Vorbemerkung

Es gingen viele wertvolle Ergänzungen zu dem kürzlich verschickten vorläufigen Ergebnispapier ein. Die vorliegende Fassung zeigt nun eine wesentlich umfassendere Sicht auf den Workshop als die doch sehr selektive vorläufige Fassung.

Ich habe versucht, eine in sich schlüssige Darstellung der Vorträge und Diskussionen des Workshops zu kompilieren. Falls noch etwas Wichtiges fehlt oder falsch dargestellt ist, bitte ich um entsprechende Hinweise.

Es gab auch Rückmeldungen von Vortragenden, welche die Darstellung ihrer Präsentation im vorläufigen Ergebnispapier zu Recht als verkürzt und verzerrt empfanden. Es war kein sinnvoller Ansatz, auch zu den Vorträgen nur das zu sammeln, was einzelnen Zuhörern als bemerkenswert auffiel. Daher sind in der vorliegenden Fassung vor allem auch die Aspekte genannt, deren Vermittlung den jeweiligen Vortragenden besonders wichtig war. Es sind aber nach wie vor keine vollständigen Inhaltsangaben der Vorträge, denn inzwischen stehen alle Präsentationen auf der Website zur Verfügung. → www.diw.de/suf-workshop2015

Vorträge "Sozialdatenschutz und Nutzung der Daten durch die Wissenschaft aus Sicht…"

a) … des Datenschutzreferates der DRV

siehe hierzu die Präsentation http://www.diw.de/suf-workshop2015/suf-workshop_zusammenfassung-gerold.506715.pdf

- Die Wissenschaft braucht eigentlich keine personenbezogenen Daten. In den allermeisten Fällen kann auf Namen oder Realidentifikatoren vollkommen verzichtet werden.
- Auch das FDZ der DRV, welches SUFs und PUFs für die Forschung erstellt und bereit stellt, kennt den konkreten Personenbezug nicht und hat auch nicht die Mittel, diesen zu rekonstruieren (Prinzip der mehrfachen Pseudonymisierung und im letzten Schritt Ersetzen der Pseudonyme durch systemfreie Datensatzkennungen). Die personenbezogenen Daten der DRV werden nach festgelegten Aufbewahrungsfristen gelöscht, soweit sie nicht für die Rentenberechnung relevant sind. Amtliche Statistiken werden aus doppelt pseudonymisierten Daten erzeugt. Die doppelt pseudonymisierten Rohdaten für die Statistiken sind die Basis, aus der durch Aufbereitung die SUFs und PUFs erzeugt werden. Im FDZ werden nur faktisch anomyme Daten dauerhaft aufbewahrt.
- Das Datenschutzreferat der DRV Bund empfiehlt, personenbezogene Originaldaten so früh wie möglich zu anonymisieren. Nach Möglichkeit soll die „Datenquelle“ aus der Gesamtheit aller bei ihr zu einer Person gespeicherten Daten diejenigen herausfiltern, die die Forscherin braucht (Grundsatz der Datensparsamkeit). Namen, Identifizierungs- und Kontaktdaten dürfen nicht an die Forscherin weiter gegeben, sondern müssen durch systemfreie Kennungen oder mindestens durch Pseudonyme ersetzt werden. Nur faktisch anonyme Daten dürfen an ForscherInnen ausgehändigt werden. (Echtdaten vergröbern oder in Kategorien gruppieren).
- Projekte, die Rohdaten aus verschiedenen Datenquellen erheben und Daten gleicher Personen in den verschiedenen Quellen zusammenführen wollen, sollten die Zuordnung nicht über personenbezogene Daten vornehmen, sondern über Pseudonyme.

- Die Datenbereitstellung seitens der Dateneigner ist aufwändig → ForscherInnen können nicht verlangen, dass alle Wünsche erfüllt werden → ein Miteinander von Dateneignern und Forschenden ist nötig
- Aufgabe eines Datenschutzbeauftragten bei der Beratung der ForscherInnen oder der „Datenquellen“ ist es, Lösungswege aufzuzeigen, wie ein Forschungsthema ohne *personenbezogene* Daten untersucht werden kann. Im Rahmen der Einhaltung der gesetzlichen Bestimmungen gilt es, Dinge möglich zu machen, Unterstützer für die Forschung zu sein.
- Wenn Daten der DRV mit Befragungsdaten gekoppelt werden sollen, dann wird von den Befragten nach Möglichkeit eine umfassende Einwilligung nach § 67 b SGB X eingeholt. In diesen Fällen liegen einige personenbezogene Daten (Versicherungsnummer, Geburtsdatum, Geburtsort, Geburtsname) im FDZ der DRV vor. Diese werden jedoch nicht dem Befragungsinstitut mitgeteilt, das andere personenbezogene Daten (z. B. die aktuelle Adresse) hält, und zu keinem Zeitpunkt mit inhaltlich aussagekräftigen weiteren Daten zusammengebracht.

b) … der Datenschutzabteilung des IAB und des FDZ der BA im IAB

siehe hierzu die Präsentationen http://www.diw.de/suf-workshop2015/suf-workshop_ross.pdf und http://www.diw.de/suf-workshop2015/suf-workshop_antoni_schmucker.pdf

- Die Nutzung von Sozialdaten für wissenschaftliche Zwecke steht in dem Spannungsverhältnis zwischen dem Recht auf informationelle Selbstbestimmung und dem Artikel 5 GG (Wissenschafts- und Forschungsfreiheit).
- Die aus Prozessdaten der BA resultierenden Forschungsdaten des SGB II und III unterliegen einem strengen gesetzlichen Schutz, denn diese Daten sind als "ohne Einwilligung der Betroffenen" - also zwangsweise - erhoben anzusehen.
- Die Weitergabe von Forschungsdaten an externe Forschungseinrichtungen ist nur für einen ganz bestimmten Forschungszweck möglich (sog. Zweckbindung).
- Die Verantwortlichen im IAB begehen als übermittelnde Stelle eine Straftat oder Ordnungswidrigkeit, wenn sie dabei gegen die gesetzlichen Schutzvorschriften verstoßen.
- Auch faktisch anonymisierte Forschungsdaten der BA unterliegen einem gesetzlichen Schutz: Faktisch anonymisierte Daten dürfen nach § 282 Abs. 7 SGB III nur für Zwecke der Arbeitsmarkt und Berufsforschung bereitgestellt werden.
- Das FDZ übermittelt diese Daten nur befristet für jeweils ein konkretes Projekt.
- Derzeit sind alle vom FDZ der BA angebotenen verknüpften Datensätze (Survey- mit Prozessdaten verknüpft) als schwach anonymisiert (und damit nicht faktisch anonymisiert) eingestuft und daher nur nach einem vereinfachten Genehmigungsverfahren gemäß § 75 SGB X im Rahmen von Gastaufenthalten und anschließendem Remote Execution mit Outputkontrolle nutzbar.

Die anschließende Diskussion drehte sich u.a. darum, wie faktisch anonymisierte Daten und schwach anonymisierte Daten zu unterscheiden sind, ob diese Unterscheidung gesetzlich begründet ist, und welche rechtlichen Konsequenzen sich aus den Unterschieden ergeben. Die Möglichkeit von Zertifizierungsverfahren mit dem Ziel, Datensätze als ‚faktisch anonym‘ einzustufen, wurde grundsätzlich befürwortet.

c) … des Referats II a 2 des Bundesministeriums für Arbeit und Soziales

siehe hierzu die Präsentation http://www.diw.de/suf-workshop2015/suf-workshop_pr%C3%A4sentation-solka.pdf

(Das Referat IIa 2 des BMAS ist u.a. zuständig für die Genehmigung von Anträgen auf Datennutzung für wissenschaftliche Zwecke nach § 75 SGB X)

- Faktisch anonymisierte Daten unterliegen keinen datenschutzrechlich begründeten Zugangsbeschränkungen. Jedoch bereitet die Abgrenzung Sozialdaten - faktisch anonymisierte Daten in der Praxis Schwierigkeiten.
- Daten der Bundesagentur für Arbeit unterliegen jedoch auch wenn sie faktisch anonymisiert sind, der Zweckbindung nach § 282 Abs. 7 SGB 3.
- Die Nutzungsgenehmigung setzt voraus, dass ein klar bestimmter und anders nicht zu erlangender Forschungszweck benannt werden kann. Diese Zweckbindung ist in manchen Fällen problematisch, etwa wenn es darum geht, eine Forschungsdatenbank aufzubauen.

***

In diesem Block wurde deutlich, dass das FDZ-RV und das FDZ-BA zwei unterschiedliche Herangehensweisen an die Aufgabe "Datenzugang für die Wissenschaft" pflegen. Diese sind auf unterschiedliche Gegebenheiten zurückzuführen. Zum einen ist das IAB als das Forschungsinstitut der Bundesagentur für Arbeit ausdrücklich mit Arbeitsmarkt- und Berufsforschung u.a. anhand der anfallenden Prozessdaten beauftragt. Die Daten, die bei der BA anfallen, werden dauerhaft aufbewahrt, so dass für verschiedene Projekte unterschiedliche Datenquellen zur Generierung der Forschungsdaten herangezogen werden können.

Bei der Rentenversicherung gibt es kein Pendant zum IAB. Bei der RV werden die Daten zu den Versicherten nach der gesetzlichen Aufbewahrungspflicht nur aufbewahrt, soweit sie zur Rentenberechnung erforderlich sind. Alle anderen Informationen werden nur als Statistikfiles oder als aufbereitete mehrfach pseudonymisierte oder anonymisierte Daten aufbewahrt.

Daraus folgend unterscheiden sich die Verfahren für Verknüpfungsprojekte: Beim FDZ der RV steht ein schlankerer und vorab festgelegter Datenkranz zur Verfügung, aus dem ein SUF bereitgestellt wird. Daher entfallen Genehmigungsverfahren durch die Aufsichtsbehörde in den meisten Fällen. Beim FDZ der BA können die gewünschten Forschungsdaten flexibel zusammengestellt werden, allerdings muss jedes von externen ForscherInnen geplante Projekt das Genehmigungsverfahren nach § 75 SGB X durchlaufen.

Vortrag: Synopse der Anonymisierungsmaßnahmen in verschiedenen Standard-SUFs und der Empfehlungen von Müller et al. (1991)

siehe hierzu die Präsentation http://www.diw.de/suf-workshop2015/suf-workshop_vortrag-erhardt.pdf
sowie die Synopse http://www.diw.de/suf-workshop2015/suf-workshop_synopse.pdf

- Die meisten der vorgestellten Datenbestände richten sich nach dem Empfehlungen von Müller et al. (1991).

- Die überraschende Erkenntnis aus der Studie von Müller et al. (1991), in der das Deanonymisierungsrisiko für den Mikrozensus (und die Einkommens- und Verbrauchsstichprobe) empirisch überprüft wurde, ist jedoch weitgehend vergessen worden: Alle geprüften Angriffsszenarien ergaben die faktische Anonymität der geprüften Daten. Die Ergebnisse erwiesen sich außerdem als übertragbar auf andere Datenbestände, sofern es sich um eine Stichprobenerhebung handelt und sofern kein personenbezogenes Zusatzwissen zur Verfügung steht, das aufgrund von kompatiblen Überschneidungsmerkmalen zum "Knacken" der Daten benutzt werden könnte.
- Die Empfehlungen von Müller et al. (1991) sollten daher eigentlich nur noch ein fiktives Restrisiko abfangen, welches sich theoretisch beim Zusammentreffen einer äußerst seltenen Risikokonstellation in den Daten mit einem irrationalen (weil in Gegensatz zu seiner beruflichen Handlungslogik und zu rationalen Nutzenerwägungen auf eine Deanonymisierung von Einzelfällen abzielenden) Wissenschaftler ergibt. (Siehe Auszug aus Müller et al., http://www.diw.de/suf-workshop2015/suf-workshop_auszug-anonymisierung_mueller_blien_ea.pdf)
- In der anschließenden Diskussion wurde der Hinweis eingebracht, dass die Empfehlungen von Müller et al. (1991) nicht in erster Linie durch die Ergebnisse des damaligen Forschungsprojekts begründet waren, sondern vor allen Dingen einen Kompromiss zu den Vorstellungen des Statistischen Bundesamts und des Datenschutzbeauftragten darstellten.
- Außerdem wurde darauf hingewiesen, dass es seit Müller et al. (1991) kaum Projekte oder zitierte Publikationen gab, die auf abweichende oder geänderte Anforderungen hindeuten würden. Auch Bacher/Brand/Bender (2002)[1] sind nicht zu anderen Schlussfolgerungen gekommen und konnten die 1991 aufgestellten Prinzipien nicht widerlegen.
- Es wird mehrfach als bemerkenswert genannt, dass es in den vergangenen 40 Jahren keinen (bekannten) Datenschutzverstoß gegeben hat.

Vortrag: Weitergabe von Forschungsdaten zwischen Risiko und Ungewissheit

siehe hierzu die Präsentation http://www.diw.de/suf-workshop2015/suf-workshop_vortrag-lenz.pdf

- Die erfolgreichen Zuordnungen zwischen den originalen Daten und den daraus erstellten anonymisierten Mikrodaten ergeben das theoretische Maximalrisiko -das "worst case"-szenario - , das in der Realität aber nie vorkommt. Denn einem realen Datenangreifer stehen die nicht anonymisierten Originaldaten nicht zur Verfügung, und wenn doch, bräuchte er den anonymisierten Datensatz nicht zu deanonymisieren.
- Ein *reales* Reidentifikationsrisiko kann dagegen nur anhand der erfolgreichen Zuordnungen zwischen einem externen Identifikationsfile und den anonymisierten Mikrodaten abgeschätzt werden.

[1] s. Bender, Stefan / Brand, Ruth / Bacher, Johann (2010): Re-identifying register data by survey data: An empirical study.Statistical Journal of the United Nations ECE 18 (2001) 373–381. https://www.researchgate.net/publication/228434618_Re-identifying_register_data_by_survey_data_An_empirical_study

- Absolute Anonymität bedeutet nicht eine Quote richtiger Zuordnungen von Null, da schon eine rein zufällige Zuordnung richtige Treffer erbringen kann. (Ein realer Datenangreifer könnte aber richtige von falschen Zuordnungen nicht unterscheiden).

Diskussionen am 2. Tag

siehe hierzu den vorab versendeten Input zu den Diskussionen http://www.diw.de/suf-workshop2015/suf-workshop_input_diskussionen.pdf

a) Big Data

- In der Diskussion wurde einerseits die Ansicht vertreten, dass Big Data nicht dazu benutzt werden kann, die von der eigenen Institution herausgegebenen SUFs zu "knacken". Andererseits gab es auch TeilnehmerInnen, die sich dessen nicht so sicher waren. Am Ende schien aber Konsens darüber zu bestehen, dass der Verweis auf Big Data für sich kein Grund sein kann, den Zugang der Wissenschaft zu amtlichen und Sozialdaten restriktiver zu handhaben.
- Mehrere TeilnehmerInnen äußerten Zweifel daran, dass es großes kommerzielles Interesse geben könnte, den Aufwand zu betreiben, um Sozialdaten zu deanonymisieren. Diese Daten haben häufig erhebliche Timelags, was ihren Nutzen für die meisten Arten kommerzieller Verwertung deutlich reduziert. Auch sind Stichprobendaten für kommerzielle Nutzung i.d.R. ungeeignet.
- Setzt "Big Data" die Bedingungsfaktoren der Deanonymisierungsrisiken, wie sie in Müller et al. beschrieben sind (nämlich: Vorhandensein eines personenbezogenen Identifikationsfiles mit kompatiblen Überschneidungsmerkmalen zu den zu identifizierenden Mikrodaten, beide Files müssen sich mindestens teilweise auf die selben Personen beziehen) außer Kraft?
  Falls dem nicht so ist, könnte das Big Data-Thema im Hinblick auf Deanonymisierungsrisiken auf die Frage reduziert werden: Liefert Big Data (verfügbare) kompatible Überschneidungsmerkmale zu anonymisierten amtlichen und Sozialdatenbeständen?
- Es gab Einigkeit darüber, dass wir alle nicht genug über die Verfügbarkeit und die Risken von "Big Data" wissen. Es sollte ein Forschungsprojekt zu diesem Thema initiiert werden (siehe Punkt "Fazit" weiter unten).

b) Vertrauen in die externen Wissenschaftler

Aufklären, Wissen vermitteln, Bewußtsein schaffen

Dieses ist auf beiden Seiten erforderlich:

- Die externen Wissenschaftler sollten stärker über Datenschutzprobleme und Möglichkeiten, sie zu vermeiden, aufgeklärt werden (s.u., Abschnitt "Externe Wissenschaftler in Datenschutzanforderungen schulen").
- Die Datenhalter und Datenschutzmitarbeiter in den FDZs sollten bei ihren postulierten Angriffsszenarien die Ergebnisse von Müller et al. sowie die Handlungslogik von Wissenschaft stärker berücksichtigen:
  - Nach Müller et al. ergeben mehr Merkmale (z.B. durch Verknüpfung von Survey- und Sozialdaten) nicht automatisch ein größeres Reidentifikationsrisiko für den verknüpf-

ten File. → "Big Data" ergibt nicht automatisch ein größeres Reidentifikationsrisiko für anonymisierte Sozial- und amtliche Mikrodaten. Weiterhin: Die technische Entwicklung für sich genommen verändert das Reidentifikationsrisiko nicht, da dieses nicht von der Rechnerkapazität abhängt, sondern von der Verfügbarkeit von kompatiblem Zusatzwissen.

Dazu gab es jedoch keinen Konsens. Einige TeilnehmerInnen haben den Standpunkt vertreten, dass das Deanonymisierungsrisiko durch deutlich mehr Merkmale steigt. Andere haben das in Frage gestellt. Die Frage verblieb kontrovers.

Zudem kam in der Feedbackrunde zum Ergebnispapier der Einwand: Die Konzentration auf einzelne sich genau überschneidende Merkmale blendet die sich aus der Kombination vieler ungefähr ähnlicher Merkmale ergebenden Analysepotentiale aus. Die sich daraus ergebenden Risiken sollten expliziter dargestellt werden.

Diese Kontroverse verweist ebenfalls auf die Notwendigkeit, den Sachverhalt unter heutigen Bedingungen erneut wissenschaftlich zu untersuchen.

  - "Wissenschaftler brauchen den Personenbezug nicht" (Gerold) → jedenfalls nicht quantitativ empirisch Arbeitende. Die Rede vom "Datenangreifer" verschiebt das Denken in eine falsche Richtung. Der Handlungslogik der Human-, Sozial- und Wirtschaftswissenschaften liegt das Ziel der Deanonymisierung vollkommen fern - es wäre eine pure Verschwendung von Ressourcen ohne einen erkennbaren Nutzen.

- Regelverletzungen durch Wissenschaftler kommen ab und zu vor: durch Nachlässigkeit, aber vermutlich auch durch fehlende Einsicht in Sinn und Zweck der einzelnen Bestimmungen. Es wäre erforderlich, die Handlungslogik der jeweiligen Gegenseite besser zu verstehen, um a) auf Seiten der Wissenschaft die Regeln als sinnvoll zu erkennen und aktiv zu bejahen und b) auf Seiten der Datenhalter die Regeln auf das durch Gesetze und Datenschutz erforderliche Maß zu beschränken.

Externe Wissenschaftler in Datenschutzanforderungen schulen

- Wie Rainer Lenz in seinem Vortrag erwähnte, ist Eurostat dabei, Schulungsmaterialien für eine spätere Lernplattform zur Sensibilisierung der Nutzer auf Probleme der Datengeheimhaltung zu entwickeln. Nach der Auseinandersetzung mit dem Material können die Nutzer anhand von Testfragen ihr Wissen prüfen. Mehrere Teilnehmer bekundeten Interesse, dieses oder ähnliches Material für die "Einarbeitung" neuer Datennutzer zu verwenden. Die Lernplattform von Eurostat ist aber noch im Entwicklungsstadium, und passt nicht umstandslos auf deutsche Gegebenheiten.
- Auch an der MPG wird eine Lernplattform entwickelt, diese ist aber nur intern zugänglich, und kann nicht an Dritte weiter gegeben werden (außer vielleicht im Rahmen expliziter Kooperationsprojekte).
- Jan Goebel: es wäre anzustreben, dass die FDZs eine solche Plattform gemeinsam entwickeln. Das FDZ im IAB unterstützt diese Idee. Wäre der RatSWD ein möglicher Organisator? Ansprechpartner? Jan Goebel wird das Thema im Ständigen Ausschuß FDI des RatSWD einbringen.

Androhung von Sanktionen im Datennutzungsvertrag als Schutz vor Deanonymisierungsversuchen?

- Die Diskussion erbrachte: angedrohte Geldstrafen sind ein zahnloser Tiger, weil man sie im Zweifelsfall wahrscheinlich nicht durchsetzen kann. Der Verlust von Reputation sowie der

Ausschluss von der Datennutzung für einen bestimmten Zeitraum sind wirkungsvollere Maßnahmen. Jedoch hat ein Reputationsverlust („Forschertod“) nicht für jeden das gleiche Drohpotential, denn viele verlassen nach (oder gar während) der Promotion die Wissenschaft.

- Regelverstöße kommen offenbar vor, jedoch ist kein Fall einer versuchten (absichtlichen) Deanonymisierung bekannt.
- Da absichtliche Deanonymisierungsversuche offenbar selten bis gar nicht vorkommen, sondern eher versehentliche Verstöße oder Verstöße wegen mangelnder Einsicht in die Notwendigkeit bestimmter Regeln, sollte Aufklärung / Wissen vermitteln im Vordergrund stehen.
- Neben den vertraglichen Möglichkeiten, hier sind sich alle Teilnehmer einig, kann eine zusätzliche Aufklärung der Nutzer über den Sinn der Datengeheimhaltung und die möglichen Folgen bei Verletzung derselben schon a priori Fälle unbeabsichtigten Datenmissbrauchs reduzieren.

Weitere Ergebnisse (nicht konkret einem Programmpunkt zuordenbar bzw. wurde mehrfach angesprochen)

- Faktisch anonymisierte Daten sind keine Sozialdaten mehr und unterliegen keinen Beschränkungen durch Datenschutzgesetze (so Gerold), jedoch ist für Daten der Bundesagentur für Arbeit die Zweckbindung § 282 Abs. 7 SGB III zu beachten (Roß, Solka)
- Die Einordnung als faktisch anonym bezieht dabei aber den Zugangsweg mit ein. Ein bestimmter Datensatz kann im Zugangsweg Gastarbeitsplatz (keine unkontrollierte Zuspielung von anderen Datenquellen möglich) als faktisch anonym angesehen werden, als an die Wissenschaftler ausgelieferter Datensatz aber nicht.
- Die Erstellung von SUFs ist aufwändig und teuer. Der Daten-Fernzugriff ist ebenfalls - und dauerhaft - aufwändig und teuer, z.B. wegen der erforderlichen Outputkontrolle → Finanzierung? Minimierung des Aufwands?
- Hauptproblem /-schaden eines erfolgten und publik gewordenen Datenschutzverstoßes ist ggf. nicht zwingend der Schaden für die einzelnen Personen. Vielmehr ist aus Sicht der Datenhalter und -anbieter (z.B. BA oder DRV) ein massiver Reputationsverlust zu befürchten. Denn die Öffentlichkeit interessiert im Zweifel nicht, dass es 40 Jahre lang ohne Vorfälle gelaufen ist. Der (Vertrauens-)Schaden ist dadurch kaum zu begrenzen.

Offen gebliebene Fragen bzw. Punkte, an denen unterschiedliche Ansichten herrschen

Datenzugänge

- Wie könnten die bestehenden Zugangsformen zu Sozial- und amtlichen Mikrodaten für die Wissenschaft so weiter entwickelt werden, dass Personaleinsatz und Kosten auf Seiten der Datenhalter minimiert werden und gleichzeitig die Usability für die Wissenschaft maximiert wird (Usability meint sowohl Analysepotential als auch Integrierbarkeit in die übliche wissenschaftliche Arbeitsweise)?

  Z.B. scheint einerseits die SUF-Erstellung aufwändig und ressourcenintensiv zu sein. Andererseits bringt aber auch der Datenfernzugriff Kapazitätsprobleme in technischer wie personeller Hinsicht mit sich, z.B. wegen der damit verbundenen Outputkontrolle.

- Wer finanziert auf lange Sicht die Kosten des Datenzugangs für die Wissenschaft? Wie ist damit umzugehen, dass in den FDZs neue Daten oft ganz oder anfänglich über zeitlich begrenzt fließende Drittmittel erschlossen werden, das Datenangebot anschließend aber über Stammmittel finanziert werden muss und für Forscher idealerweise kostenlos sein soll?

Wann verlieren Sozialdaten ihren Personenbezug und somit ihre Sozialdaten-Eigenschaft?

- Vortrag Gerold (mit Bezug auf DRV-Daten): Anonyme Rohdaten dürfen ohne Beschränkung durch Datenschutzgesetze verwendet werden.
- Vortrag Roß (mit Bezug auf BA-Daten): Faktisch anonymisierte Daten unterliegen immer noch einem - wenn auch abgeschwächten - gesetzlichen Schutz.
  Ergänzung von Frau Roß aus der Feedback-Runde hierzu: Das Gesetz sieht einen Schutz der informationellen Selbstbestimmung für alle Menschen vor und schwächt dabei nicht ab für spezielle Situationen. Das Risiko einer Deanonymisierung von faktisch anonymisierten Daten ist in bestimmten kontrollierten Umgebungen geringer als bei z.B. lediglich pseudonymisierten Einzeldaten. Daher können die Bedingungen für den Zugang unterschiedlich gestaltet werden.
- Vortrag Solka (mit Bezug auf Sozialdaten allgemein): Datenschutz greift nicht bei anonymisierten Daten.

Die gesetzlichen Grundlagen zu diesem Punkt:

- § 67 SGB X, Begriffsbestimmungen:
  (1) Sozialdaten sind Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person (Betroffener), die von einer in § 35 des Ersten Buches genannten Stelle im Hinblick auf ihre Aufgaben nach diesem Gesetzbuch erhoben, verarbeitet oder genutzt werden.
  (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.
  (8a) Pseudonymisieren ist das Ersetzen des Namens und anderer Identifikationsmerkmale durch ein Kennzeichen zu dem Zweck, die Bestimmung des Betroffenen auszuschließen oder wesentlich zu erschweren.
- § 75 SGB X, Übermittlung von Sozialdaten für die Forschung und Planung:
  (1) Eine Übermittlung von Sozialdaten ist zulässig, soweit sie erforderlich ist für ein bestimmtes Vorhaben
  1. der wissenschaftlichen Forschung im Sozialleistungsbereich oder der wissenschaftlichen Arbeitsmarkt- und Berufsforschung oder
  2. der Planung im Sozialleistungsbereich durch eine öffentliche Stelle im Rahmen ihrer Aufgaben

  und schutzwürdige Interessen des Betroffenen nicht beeinträchtigt werden oder das öffentliche Interesse an der Forschung oder Planung das Geheimhaltungsinteresse des Betroffenen erheblich überwiegt.
- § 282 SGB III Arbeitsmarkt- und Berufsforschung:
  (7) Die Bundesagentur übermittelt wissenschaftlichen Einrichtungen auf Antrag oder Ersuchen anonymisierte Daten, die für Zwecke der Arbeitsmarkt- und Berufsforschung erforderlich sind."

Weiterer kontroverser Punkt

- Das Wissen um die von Müller et al. (1991) herausgearbeiteten Bedingungsfaktoren einer Reidentifizierung (die auch noch einmal sehr klar und prägnant bei Heike Wirth (2006)[2] zusammengestellt sind), ist bei manchen der Akteure, die über die faktische Anonymität von Daten zu befinden haben, nicht ausreichend vorhanden. Daher werden Risiken tendenziell überschätzt: weder Big Data noch die technische Entwicklung stellen *per se* ein erhöhtes Deanonymisierungsrisiko für anonymisierte Mikrodaten dar.

  Die Gegenposition: Auf dem Workshop wurden die Risiken und Gefahren zu sehr ausgeblendet. Z.B. fehlte es an Offenheit bei der Suche nach den Risiken und Gefahren für die informationelle Selbstbestimmung.

  Aus beiden Positionen ergibt sich die Synthese, dass die möglichen Risiken unter heutigen Bedingungen erneut erforscht werden sollten.

a) Fazit oder: wie könnte es weiter gehen?

- Es sollte ein Projekt gestartet werden, welches das unter dem Stichwort "Big Data" für WissenschaftlerInnen verfügbare Zusatzwissen und dessen Struktur eruiert. Damit die Ergebnisse nicht schon veraltet sind, bevor sie veröffentlicht werden, sollte das Gewicht auf solchen Quellen gefährlichen Zusatzwissens liegen, die Schlüsselmerkmale zu den vertraulichen Daten enthalten bzw. aus denen solche Merkmale generiert werden können. Derart, dass das Projekt einen Maßstab für die Bewertung von "neuen" Big-Data-Sachverhalten liefert (Nachhaltigkeit).
- Angedacht wurde auch, eine Neuauflage des Projekts von Müller et al. unter heutigen technologischen Bedingungen und heutigem möglichen Zusatzwissen (Big Data?) zu initiieren.
- Könnten nicht die Maßnahmen zur Herstellung faktischer Anonymität im SUF sowie die Outputkontrolle im Licht realistischer Angriffsszenarien und der bisherigen Erfahrungen mit wissenschaftlicher Nutzung von Sozialdaten deutlich verschlankt werden, ohne dass Datenschutzgesichtspunkte dabei zu kurz kommen? Die frei werdenden Ressourcen könnten für die Entwicklung von Strukturen zur Schulung und Zertifizierung externer DatennutzerInnen aus der Wissenschaft und in die Weiterentwicklung von sicheren Lösungen für einen nutzerfreundlicheren Datenzugang eingesetzt werden.
- Andere Zugangswege wie Remote Access / Remote Execution wurden auf dem Workshop eher ausgeblendet, da der Fokus auf dem Zugangsweg Scientific Use File lag. Diese Alternativen sollten jedoch parallel zu Verbesserungen im Bereich SUF ebenfalls weiterentwickelt werden. Man müsste zu diesem Thema eigentlich einen weiteren Workshop veranstalten, mit dem Ziel, über technische und organisatorische Lösungen zu beraten, welche möglichst gut vereinbar sind mit der wissenschaftlichen Arbeitsweise, den Aufwand für die FDZs niedrig halten und gleichzeitig den erforderlichen Datenschutz gewährleisten.

---

[2] Heike Wirth (2006): Anonymisierung des Mikrozensuspanels im Kontext der Bereitstellung als Scientific-Use-File. Methodenverbund "Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe", Arbeitspapier Nr. 11, https://www.destatis.de/DE/Methoden/Methodenpapiere/Mikrozensus/Arbeitspapiere/Arbeitspapier11.pdf?__blob=publicationFile