

24.02.2015

Stellungnahme des Rates für Sozial- und Wirtschaftsdaten

zu ausgewählten Punkten des aktuellen Entwurfs
der **EU-Datenschutz-Grundverordnung (DSGVO)**

- *Diese Stellungnahme basiert auf der Fassung der Legislativen Entschließung des Europäischen Parlaments vom 12. März 2014 zum Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Datenverkehr (allgemeine Datenschutzverordnung) – COM(2012)0011 – C7 - 0025/2012 – 2012/0011(COD) – Erste Lesung*
- *Diese Stellungnahme wurde verfasst in Kenntnis, aber ohne besondere Berücksichtigung, der Änderungen der Italienischen Präsidentschaft im Ratsdokument 15395/14 vom 19. Dezember 2014 – Interinstitutional File 2012/0011(COD) Limite*

Hintergrund

Mit der im Januar 2012 durch die EU-Kommission vorgelegten „Verordnung zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Datenverkehr“, bekannt als EU-Datenschutz-Grundverordnung, soll die Europäische Union erstmals eine einheitliche Datenschutzregelung erhalten, die für alle Mitgliedsstaaten eine bindende rechtliche Wirkung hat. Die im Jahr 1995 verabschiedete Datenschutzrichtlinie 95/46/EG wird damit ersetzt und eine Modernisierung und Anpassung an die technologischen Entwicklungen vollzogen.

Die über 3.000 im Innen- und Justizausschuss des Europäischen Parlaments eingereichten Änderungsvorschläge zum Kommissionsentwurf der Grundverordnung belegen eindrücklich, welche Bedeutung einem zukünftig harmonisierten Datenschutzrecht beigegeben wird.

Zum aktuellen Zeitpunkt hat der Trilog des Europäischen Rates, der Europäischen Kommission und des Europäischen Parlamentes zur Verhandlung der EU-Datenschutz-Grundverordnung noch nicht begonnen. Nach Einreichen der Änderungen des Ratsdokumentes im Zuge der italienischen Ratspräsidentschaft verhandelt eine Ratsarbeitsgruppe weiter. Es ist davon auszugehen, dass eine konsolidierte Fassung des Ratsdokumentes zur Verfügung gestellt wird und der Trilog vor der Sommerpause 2015 beginnt.

Vor diesem Hintergrund bezieht der Rat für Sozial- und Wirtschaftsdaten (RatSWD) Stellung, um die Perspektive der Sozial- und Wirtschaftswissenschaften auf ausgewählte Punkte des verhandelten Entwurfs deutlich zu machen.

Der RatSWD ist ein unabhängiges Expertengremium, welches seit 2004 die Bundesregierung in Fragen der Weiterentwicklung der Forschungsdateninfrastruktur berät. Der RatSWD hat dazu beigetragen, dass der Zugang für die Wissenschaft zu amtlichen Daten und wissenschaftlichen Daten unter Wahrung des Datenschutzes maßgeblich verbessert werden konnte.

Datenschutz in den Sozial- und Wirtschaftswissenschaften

Datenschutz, im Sinne des Schutzes der persönlichen Daten von Bürgerinnen und Bürgern vor unbefugter Offenlegung oder Missbrauch, ist essentiell für die empirische Forschung. Die Bereitschaft von Menschen, an wissenschaftlichen Datenerhebungen teilzunehmen und wahrheitsgemäße Angaben zu teils sehr persönlichen Sachverhalten zu machen, steht in einem engen Zusammenhang mit der Zusicherung auf und Realisierung von Anonymität; hierzu gehört die anonyme Auswertung der Daten. Diese Grundsätze des Datenschutzes gelten auch für Sekundäranalysen von administrativen Daten (wie zum Beispiel Steuerdaten) oder Registern (wie zum Beispiel den Melderegistern). Verstöße gegen den Datenschutz widersprechen der Handlungslogik und den Interessen wissenschaftlicher Forschung, da sie unethisch wären, zu einem Vertrauensverlust in der Bevölkerung führen und damit die Forschung von ihrer empirischen Basis abschneiden könnten.

Vor diesem Hintergrund begrüßt der RatSWD das Vorhaben der EU, mittels einer Datenschutz-Grundverordnung (im Folgenden DSGVO), den Schutz persönlicher Daten europaweit zu harmonisieren und verbindlich festzulegen. Ebenfalls wird begrüßt, dass die DSGVO in vielfacher Hinsicht der deutschen Datenschutzgesetzgebung ähnelt und vom Grundsatz forschungsfreundlicher Regularien geprägt ist. Jedoch finden sich in der DSGVO auch Punkte, die in der Rechtspraxis die empirische Forschung zukünftig beeinträchtigen könnten. Dies betrifft insbesondere die medizinische Forschung bzw. die Nutzung von Gesundheitsdaten. Die diesbezüglichen Bedenken wurden bereits von Seiten der einschlägigen Fachgesellschaften vorgebracht und vom RatSWD¹ unterstützt, da auch in der Sozial-² und Wirtschaftsforschung vermehrt Gesundheitsdaten erhoben oder genutzt werden.

Die folgenden Ausführungen beziehen sich daher auf die mit Blick auf die Analyse von Sozial- und Wirtschaftsdaten im Entwurf der DSGVO enthaltenen **Detailprobleme**, die in der bisherigen Diskussion keine oder wenig Beachtung fanden. Da sich diese Details in Zukunft als eine deutliche Barriere für die empirische Sozial- und Wirtschaftsforschung erweisen können, weisen wir hier darauf hin und machen Lösungsvorschläge.

¹ Vgl. Stellungnahme des Rates für Sozial- und Wirtschaftsdaten zum aktuellen Stand der EU-Datenschutzverordnung von 27.08.2014 (www.ratswd.de/dl/RatSWD_Stellungnahme_EU_DATA.pdf).

² Sozialforschung wird Rahmen dieser Stellungnahme in ihrer breiten Definition verstanden.

Problematik

Zentrales Problem ist (sowohl im Kommissions- wie auch im Parlamentsentwurf) die lediglich implizite, zugleich aber wirkmächtige Definition der Anonymisierung.

Dem Konzept der Anonymisierung kommt in der empirischen Sozial- und Wirtschaftsforschung eine herausgehobene Bedeutung zu, da so einmal erhobene Daten im Rahmen von Sekundäranalysen Drittforschern in anonymisierter Form verfügbar gemacht werden können.³ Sogenannte Scientific Use Files (SUFs) enthalten weniger Variablen und Ausprägungen als die Original-Files, um dadurch die Re-Identifizierung unwahrscheinlich zu machen. Darüber hinaus werden die SUFs durch juristische und organisatorische Maßnahmen, d. h. durch vertraglich festgelegte Ge- und Verbote der Verarbeitung, geschützt. Beispiele für sozialwissenschaftliche Studien, die in Form von Scientific Use Files der Wissenschaftsgemeinschaft zur Verfügung stehen, sind etwa ALLBUS, NEPS, PIAAC oder SOEP.⁴ Auch ein anderer und immer mehr an Bedeutung gewinnender Zugang zu sensiblen Mikrodaten basiert im Grunde auf faktisch anonymen Daten: In sogenannten Gastaufenthalten bei Forschungsdatenzentren werden anonymisierte Daten, die besonders schützenswert sind, der Forschung zur Verfügung gestellt. Hier werden die Daten vor allem durch organisatorische und technische Maßnahmen geschützt (neben dem juristischen Schutz).

In der gegenwärtigen Praxis des Umgangs mit empirischen Sozial- und Wirtschaftsdaten werden entsprechend dem Bundesdatenschutzgesetz (BDSG)

- (a) direkte Identifikatoren (Name, Adresse etc.) nach Abschluss des Datenaufbereitungsprozesses gelöscht, sofern keine Wiederholungsbefragung geplant ist. Ist eine wiederholte Befragung derselben Personen geplant (Panel- bzw. Längsschnitt-Ansatz), werden die Daten **pseudonymisiert**. Das heißt, die direkten Identifikatoren werden getrennt von den Daten aufbewahrt und nur bei einer erneuten Befragung wieder verwendet.
- (b) Für die Sekundäranalyse bzw. wissenschaftliche Nachnutzung werden die Daten entsprechend BDSG §3 (6) so **anonymisiert**, dass ein Personenbezug allenfalls mit einem **unverhältnismäßig hohen Aufwand** möglich ist (Prinzip der **faktischen Anonymität** basierend auf dem Prinzip der **relativen Anonymität**).

³ Mit Sekundäranalysen ist gemeint, dass im Rahmen eines Forschungsvorhabens keine eigene Erhebung durchgeführt wird, sondern auf bereits verfügbare Daten zurückgegriffen wird. Dies dient nicht nur dem ökonomischen Einsatz von (meist öffentlich geförderten) Forschungsressourcen, sondern ist auch für den Wissenschaftsdiskurs elementar. Dementsprechend ist die Sekundäranalyse bzw. die Nachnutzung von Forschungsdaten auch ein wichtiges, wissenschaftspolitisches Anliegen, das von Seiten der DFG wie auch vom BMBF starke Unterstützung erfährt.

⁴ Siehe www.ratswd.de/forschungsdaten/fdz

Bezogen auf diese langjährig bewährte Praxis ist die DSGVO problematisch:

- (1) Im Unterschied zum BDSG, in welchem im Kontext der allgemeinen Begriffsbestimmungen auch die faktische Anonymisierung (§3 (6)) definiert ist, fehlt eine solche Definition in der DSGVO.
- (2) Stattdessen findet sich im Erwägungsgrund 23 DSGVO eine indirekte Definition über den Begriff der Bestimmbarkeit einer Person.

„(23) Um festzustellen, ob eine Person bestimmbar ist, **sollten alle Mittel berücksichtigt werden, die von dem für die *Verarbeitung Verantwortlichen oder einer anderen Person nach allgemeinem Ermessen aller Voraussicht nach zum unmittelbaren oder mittelbaren Identifizieren oder Herausgreifen der Person genutzt werden.*** Bei der Prüfung der Frage, ob Mittel nach allgemeinem Ermessen aller Voraussicht nach zur Identifizierung der Person genutzt werden, sollten alle objektiven Faktoren, wie die Kosten der Identifizierung und der dafür erforderliche Zeitaufwand, herangezogen werden, **wobei sowohl die zum Zeitpunkt der *Verarbeitung verfügbare Technologie als auch die technologische Entwicklung* zu berücksichtigen sind.** (...)“ (DSGVO, Parlamentsentwurf)

- (3) Während aber dem BDSG das Konzept der *relativen* Anonymität (Unverhältnismäßigkeitskriterium) zugrunde liegt, ist das Anonymitätskonzept in der DSGVO nicht ausreichend bestimmt. Zwar ist im Parlamentsentwurf mit dem Verweis auf Kosten und Zeitaufwand einer Re-Identifizierung ein relatives Konzept angelegt, aber die fehlende Eindeutigkeit hat erhebliche Rechtsunsicherheit zur Folge. Es ist nicht ausgeschlossen, dass sich in der Rechtspraxis das Konzept der *absoluten* Anonymisierung durchsetzt. Dies legen folgende Formulierungen im Erwägungsgrund 23 nahe:

- Zu berücksichtigen sind alle Mittel einer für die **„*Verarbeitung Verantwortlichen oder einer anderen Person*“**: Bei einer engen Auslegung dieser Formulierung muss nicht nur das Wissen, der mit den Daten arbeitenden Personen berücksichtigt werden, sondern auch das allen anderen beliebigen Personen theoretisch zur Verfügung stehende Wissen.
- **„*nach allgemeinem Ermessen aller Voraussicht nach*“**: Bei dieser Formulierung handelt es sich um die Übersetzung des englischen Begriffs ‚reasonable‘. Hier ist nicht nachvollziehbar, weshalb sich das Adjektiv ‚reasonable‘, das in diesem Kontext sowohl in der aktuellen EU-Datenschutzrichtlinie wie auch in der DSGVO gebraucht wird, in der deutschen Version von ‚vernünftigerweise‘ (Richtlinie) zu der deutlich umfassenderen Formulierung **„*nach allgemeinem Ermessen aller Voraussicht nach*“** (DSGVO) wandelt.
- **„*sowohl die zum Zeitpunkt der Verarbeitung verfügbare Technologie als auch die technologische Entwicklung*“**: Dem hier formulierten Anspruch bei der Anonymisierung auch zukünftige Technologien zu berücksichtigen, kann in der Realität nicht entsprochen werden.

Absolut anonymisierte Daten sind für wissenschaftliche Analysen kaum geeignet, da ihr Informationsgehalt für fast alle Forschungsfragen nicht mehr ausreichend ist (so stehen entscheidende Variablen nicht zur Verfügung oder Merkmale werden zu stark vergrößert). Aus diesem Grunde war die sukzessive Ablösung des Konzepts der absoluten Anonymität durch das Konzept der relativen Anonymität im BStatG, BDSG und Sozialgesetzbuch ein ganz erheblicher Fortschritt für die empirische Forschung in Deutschland, ohne dass hierdurch die datenschutzwürdigen Belange der Bürgerinnen und Bürger beeinträchtigt worden wären. Vor diesem Hintergrund ist die fehlende Eindeutigkeit des Anonymitätskonzepts ein Problem und wird bei Rechtsunsicherheit zu einer empfindlichen Einschränkung der Sekundärdatenanalyse in Deutschland führen.

Sekundäranalyse dient nicht nur dem ökonomischen Einsatz von (meist öffentlich geförderten) Forschungsressourcen, sondern ist darüber hinaus für den Wissenschaftsdiskurs elementar. Insbesondere dort, wo empirische Befunde eine Informationsgrundlage für politische Entscheidungen bilden, ist es wichtig, die Daten mit unterschiedlichen theoretischen und methodischen Ansätzen zu analysieren. Dementsprechend ist die Sekundäranalyse bzw. die Nachnutzung von Forschungsdaten essentiell für die Wissenschaft an sich. Vor dem Hintergrund der weiter wachsenden Bedeutung von Sekundäranalysen kann eine enge Auslegung von Erwägungsgrund 23 [absolute Anonymität] dazu führen, dass wichtige Datenquellen für die Wissenschaft nicht mehr nutzbar sind und somit wissenschaftlicher Fortschritt und Erkenntnisgewinn behindert werden. Dies wiegt umso schwerer, als sich in Deutschland ein – durch den RatSWD initiiertes und geförderter – datenschutzgemäßer Zugang zu diesen Daten als Stütze wissenschaftlicher Forschung und der wissenschaftlichen Politikberatung etabliert hat.

Abschließend sei noch auf ein weiteres Problem hingewiesen. Sowohl im BDSG wie auch in der DSGVO findet sich eine explizite Forschungsklausel. Die im BDSG enthaltene Klausel bezieht sich dabei auf Daten, die explizit für Forschungszwecke erhoben wurden und hält explizit fest, dass diese Daten auch ausschließlich nur für Forschungszwecke verwendet werden dürfen (analog zum Statistikgeheimnis, das festlegt, dass Statistikdaten nicht für andere Zwecke verwendet werden dürfen). Durch eine Forschungsklausel sind zum Beispiel personenbeziehbare Forschungsdaten, die sich mit delinquenten Verhalten beschäftigen, vor einem staatlichen Zugriff geschützt. In der Forschungsklausel (Artikel 83) der DSGVO findet sich eine solche Zweckbindung nicht.

Die Bereitschaft an wissenschaftlichen Datenerhebungen teilzunehmen und die Erlaubnis zur Verwendung von Sekundärinformationen basiert auf einem Vertrauensverhältnis zwischen Forschenden und Beforschten. Die fehlende verbindliche Zusage, dass Daten nur für Forschungszwecke verwendet werden und von anderen Bereichen abgeschottet werden, könnte sich negativ auf die Forschung auswirken.

Empfehlungen und Hinweise des RatSWD zur EU-DSGVO

- (1) Der RatSWD hält es für notwendig, dass in der DSGVO das Konzept des relativen Personenbezugs in eindeutiger Weise anerkannt wird.
- (2) Der RatSWD empfiehlt, dass bei den Begriffsbestimmungen eine Definition von „Anonymität“ aufgenommen wird, wie sie in der Europäischen Statistikverordnung bereits realisiert ist und wo sie sich bewährt hat (VERORDNUNG (EG) Nr. 223/2009).

Alternativ wäre im Hinblick auf Erwägungsgrund 23 der Kommissionsentwurf mit leichten Modifikationen (s. u.) dem Parlamentsentwurf vorzuziehen:

Um festzustellen, ob eine Person bestimmbar ist, sollten alle Mittel berücksichtigt werden, die vernünftigerweise entweder von dem für die Verarbeitung Verantwortlichen oder von einem Dritten eingesetzt werden könnten, um die betreffende Person zu bestimmen.

Der für die Wissenschaft wichtige Vorteil der o. a. Formulierung ergibt sich daraus, dass diese Definition von Anonymität bereits Praxis ist und sich bewährt hat. Darüber hinaus wäre diese Definition weitgehend in Analogie zum Europäischen Statistikgesetz (EU-Verordnung Nr. 223/2009),⁵ so dass nicht nach Datentyp unterschiedliche Anonymisierungsrichtlinien zur Anwendung kommen werden.

- (3) Der RatSWD empfiehlt, dass in der DSGVO in Artikel 4 eine klare Abgrenzung zwischen den Konzepten von personenbezogenen, pseudonymen und anonymen Daten aufgenommen wird. DSGVO, Artikel 4 (2a) Parlamentsentwurf definiert ‚pseudonyme Daten‘ wie folgt:

(2a) „pseudonymisierte Daten“ personenbezogene Daten, die ohne Heranziehung zusätzlicher Informationen keiner spezifischen betroffenen Person zugeordnet werden können, sofern diese zusätzlichen Informationen gesondert aufbewahrt werden und technischen und organisatorischen Maßnahmen unterliegen, die die Nichtzuordnung gewährleisten;“

- Da pseudonyme Daten als personenbezogene Daten definiert sind, gelten nach Auffassung des RatSWD die in Artikel 5 DSGVO festgelegten Grundsätze für die Verarbeitung personenbezogener Daten auch für pseudonymisierte Daten.

Bei einer sehr engen Auslegung der DSGVO impliziert dies, dass die Nutzung von pseudonymisierten Daten durch Drittforscher oder eine Veränderung in den Forschungsinhalten den betroffenen Personen mitgeteilt bzw. gegebenenfalls auch die Einwilligung eingeholt werden muss.

- Weiterhin ist die Abgrenzung zu anonymen Daten uneindeutig, da auch anonyme Daten „ohne Heranziehung zusätzlicher Informationen keiner spezifischen betroffenen Person zugeordnet werden können“.

- (4) Der RatSWD empfiehlt, dass in die Forschungsklausel (Artikel 83) der DSGVO eine eindeutige Zweckbindung eingefügt wird. Forschungsdaten dürfen nur der Forschung dienen und keinerlei anderen Zwecken.

⁵ „Bei der Entscheidung, ob eine statistische Einheit identifizierbar ist, sind alle Mittel zu berücksichtigen, die nach vernünftigem Ermessen von einem Dritten angewendet werden könnten, um die statistische Einheit zu identifizieren“ VERORDNUNG (EG) Nr. 223/2009; Artikel 3 (7).