www2.uwe.ac.uk/faculties/BBS/BUS/Research/Economics%20Papers%202015/1503.pdf (Zugriff: 05.11.2015)

- Frage: muss faktische Anonymität verhindern, dass sich ein Befragter in den Daten selbst identifizieren kann?

## *1.2 Bedingungen für erfolgreiche Reidentifikationsversuche*

- Auszug aus Müller et al. 1991: 41-45
  → Müller-et-al_Meth-Grundl-Reidentifizierung.pdf:
    - Zusatzwissen ist erforderlich: Personenbezogener Identifikationsfile; im Extremfall nicht codiertes Wissen über einen Einzelfall
    - Der Identifikationsfile muss sich zumindest in Teilen auf die gleichen Personen beziehen wie der Mikrodatenfile
    - Deanonymisierungsversuche können nur an den Merkmalen ansetzen, die Mikrodatenfile und Identifikationsfile gemeinsam haben (Überschneidungsmerkmale).
    - "Ein Reidentifikationsversuch ist dann relativ problemlos, wenn beide benötigten Datenfiles, also Zusatzwissen und Mikrodatenfile, die in ihnen enthaltenen Einzelfälle in den Datensätzen völlig kompatibel [in Bezug auf die Überschneidungsmerkmale] abbilden und mindestens einer der beiden Datenfiles die gesamte Population enthält."
- Hinzu kommt aber ein rationales Deanonymisierungsmotiv, das sich am Nutzen (Wert der Daten), in Verhältnis zum Aufwand der Deanonymiseriung bemisst: Wenn die Daten keinen Wert für Datenangreifer haben, unterbleiben Deanonymisierungsversuche, auch wenn sie technisch leicht zum Erfolg führen würden.

# 2 Thema: Big Data als Zusatzwissen brauchbar? Welche Datensammlungen sind in Deutschland verfügbar?

Das Wissen über externe Datenquellen und ihre Verfügbarkeit ist relevant für die realistische Abschätzung des Deanonymisierungsrisikos.

## *2.1 Allgemeine Informationen*

- Es gibt eine neue Arbeitsgruppe beim RatSWD "Relevanz von Big Data für die Wissenschaft" → http://www.ratswd.de/themen/bigdata

## Verfügbare Datensammlungen (Auswahl):

### Angereicherte Adressdaten (käuflich)

werden hauptsächlich für gezieltes Marketing verwendet. Die Anreicherungsmerkmale gehen teilw. ziemlich weit.

Beispiele:

- Schober
  auch Privatadressen angereichert mit Merkmalen zum Konsumerverhalten
  http://shop.schober.com/mall/1/files/sao_documentation_de.pdf
- Deutsche Post
  https://www.deutschepost.de/content/dam/dpag/images/D_d/DDP/Downloads/20150115_Zielgruppenadressen_Broschuere.pdf
    - Mikrogeografische Datenbank microdialog
      https://www.deutschepost.de/content/dam/dpag/images/D_d/DDP/Downloads/Oktober-2014/20141106_microdialog_Broschuere.pdf
- Mikrom
  http://www.microm-online.de/fileadmin/media/image/Datenuebersicht.pdf
- infas 360
  http://infas360.de/
  u.a.:
    - Migrationshintergrund
      http://infas360.de/service/news/neu-lokalisiert-deutsche-mit-migrationshintergrund/
    - Public Data Package
      http://infas360.de/service/news/endlich-public-data-fuer-alle/
- Nexiga
  http://www.nexiga.com/infocenter/produktbroschueren/

## Sonstige Datenbestände

- Europäische Meldeauskunft Riser
  "Die RISER ID Services GmbH ist ein One-Stop-Service Provider für die Melderegister-auskunft aus Deutschland, Österreich und der Schweiz"
  außerdem gibt es eine Umzugsdatenbank, in der auch Todesfälle eingetragen sind
  http://www.riserid.eu/fileadmin/user_upload/Benutzeranleitung/RISER_Benutzeranleitung_1503.pdf

  Kann eigentlich nur zur Aktualisierung von Adressbeständen genutzt werden, nicht zur Deanonymisierung von Mikrodaten (wegen fehlender Überschneidungsmerkmalen).

- Kürschners Handbücher für bestimmte Berufsgruppen
  Gelehrte, Künstler, Abgeordnete
  https://de.wikipedia.org/wiki/K%C3%BCrschners_Handb%C3%BCcher

  Stellen ein (nahezu) Vollregister der betreffenden Berufsgruppe dar, und könnten daher als Identifikationsfile benutzt werden, jedoch ist lt. der empirischen Überprüfung von Müller et al. anhand des Gelehrtenkalenders wegen Dateninkompatibilität eine Reidentifizierung kaum zweifelsfrei möglich. Die in solchen und ähnlichen Registern ausgewiesenen Berufsgruppen sollten jedoch in den Mikrodaten nicht abgrenzbar sein.

- Monitoring der Sozialen Netzwerke coosto
  http://www.coosto.com/de/module/

- google Suchbegriffe (google-Trends)
  http://marketingland.com/google-trend-goldmine-117626
  google trends help center: https://support.google.com/trends/?hl=en

  Die googles-Suchbegriffe spielten eine Zeitlang eine große Rolle in der Diskussion über Big Data, aber: KEINE MÖGLICHKEIT, DIE DATEN ZUR DEANONYMISIERUNG VON MIKRODATEN ZU NUTZEN

- Aggregierte Mautdaten des Bundesamt für Güterverkehr BAG
  Askitas/Zimmermann (2011): Nowcasting Business Cycles Using Toll Data, IZA Discussion Paper Series 5522
  http://ftp.iza.org/dp5522.pdf

  Die Daten wurden vom Bundesamt für Güterverkehr als Monatsaggregation zur Verfügung gestellt. Aus den Daten macht das IZA den "Toll Index".
  KEINE MÖGLICHKEIT, DIE DATEN ZUR DEANONYMISIERUNG VON MIKRODATEN ZU NUTZEN

- Daten der Intelligenten Stromzähler
  → Buchmann (2015)[3]
  Wer ist "Besitzer" der Daten? Wie ist der Datenschutz geregelt?
  KEINE MÖGLICHKEIT, DIE DATEN ZUR DEANONYMISIERUNG VON MIKRODATEN ZU NUTZEN

[3] Buchmann, Erik (2015): Wie kann man Privatheit messen? Privatheitsmaße aus der Wissenschaft. DuD 8/2015. S. 510-514, http://dx.doi.org/10.1007/s11623-015-0461-1

## *2.2 Relevante Aspekte*

- Hohe Kosten der Datenbeschaffung wirken sich prohibitiv für ihre wissenschaftliche Nutzung aus, wie Schimpl-Neimanns / Weiss (2014: 209) anhand des Zusammenhangs zwischen den Kosten für Daten der amtlichen Statistik und deren Nutzung durch die Wissenschaft aufzeigen[4].

## *2.3 Stimmen zum Thema:*

- Das Thema BigData und die Unsicherheit darüber, welcher Mißbrauch mit den überhand nehmenden Datensammlungen möglich ist, verunsichert die für den Schutz von Sozialdaten und Daten der amtlichen Statistik Verantwortlichen:

  "Aus meiner Sicht als nicht auf ein bestimmtes Gebiet des öffentlichen Rechts spezialisierten Verwaltungsjurist einer Datenschutzaufsichtsbehörde liegt eine Ursache für eine solche Zurückhaltung [der amtlichen Statistik, Daten zur Forschung bereitzustellen] eher in der Schwierigkeit, aufgrund der enormen technischen Veränderungen mit zunehmender Digitalisierung des Alltags die Risiken für eine Deanonymisierung Betroffener über Verschneidungen mit anderen Datensätzen kaum noch einschätzen bzw. überblicken zu können. Das in dieser Hinsicht vorhandene Gefahrenpotenzial wird in Zeiten von „Big Data“ weiter steigen." (Smolle 2015: 73)[5]

- Erik Buchmann (2015)[6] warnt: "Unter anderem hat eine Studie [1] im Jahr 2006 gezeigt, dass 63% der US-Amerikaner durch die Kombination der Attribute „Geschlecht“, „Geburtsdatum“ und „Postleitzahl“ (ZIP-Code) eindeutig zu identifizieren sind. Anhand dieser Attribute war es möglich, die frei erhältlichen US-Wählerlisten mit den ebenso frei verfügbaren medizinischen Daten der Group Insurance Commission zu verknüpfen, und so zahlreichen – eigentlich anonymen – Patientendaten eindeutige Namen und Anschriften zuordnen."

  Er beschreibt, dass sich beispielsweise aus den Daten der seit 1. Januar 2010 für Neubauten vorgeschriebenen 'intelligenten Stromzähler' u.a. "der Tagesablauf des Haushalts, Beschäftigungsverhältnisse, oder die Zahl der Bewohner rekonstruieren lassen." Und dass es zudem möglich ist "individuelle elektrische Geräte desselben Fabrikats auseinanderzuhalten, aus dem Stromverbrauch des Fernsehers das Fernsehprogramm abzuleiten oder anhand der im Stromverbrauch repräsentierten Gewohnheiten bestimmte Stromkunden wiederzuerkennen."

- "Eine Diskussion darüber, ob die Regeln von Müller et al. (1991) zur Erstellung von Scientific Use Files vor dem Hintergrund eines stetig wachsenden, immer detaillierten und immer leichter zugänglich werdenden Zusatzwissens noch gelten können, erscheint mir überfällig. Vielleicht sind auch deshalb die Zugangswege „Datenfernverarbeitung“ und „Gastaufenthalte“ zu präferieren. Hier spielt ein mögliches Zusatzwissen nur eine

---

4 Schimpl-Neimanns, Bernhardt / Weiss, Felix (2014): Zur Bereitstellung amtlicher Mikrodaten für die Wissenschaft aus sozialwissenschaftlicher Perspektive, AStA Wirtsch Sozialstat Arch (2014) 8: 205-219, http://dx.doi.org/10.1007/s11943-014-0156-3

5 Smolle, Michael (2015): Datenschutzrechtliche Anmerkungen zum Artikel „Vom Datenangreifer zum zertifizierten Wissenschaftler“ von Ulrich Rendtel, AStA Wirtsch Sozialstat Arch (2015) 9:73–78, http://dx.doi.org/10.1007/s11943-015-0163-z

6 s. FN 3

untergeordnete Rolle, weil es für eine mögliche Deanonymisierung nicht direkt eingesetzt werden kann." (Bender 2015: 246, FN 16)[7]

## *2.4 Mögliche Fragen für die Diskussion*

- Sind in Deutschland personenbezogene "Big Data" Bestände oder andere Datenbestände zugänglich, die sich zur Deanonymisierung von Mikrodaten verwenden lassen? Welche? Zu welchen Kosten?
- Wie verhält es sich mit dem besonderen Reidentifikationsrisiko, das lt. Müller et al. 1991 mit dem Merkmal Staatsangehörigkeit verbunden ist?

  "Da die Staatsangehörigkeit ein in der Regel leicht und kompatibel mit dem Mikrodatenfile erfahrbares Merkmal ist, wird zusätzlich empfohlen, dieses Merkmal nur stark vergröbert weiterzugeben". (Müller et al. 1991: 447)[8]

  Im SOEP wird die Staatsangehörigkeit z.B. unvergröbert weiter gegeben. Die kleinste Regionaleinheit im offline nutzbaren SOEP ist das Bundesland. Welche Bedingungen müssten erfüllt sein, welches Zusatzwissen müsste vorhanden sein, um anhand der Staatsangehörigkeit erfolgreiche Reidentifizierungen vorzunehmen?

---

7 Bender, Stefan (2015): Datenzugang in Deutschland: Der Paradigmenwechsel hat bereits stattgefunden. AStA Wirtsch Sozialstat Arch (2014) 8:237–248, http://dx.doi.org/10.1007/s11943-014-0158-1

8 s. FN 1

# 3 Thema: realistische Angriffsszenarien

Als erfolgreiche Deanonymisierung soll die Herstellung eines konkreten Personenbezugs verstanden werden. Das Reidentifikationsrisiko ist demnach das Risiko, Daten zweifelsfrei (d.h. eindeutig) Personen zuordnen zu können.

## *3.1 Grundlagen*

- Exzerpt von Müller et al. 1991 (→ Exzerpt_Müller_Blien_ea_Faktische-Anonymität-Mikrodaten.pdf)
- Auszug aus Müller et al. 1991: 41-45 (→ Müller-et-al_Meth-Grundl-Reidentifizierung.pdf)

## *3.2 Mögliche Fragen für die Diskussion*

- Welches personenbezogene und für die Reidentifizierung der Mikrodaten benutzbare Zusatzwissen ist tatsächlich verfügbar?
- Wie gefährlich sind einzigartige Merkmalskombinationen im Mikrodatenfile, bzw. unter welchen Bedingungen sind sie gefährlich?

  Antwort in → Müller et al.: ihre Gefährlichkeit wird stark überschätzt. Entscheidend ist nicht die Uniqueness von Merkmalsausprägungen in der Stichprobe, sondern die Uniqueness in der Population. Sowie, dass der Datenangreifer um die Uniqueness in der Population weiß.

  Wenn Identifikationsfile oder Mikrodatenfile eine (annähernde) Vollerhebung darstellen, sind diese Bedingungen erfüllt. Ebenso bei Stichproben, wenn der Datenangreifer über response knowledge verfügt, also weiß, dass eine bestimmte Person mit einzigartigen Merkmalen im Mikrodatenfile vertreten ist. (Daher ist die strikte Geheimhaltung der Stichprobenpläne eine grundlegende Regel zum Datenschutz).
- Verschärft oder entschärft die Verknüpfung von Register- mit Surveydaten das Deanonymisierungsrisiko?

  Argumente für Entschärfung

  - Es geht um die Verknüpfung von Surveydaten mit Sozialdaten → Surveydaten haben grundsätzlich einen verhältnismäßig niedrigen Stichproben-Auswahlsatz.

    Je kleiner dieser ist, desto kleiner ist auch das Deanonymisierungsrisiko. Ein Datenangreifer müsste über eine Gesamtdatenbank verfügen um sicher zu sein, dass eine bestimmte, singulär in der Stichprobe vorhandene Merkmalskombination auch in der Grundgesamtheit singulär ist. Wenn jedoch der Stichprobenauswahlsatz schon nah an einer Vollerhebung ist, sind singuläre Merkmalskombinationen mit viel höherer Wahrscheinlichkeit auch in der Grundgesamtheit (bzw. in der verfügbaren kleinsten regionalen Einheit) ebenfalls singulär. → Müller et al. 1991: 164f.

Argumente für Verschärfung

- Es stehen mehr Merkmale zur Verfügung, die zur Deanonymisierung genutzt werden können

  Jedoch: nur die Überschneidungsmerkmale zwischen Mikrodaten und Zusatzwissen können für Reidentifizierungsversuche benutzt werden.

  Jedoch: Müller et al.: mehr Merkmale erhöhen auch die Dateninkompabilität der Überschneidungsmerkmale.
- Mehr Merkmale könnten den Datennutzen für potentielle Angreifer erhöhen.

- Hat die Weiterentwicklung der Techniken des Statistical Matching einen Einfluss auf die Deanonymisierungswahrscheinlichkeit? Anders gefragt: Ist ein statistisches Matching, das mit einer Wahrscheinlichkeit oberhalb eines zu definierenden Schwellenwertes zutreffend ist, eine Deanonymisierung? → Fischzugszenario

  Antwort in → Müller et al.: Nein, denn der Datenangreifer kann mit diesen Techniken nicht unterscheiden, ob tatsächlich eine Identifizierung stattgefunden hat oder ob nur statistische Zwillinge einander zugeordnet wurden.
- Worin liegt ein denkbarer Nutzen der Deanonymisierung von DRV- und BA-Personendaten für potentielle Angreifer aus dem Kreis der Wissenschaft? Von anderen, "brisanteren" Daten (z.B. Gesundheitsdaten?)

# 4 Thema: Vertrauen statt Kontrolle?

## *4.1 Stimmen zum Thema*

**gezielte Deanonymisierungsversuche durch die Forschung?**

- Hafner / Ritchie / Lenz (2015)[9]: Das Risiko einer gezielten Enthüllung aufgrund persönlicher Neugierde des Forschers ist außerordentlich niedrig, so das Ergebnis am Beispiel des Community Innovation Surveys (CIS), eines europaweiten Betriebssurvey

  "In summary, spontaneous recognition is feasible but unlikely to have sufficient certainty to be worthwhile; a successful and informative match is theoretically possible but the practical problems are large. Most importantly, matching requires the researcher to actively search for the company; it is not an outcome of spontaneous recognition. The SUF licence agreement forbids attempting to identify any respondent; evidence suggests this is credible. Therefore, it appears that the risks of deliberate disclosure associated with researcher inquisitiveness are of a very low order." (Hafner / Ritchie / Lenz 2015: 7)

- Die bestehende Gesetzesgrundlage des BStatG erkennt zwar ein Wissenschaftsprivileg an, das dem Wissenschaftler gegenüber dem Rest der Datennutzer ein vergrößertes Analysepotential zur Verfügung stellt. Neben der Verpflichtung auf die Einhaltung der Datenschutzbestimmungen mit einer expliziten Strafandrohung im Falle einer Missachtung dieser Bestimmungen wird zusätzlich das Modell der faktischen Anonymität benutzt. Implizit wird damit dem Wissenschaftler die Absicht einer Deanonymisierung der Daten unterstellt: Allein die faktische Anonymität der Daten, die den Aufwand zur Deanonymisierung hoch treibt, schaffe die Gewähr, dass der Wissenschaftler von dem ohnehin verbotenen Treiben Abstand nimmt.." Rendtel 2014: 191[10]

- Die Empfehlungen von Müller et al. dienen nur zum Abfangen des im folgenden beschriebenen Restrisikos:

  "Bei der Unterstellung von Teilnahmekenntnis ergaben sich dagegen Anhaltspunkte für eine Risikokonstellation, bei denen unter Umständen in Einzelfällen eine erfolgreiche Reidentifikation mit vergleichsweise niedrigem Aufwand möglich erscheint. Dieser - allerdings äußerst seltene Fall (…) setzt das Zusammentreffen sehr spezifischer Risikofaktoren voraus und kann allgemein wie folgt charakterisiert werden: [FN 2]

  - Die im Mikrodatenfile gesuchte Person gehört einer sehr kleinen, durch ein spezifisches Merkmal eingrenzbaren Subpopulation an (…) (sachliche Tiefengliederung)
  - das Mikrodatenfile enthält tiefgegliederte Regionalinformationen, so daß in den jeweiligen Regionaleinheiten nur wenige Angehörige dieser spezifishen Subpopulation leben (regionale Tiefengliederung)
  - ein Forscher, der Zugang zu den Einzelangaben des Mikrodatenfile hat, kann sich Kenntnisse über einen Angehörigen dieser spezifischen Subpopulation beschaffen

---

9 s. FN 2

10 Rendtel, Vom potenziellen Datenangreifer zum zertifizierten Wissenschaftler – Für eine Neugestaltung des Wissenschaftsprivilegs beim Datenzugang, AStA Wirtsch Sozialstat Arch (2014) 8:183–197, http://dx.doi.org/10.1007/s11943-014-0148-3

und weiß, daß diese Person an der Mikrodatenerhebung (…) teilgenommen hat (Teilnahmekenntnis)

- Die Merkmale der Person sind genau in der Weise im Mikrodatenfile erfaßt, wie es der Forscher vermutet (Kompatibilität)

(…) Es ist wichtig darauf hinzuweisen, daß alle vier Bedingungen gleichzeitig erfüllt sein müssen. Bereits wenn eine der Bedingungen nicht gegeben ist, kann eine sichere Reidentifikation ohne den Aufwand unverhältnismäßig hoher Kosten nach den durchgeführten Experimenten als äußerst gering betrachtet werden. Das gleichzeitige Zusammentreffen aller Bedingungen kann bei Stichprobenerhebungen als außergewöhnlich seltenes Ereignis betrachtet werden."

[FN 2: "Damit es zu einem Datenangriff kommt, muss außerdem vorausgesetzt werden, daß ein Datenangreifer ein subjektives Interesse daran hat, das die denkbaren Kosten der Konsequenzen des Angriffs (Reputationsverlust, Vertragsstrafen, gesetzliche Strafen) übersteigt."] (Müller et al. 1991: 387-388)

**Unsicherheit über das, was "draußen" mit den Daten unternommen werden könnte, als Ursache für eine restriktive Auslegung der Datenweitergabe an die Forschung**

- Michael Smolle 2015: 73:[11] "Aus meiner Sicht als nicht auf ein bestimmtes Gebiet des öffentlichen Rechts spezialisierten Verwaltungsjurist einer Datenschutzaufsichtsbehörde liegt eine Ursache für eine solche Zurückhaltung [der amtlichen Statistik, Daten zur Forschung bereitzustellen] eher in der Schwierigkeit, aufgrund der enormen technischen Veränderungen mit zunehmender Digitalisierung des Alltags die Risiken für eine Deanonymisierung Betroffener über Verschneidungen mit anderen Datensätzen kaum noch einschätzen bzw. überblicken zu können. Das in dieser Hinsicht vorhandene Gefahrenpotenzial wird in Zeiten von „Big Data“ weiter steigen."
- Stefan Bender 2015: 240:[12] "Ein Output, der das FDZ der BA im IAB verlässt, kann potentiell von jedem angesehen werden. Daher ist dieser Output auf absolute Anonymität zu prüfen."

**Fehlendes Wissen der Forscher über Datenschutzbelange?**

- Stefan Bender 2015: 239: "Zwei ketzerische Fragen seien daher erlaubt: Warum soll ein Datenanbieter einem Forscher einen Vertrauensvorschuss geben, der diesen dann aus „Uninformiertheit“ missbraucht? Warum soll man einer Forscherin noch mehr Vertrauen entgegen bringen, wenn sie sich unter den gegebenen Bedingungen jetzt schon nicht um Datenschutzbelange kümmert, weil sie es nie gelernt hat? Es geht hier nicht um die Unterscheidung zwischen „vorsätzlich“ und „uninformiert“, es geht um den richtigen Umgang mit sensiblen Daten."
-

11 s. FN 5

12 s. FN 7

## *4.2 Mögliche Fragen für die Diskussion*

- Was unterscheidet die "drinnen" (Mitarbeiter der Prozessdaten produzierenden Institutionen, der Forschungsdatenzentren, der statistischen Ämter) und die "draußen" (Wissenschaft und Forschung) hinsichtlich der:
    - gesetzlichen/vertraglichen/sanktionsbewehrten Verpflichtung zu Geheimhaltung und Datenschutz?
    - Motivation zu gezielten Deanonymisierungsversuchen?
- Wäre die "schwedische Lösung" eine Option auch für Deutschland?
- Was müsste geschehen, um das Vertrauen in die Zuverlässigkeit externer Wissenschaftler hinsichtlich der Einhaltung der Datenschutzbelange gestärkt wird?

# 5 Wie geht es weiter - Remote Access statt SUFs?

## *Zu klärende Rahmenbedingungen des Remote Access:*

- § 75 SGB X-Antrag erforderlich für Remote Access zu Sozialdaten?
- Outputkontrolle manuell oder automatisch? Output faktisch oder absolut anonymisiert?
- Wer zahlt für die Kosten?
- Wie die Replizierbarkeit der Forschung sicher stellen?
- Kann ein zum Remote Access zugelassener Forscher von seinem Arbeitsplatzrechner aus zugreifen, oder muss er einen "FDZ-im-FDZ-Arbeitsplatz" buchen?

**Beispiel Schweden:**

"Den konsequentesten Weg hat sicherlich Schweden beschritten. In Schweden sind Wissenschaftler den gleichen Datenschutzgesetzen unterworfen wie die Mitarbeiter von Statistics Sweden. Dies hat in Schweden die Konsequenz, dass die Wissenschaftler selber für die Einhaltung der Datenschutzregeln verantwortlich sind. Es bedarf also keiner weiteren Beaufsichtigung durch die Mitarbeiter von Statistics Sweden. Unter diesen Rahmenbedingungen ist ein Remote-Access das Mittel der Wahl für beide Seiten (…). Beim Remote Access kann der Wissenschaftler direkt auf dem formal anonymisierten Datensatz von Statistics Sweden rechnen. Der Wissenschaftler hat also immer Zugang auf die neueste Fassung des Datensatzes und muss keine Vergröberung der Merkmale in Kauf nehmen. Weiterhin erhält er die Ergebnisse der Analyse sofort an seinem Bildschirm. Der Remote Server sichert jedoch den Datenbestand. Es kann lediglich auf dem Datensatz gerechnet werden. Statistics Sweden braucht keine separaten Scientific Use Files mehr erstellten. Es gibt keinen Versand der Files mehr und die Dokumentation wird zentral betrieben. Für Deutschland erscheint diese Verfahrensweise derzeit kaum denkbar. (Rendtel 2014: 191)[13]

## *Stimmen zum Thema*

- Schimpl-Neimanns / Weiss (2014)[14] bringen detaillierte Argumente für SUFs und gegen deren Ersetzung durch Remote Access vor, zumindest unter den für Remote Access zu Daten der amtlichen Statistik geltenden Rahmenbedingungen
- Stefan Bender (2014)[15] hält dagegen die Erstellung von SUFs aus verschiedenen Gründen nicht für den Königsweg der Datenbereitstellung und setzt eher auf die Ausweitung von Gastaufenthalten im Rahmen des "FDZ-im FDZ-Konzepts" sowie auf zukünftige Entwicklungen im Bereich remote access.

---

13 s. FN 10

14 s. FN 4

15 s. FN 7