Stellungnahme im öffentlichen Konsultationsverfahren des BfDI zum Thema „Anonymisierung unter der DSGVO unter besonderer Berücksichtigung der TK-Branche“

Christian Burkert, Hannes Federrath, Matthias Marx, Monina Schwarz

Universität Hamburg, FB Informatik, Arbeitsbereich Sicherheit in verteilten Systemen
https://svs.informatik.uni-hamburg.de

20.03.2020

1 Einleitung

Es ist ausdrücklich zu begrüßen, dass der in der EU-Datenschutzgrundverordnung (DSGVO) nicht mehr näher ausgestaltete Begriff der Anonymisierung im Rahmen eines Konsultationsverfahrens diskutiert wird. Im Folgenden gehen wir auf das Positionspapier [1] des BfDI vom 10. Februar 2020 ein, welches als Diskussionsgrundlage zu den Anforderungen an die Anonymisierung dient und die Anonymisierung als Alternative zur Löschung diskutiert.

In dem Terminologie-Papier [2] aus dem Jahre 2010 haben Pfitzmann und Hansen bereits versucht, die technischen und die rechtlichen Aspekte des Begriffs Anonymität möglichst präzise zu fassen. Dort wird Anonymität allgemein definiert als **Eigenschaft eines Subjekts**, innerhalb einer Menge von Subjekten, der sog. Anonymitätsgruppe, nicht identifizierbar zu sein. Bei dieser Definition spielt es zunächst keine Rolle, ob es sich um Daten in einer Datenbank handelt oder um Verbindungs- bzw. Metadaten, die während einer Kommunikationsverbindung entstehen und typischerweise in Protokolldateien gespeichert werden.

Zur zuverlässigen Entfernung des Personenbezugs müssen mitunter mehrere Methoden miteinander kombiniert werden und – je nach Verwendungszweck der anonymisierten Daten – auf *alle* Attribute angewendet werden, d.h. es werden im Prozess der Anonymisierung

- bestimmte **Attribute** eines Datensatzes **weggelassen** bzw. entfernt (Datenminimierung),
- mehrere **Einzelwerte** zu einem gruppierten Wert (z.B. Durchschnitt, Median, Wertebereich) **zusammengefasst** (Generalisieren bzw. Aggregieren),
- Einzelwerte durch Hinzufügen von indeterministischem Rauschen jeweils **ungenau gemacht** (Perturbation).

Darüber hinaus können mittels eines Algorithmus (oder auch einer Zuordnungsregel) die **Attribute** eines Datensatzes **durch einen anderen eindeutigen Wert ersetzt** (Pseudonymisierung, Verschlüsselung) werden, was jedoch zumindest bei Kenntnis der Zuordnungsregel (z.B. Hashverfahren) bzw. seiner Ersetzungsparameter (z.B. kryptographischer Schlüssel, Seed, Salt) zu keiner wirksamen Anonymisierung führen wird. Insoweit sind die Pseudonymisierung oder die Verschlüsselung **keine geeigneten Mechanismen zur Anonymisierung**.

2 Anforderungen an die Anonymisierung

Maßgeblich für eine Beurteilung der Identifizierbarkeit einer Person und damit im Umkehrschluss auch der Anonymisierung ist nach Erwägungsgrund 26 Satz 3 der DSGVO die Berücksichtigung **aller Mittel**, die von **dem Verantwortlichen oder einer anderen Person** nach **allgemeinem Ermessen wahrscheinlich** genutzt werden. Erwägungsgrund 26 Satz 4 der DSGVO konkretisiert die wahrscheinliche Nutzung wie folgt:

> „Bei der Feststellung, ob Mittel nach allgemeinem Ermessen wahrscheinlich zur Identifizierung der natürlichen Person genutzt werden, sollten alle objektiven Faktoren, wie die **Kosten der Identifizierung** und der dafür **erforderliche Zeitaufwand**, herangezogen werden, wobei die **zum Zeitpunkt der Verarbeitung** verfügbare Technologie und technologische Entwicklungen zu berücksichtigen sind."

Diese Maßgaben der DSGVO genügen in der Praxis nicht, um Verfahren zu Anonymisierung objektiv und einheitlich zu bewerten, da

a) das objektive Treffen von Annahmen über wahrscheinlich genutzte Mittel Dritter und

b) die Antizipation zukünftiger technischer Entwicklungen

wesentliche Einflussgrößen des Ermessens sind, die individuelle Verantwortliche überfordern und Spielraum für missbräuchliche Auslegungen lassen. Nachfolgend werden beide Aspekte betrachtet.

2.1 Ermessen wahrscheinlich genutzter Mittel Dritter

Mittel der Re-Identifikation können sowohl fortschreitende technologische Möglichkeiten (etwa Big-Data-Verfahren) als auch andere Daten (z.B. Kontextinformationen, die man über Auskunfts- und Einsichtsrechte erlangen kann) sein, die mit den zu bewertenden Daten verknüpft werden können, um aus deren Kombination eine Re-Identifikation zu ermöglichen.

Wenn gemäß Erwägungsgrund 26 Satz 3 der DSGVO nicht nur wahrscheinlich genutzte Mittel des Verantwortlichen zur Beurteilung herangezogen werden müssen, sondern auch alle Mittel, die von anderen Personen (Dritten) „nach allgemeinem Ermessen wahrscheinlich genutzt werden", dann muss man im Allgemeinen davon ausgehen, dass genug Ressourcen zur Verfügung stehen, um einen Datensatz zumindest teilweise re-identifizieren zu können. Jedenfalls dürfen die Verantwortlichen das Risiko einer Re-Identifizierung nicht zu optimistisch einschätzen. Immer wieder zeigt sich, dass vermeintlich anonymisierte Datensätze unter Hinzuziehung von (öffentlichen) Kontextinformationen re-identifizierbar [3] sind bzw. sensible Betriebs- und Geschäftsgeheimnisse offenlegen [4] können. So zeigte etwa Sweeney, dass ein vermeintlich anonymes medizinisches Register, das jeweils nur das Geschlecht, das Geburtsdatum und die Postleitzahl einer Person enthält, in den USA geeignet ist, ca. 87 Prozent der US-Bevölkerung zu re-identifizieren, indem Daten mit den dort öffentlich zugänglichen Wählerverzeichnissen verknüpft [5] wurden.

Im Positionspapier des BfDI wird etwa auf Seite 10 zu „(2) § 98 Abs. 1 TKG" ausgeführt, dass gemäß § 98 Abs. 1 TKG „Standortdaten der Verarbeitung in Form einer Anonymisierung zugeführt werden dürfen". Selbst wenn es sich hier um aggregierte Einzeldaten handelt, also im Zuge der Anonymisierung nur noch ungenaue Daten über die konkreten Zeitpunkte und/oder die genauen Aufenthaltsorte einzelner Nutzer (z.B. durch Perturbation oder Aggregation), darf dennoch davon ausgegangen werden, dass diese Daten weiterhin einen Personenbezug haben und sie somit keine anonymisierten Daten darstellen:

Wie bei dynamisch veränderlichen Einzeldaten stehen auch noch die (perturbierten oder aggregierten) Einzelereignisse eines Nutzers in einem funktionalen Zusammenhang zueinander, der es

in vielen Fällen erlauben wird, aus den ungenau gemachten Daten die Einzeldaten annähernd zu rekonstruieren. Wenn beispielsweise die Geokoordinaten zweier unterschiedlich schneller und in unterschiedliche Richtungen fahrender PKWs durch Anonymisierung zusammengefasst werden, könnten dennoch aus der zeitlichen Abfolge dieser anonymen Datenpunkte die Bewegungsvektoren der beiden Fahrzeuge mit hoher Wahrscheinlichkeit rekonstruiert werden.

Hinzu kommt, dass gerade vor dem Hintergrund der Verkettbarkeit von geographischen Einzeldaten der Telekommunikationsdienstleister mit den etwa bei den großen Technologiekonzernen vorhandenen Daten, die Geokoordinaten nach unserer Auffassung konsequent als personenbeziehbar einzustufen sind und somit im Anwendungsbereich der DSGVO verbleiben sollten.

2.1.1 Absolute Anonymisierung

Für eine **absolute**, also eine unter allen Umständen unumkehrbare Anonymisierung dürfen die anonymisierten Daten

a) entweder keinen Informationsgehalt mehr haben

b) oder ihr verbleibender Informationsgehalt lässt eine Re-Identifizierung tatsächlich nicht mehr zu.

Hierbei sind die unterstellten verfügbaren Mittel des Angreifers (fortschreitende technologische Möglichkeiten, Kontextinformationen und ggf. zur Verfügung stehende Zeit zur Re-Identifizierung) **im Rahmen einer Datenschutzfolgenabschätzung** konkret zu benennen.

Das Positionspapier des BfDI stellt richtigerweise heraus, dass eine absolute, also eine unter allen Umständen unumkehrbare Anonymisierung häufig nicht möglich ist.

2.1.2 Praktische Anonymisierung

Für eine **praktische**, also unter Berücksichtigung der verfügbaren Mittel des Angreifers durchgeführte Anonymisierung genügt es nicht, wenn der Verantwortliche für das jeweils vorgeschlagene Anonymisierungsverfahren pauschal annimmt, dass der Personenbezug nur mit unverhältnismäßigem Aufwand wiederhergestellt werden kann. Vielmehr ist es notwendig, die unterstellten verfügbaren Mittel des Angreifers (Technik, Kontextinformationen, Zeit) **im Rahmen einer Datenschutzfolgenabschätzung** konkret zu benennen.

Seitens der Aufsichtsbehörden sollten konkrete

a) Vorgaben zur Konkretisierung der praktischen Nicht-Durchführbarkeit einer Re-Identifizierung sowie

b) Vorgaben zur Objektivierung und Vereinheitlichung der wahrscheinlichen Mittel des Angreifers

entwickelt werden. Darüber hinaus ist es notwendig, den Unverhältnismäßigkeitsbegriff konkret auszugestalten.

Hierbei muss die Frage beantwortet werden, wie ein Verantwortlicher objektiv die wahrscheinlich genutzten Mittel einer **anderen Person** ermessen und angemessen berücksichtigen kann. Dieses Ermessen kann und sollte nicht einzelnen Verantwortlichen überlassen bleiben, um eine bewusst oder unbewusst enge Auslegung der wahrscheinlich genutzten Mittel zulasten des Anonymisierungsniveaus zu vermeiden.

Weil eine praktische Anonymisierung keinen hundertprozentigen Schutz vor Re-Identifizierung bieten kann, bezeichnet wir sie auch als imperfekte Anonymisierung.

2.2 Antizipation zukünftiger technischer Entwicklung: Anonymisierung als untaugliches Konzept

In der Literatur wird die Anonymisierung als Schutzkonzept und als Ausweg aus dem Personenbezug aufgrund der gewachsenen Möglichkeiten von Datenanalysen mit Hilfe von „Big Data“ und künstlicher Intelligenz zum Teil gänzlich in Frage gestellt. Hoffmann-Riem [6] etwa stellt fest, es könne „nicht mehr daran festgehalten werden, dass der Personenbezug stets schon durch Anonymisierung von Daten entfällt.“

Wir weisen darauf hin, dass die Möglichkeiten der Re-Identifizierung und Ableitung von Eigenschaften nicht unterschätzt werden dürfen. Solange noch Einzeldaten vorhanden sind, muss davon ausgegangen werden, dass diese auch weiterhin personenbeziehbar sein können. Zumindest ist die Wahrscheinlichkeit einer Re-Identifizierung niemals Null.

Bezugnehmend auf Erwägungsgrund 26 Satz 4 der DSGVO halten wir es für dringend erforderlich, bei der Beurteilung der Anonymisierung nicht bloß auf aktuell mögliche und wahrscheinlich genutzte Technologien abzustellen, sondern auch zukünftige technische Entwicklungen zu antizipieren. Andernfalls könnten – wegen einer zu engen Auslegung der wahrscheinlich genutzten Mittel – vermeintlich anonymisierte Daten innerhalb weniger Jahre re-identifiziert und personenbeziehbar weiterverarbeitet werden.

In diesem Sinne sollte gerade bei der Übermittlung anonymisierter Daten besondere Rücksicht auf zukünftige Entwicklungen genommen werden. Hier genügt es aus unserer Sicht auch nicht, Anonymisierungsverfahren regelmäßig einer kritischen Aktualitätsprüfung zu unterziehen. Vielmehr müssen alle in der Vergangenheit anonymisierten Daten vom Verantwortlichen regelmäßig, etwa im Rahmen der **Fortschreibung der Datenschutzfolgenabschätzung**, auf ihre Re-Identifizierbarkeit geprüft und ggf. nachanonymisiert werden. Dabei ist jedoch zu beachten, dass einmal verloren gegangene Vertraulichkeit aufgrund der Monotonieeigenschaften [7] der Schutzziele der mehrseitigen IT-Sicherheit niemals wieder zurückgeholt werden kann.

Die Anonymitätseigenschaft von Daten sollte von den Verantwortlichen auch deshalb überwacht werden, weil durch eine wiedererlangte Identifizierbarkeit der Personenbezug wieder hergestellt sein könnte und damit die DSGVO erneut Anwendung fände. Wir schlagen daher in Anlehnung an Hoffmann-Riem vor, auch anonymisierte Daten weiterhin als personenbeziehbare Daten und damit personenbezogene Daten i.S.d. Art. 4 Nr. 1 DSGVO zu behandeln und zu schützen, wenngleich der Schutzbedarf geringer ausfallen wird als bei den personenbezogenen Daten, bei denen der Personenbezug offensichtlich ist.

3 Anonymisierung untauglich als Verfahren zur Umsetzung der Löschung

Das Positionspapier des BfDI kommt auf Seite 3 zu der optimistischen Feststellung, dass „eine Verpflichtung zur unverzüglichen Löschung [...] durch eine Anonymisierung erfüllbar“ ist. Unter Berücksichtigung der zukünftigen technischen Möglichkeiten zur Re-Identifizierung dürfte somit eine Löschverpflichtung bestenfalls einer **praktischen Anonymisierung** gleichkommen. Legt man die o.a. strengen Maßstäbe an eine Risikobeurteilung an, d.h. Durchführung einer Datenschutzfolgenabschätzung, fortlaufende Risikobeurteilung und Fortschreibung der Datenschutzfolgenabschätzung, führt die Anonymisierung regelmäßig zu einem hohen Aufwand der Verantwortlichen. Währenddessen ist eine (echte) Löschung von Daten vergleichsweise kostengünstig umsetzbar.

Insgesamt können wir der Einschätzung, dass Löschung und Anonymisierung vergleichbare Risiken für die Re-Identifizierbarkeit haben, nicht zustimmen. Selbst wenn man der Argumentation des Positi-

onspapiers des BfDI folgt, die DSGVO mache eine bewusste Unterscheidung zwischen Löschung und Vernichtung, und daraus schwächere Ansprüche an die Endgültigkeit der Löschung ableitet, folgt daraus nicht, dass die Anonymisierung als Verfahren für die ebenfalls potentiell imperfekte Entfernung des Personenbezugs der Löschung gleichzustellen ist. Ein wesentlicher Unterschied liegt darin, dass aus einer imperfekten Löschung keine Daten resultieren, die beliebig übermittelt werden können und dem Anwendungsbereich der DSGVO entzogen sind: Bei einer Anonymisierung – anstelle der Löschung – verbleiben Daten, die vermeintlich nicht mehr personenbeziehbar sind und somit (zunächst) nicht mehr den Regelungen der DSGVO unterliegen, was wie o.a. trügerisch sein kann.

Aus der Annahme, dass das Gebot der Löschung gewissen Verhältnismäßigkeitsschranken unterliegt, kann zwar gefolgert werden, dass die Löschung nicht auf allen Kopien (wie etwa Backups) erfolgen muss, sofern dies unverhältnismäßig hohen Aufwand verursachen würde, andererseits folgt daraus aber nicht, dass die Datenträger dieser nicht gelöschten Kopien frei von personenbezogenen Daten wären und übermittelt werden dürften. Vielmehr erstreckt sich die Zumutbarkeitsschranke nur auf den alltäglichen Prozess der Löschung. Die Auslistung oder Entsorgung der (gelöschten) Speichermedien würde weiterhin den Bestimmungen für personenbezogene Daten unterliegen. Anders stellt sich die Situation bei der praktischen Anonymisierung (siehe Abschn. 2.1.2) dar, die ebenfalls nur imperfekt den Personenbezug aus den Daten entfernt. Hier entstehen allerdings verkehrsfähige Daten, die unkontrolliert übermittelt werden könnten.

Zum Zeitpunkt des Löschens mag eine Anonymisierung aus Sicht des Verantwortlichen eine äquivalente Unzuordenbarkeit der Daten zu Personen bedeuten und daher dem Effekt einer Löschung gleich kommen. Jedoch kann etwa durch einen anderen Kontext, eine weitere Datenquelle oder eine zeitbedingte Abschwächung des Anonymisierungsalgorithmus der Personenbezug wiederhergestellt werden. [8]

Im Positionspapier des BfDI wird beispielsweise auf Seite 10f zu „(3) § 96 Abs. 1 S. 2 Alt. 2/Art. 6 Abs. 1 Buchst. c) DSGVO i.V.m. § 96 Abs. 1 S. 3 TKG“ ausgeführt, dass „die personenbezogenen Daten auch durch deren Anonymisierung gelöscht werden können“. Hier geht es um Verkehrs- bzw. Metadaten, die der Telekommunikationsdienstleister nach Beendigung der Verbindung unverzüglich zu löschen hat. In der Praxis genügen mit hoher Wahrscheinlichkeit die konkreten (minuten- oder sekundengenauen) Zeitstempel von Beginn und Ende einer Verbindung, um einen Personenbezug mit Hilfe von Kontextinformationen herzustellen. Insoweit wird es sich bei einer (praktischen) Anonymisierung selten bis nie um eine zum Löschen äquivalente Möglichkeit handeln.

Unter der Annahme, eine perfekte Anonymisierung leisten zu können, könnte man eine Anonymisierung tatsächlich mit einer Löschung gleichsetzen. Da in der Praxis jedoch nur eine Anonymisierung in Bezug auf bestimmte verfügbare Mittel (Technik, Kontextinformationen, Zeit) erreicht werden kann, sind die Löschung und die Anonymisierung hinsichtlich der Risiken für die Betroffenen nicht vergleichbar. Die Anonymisierung kann daher nur in wenigen Fällen als gleichwertigen Ersatz für die Löschung angesehen werden.

4 Schlussbemerkungen

Aus wissenschaftlicher Sicht möchten wir davor warnen, die Möglichkeiten einer Anonymisierung von Datensätzen (etwa aus dem medizinischen Bereich) oder von Protokolldaten (etwa aus dem Telekommunikationsbereich) zu optimistisch zu betrachten. Das einfache Entfernen von offensichtlich personenbezogenen Attributen eines Datensatzes genügt regelmäßig allein nicht, um Daten zu anonymisieren.

Vielmehr lässt sich – zumeist durch Hinzunahme von Kontextinformationen – aus den vermeintlich

nicht unmittelbar personenbeziehbaren Attributen eines Datensatzes aufgrund von eindeutigen Merkmalskombinationen zumindest für einzelne Datensätze häufig der Personenbezug mit einer hohen Wahrscheinlichkeit wieder herstellen.

Praktische Anonymisierung (siehe Abschn. 2.1.2) wird zumeist nicht wirkungsvoll genug sein, um die entstandenen Daten dem Anwendungsbereich der DSGVO zu entziehen, da nicht davon ausgegangen werden kann, dass die Personenbeziehbarkeit entfällt. Nur die absolute Anonymisierung (siehe Abschn. 2.1.1) entspricht der wirksamen und unumkehrbaren Entfernung des Personenbezugs, so dass damit auch der Anwendungsbereich der DSGVO nicht mehr eröffnet ist.

Literatur

[1] Der Bundesbeauftragte für den Datenschutz und die Informationsfreiheit: Anonymisierung unter der DSGVO unter besonderer Berücksichtigung der TK-Branche. Positionspapier im Konsultationsverfahren des BfDI, 10. Februar 2020. https://www.bfdi.bund.de/SharedDocs/Konsultationsverfahren/2020/01_Anonymisierung-TK.pdf?__blob=publicationFile&v=6 (letzter Abruf am 06.03.2020)

[2] Andreas Pfitzmann, Marit Hansen: A terminology for talking about privacy by data minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management. Version v0.34, August 2010. https://dud.inf.tu-dresden.de/Anon_Terminology.shtml (letzter Abruf am 06.03.2020)

[3] J. K. Trotter: Public NYC Taxicab Database Lets You See How Celebrities Tip. 23. Oktober 2014. https://gawker.com/the-public-nyc-taxicab-database-that-accidentally-track-1646724546 (letzter Abruf am 06.03.2020)

[4] The Guardian: Pentagon to review security after Strava reveals sensitive information. 29. Januar 2019. https://www.theguardian.com/us-news/2018/jan/29/pentagon-strava-fitness-security-us-military (letzter Abruf am 06.03.2020)

[5] Latanya Sweeney: Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000. https://dataprivacylab.org/projects/identifiability/paper1.pdf (letzter Abruf am 06.03.2020)

[6] Wolfgang Hoffmann-Riem: Big Data - Regulative Herausforderungen. Nomos 2018, S. 56

[7] Gritta Wolf, Andreas Pfitzmann: Empowering Users to Set their Protection Goals. In: Günter Müller, Kai Rannenberg (Ed.): Multilateral Security in Communications, Addison-Wesley-Longman, 1999, S. 113-136

[8] Dániel Kondor, Behrooz Hashemian, Yves-Alexandre Montjoye, Carlo Ratti: Towards matching user mobility traces in large-scale datasets. IEEE Transactions on Big Data. (2017) DOI: 10.1109/TBDATA.2018.2871693