Sarah Stummer*

Personenbezogenheit vs. Anonymität

Ein Mapping des rechtlichen und technischen Begriffsverständnisses von „Personenbezogenheit", „Pseudonymität" und „Anonymität"

Um die Anonymisierung praktisch anwendbar zu machen und konkrete Anforderungen an Anonymität stellen zu können, sollte der Themenkomplex von einem interdisziplinären Standpunkt aus betrachtet werden. Hierbei gilt es bereits bei den grundlegenden Begriffen der Personenbezogenheit und Anonymität anzusetzen und diese interdisziplinär zu vereinheitlichen.

1 Problemstellung

Obwohl die Anonymisierung personenbezogener Daten überall dort, wo kein Personenbezug erforderlich ist, einen Mehrwert schaffen kann, werden deren Chancen derzeit noch nicht ausgeschöpft. Grund hierfür sind zahlreiche (vorwiegend rechtliche) Unsicherheiten im Zusammenhang mit dem Themenkomplex der Anonymisierung.[1]

Da der Themenkomplex dabei sowohl rechtlicher als auch technischer Natur ist, bedarf es zur Auflösung dieser Unsicherheiten eines interdisziplinären Ansatzes, welcher nicht nur die rechtlichen Anforderungen, sondern auch die technischen Möglichkeiten berücksichtigt. Dabei sollte bereits beim Fundament des Themenkomplexes – der Abgrenzung zwischen den Datenzuständen der Personenbezogenheit (inklusive der Pseudonymität) sowie der Anonymität[2] – angesetzt werden, indem deren rechtliches und technisches Begriffsverständnis analysiert und zu einem interdisziplinären Begriffsverständnis des Rechts und der Technik vereint werden.

* Dieser Beitrag wurde vom Bundesministerium für Bildung und Forschung (BMBF) und vom Hessischen Ministerium für Wissenschaft und Kunst (HMWK) im Rahmen ihrer gemeinsamen Förderung für das Nationale Forschungszentrum für angewandte Cybersicherheit ATHENE unterstützt.

1 *Stummer*, INFORMATIK 2022, S. 179-194.

2 *Finck/Pallas*, IDPL 2020, 11 (11 f.).

Sarah Stummer (LL.M.)

ist wissenschaftliche Mitarbeiterin der Abteilung IT Law & Interdisciplinary Privacy Research am Fraunhofer-Institut für Sichere Informationstechnologie (SIT) und am Nationalen Forschungszentrum für angewandte Cybersicherheit ATHENE.
E-Mail: sarah.stummer@sit.fraunhofer.de
https://orcid.org/0000-0003-4015-4429

2 Quellen des rechtlichen und technischen Begriffsverständnisses

Im Datenschutzrecht wird zwischen anonymen und personenbezogenen Daten unterschieden. Hintergrund hierfür ist der sachliche Anwendungsbereich des Datenschutzrechts, welcher an die Verarbeitung personenbezogener Daten anknüpft und anonyme Informationen vom Anwendungsbereich ausnimmt. Zusätzlich kennt das Datenschutzrecht den Vorgang der Pseudonymisierung. All diese Begriffe werden in der DSGVO adressiert und sind insoweit unmittelbar durch die rechtlichen Vorgaben geprägt.

Technische Begriffsbestimmungen werden hingegen vor allem durch (internationale) Standards und wissenschaftliche Publikationen geprägt. Da es üblich ist, dass diese lediglich bestimmte Branchen oder Anwendungsbereiche adressieren, gilt es zur Ermittlung und Analyse der technischen Begriffsbestimmungen international anerkannte sowie möglichst branchenunabhängige Definitionen zur Grundlage zu nehmen. Hierfür bieten sich vor allem internationale Standards der Internationalen Organisation für Standardisierung (ISO) an. Dabei sind vor dem Hintergrund des Anwendungsbereichs der ISO-Standards insbesondere die Definitionen aus der ISO/IEC 29100:2011 sowie der ISO/IEC 20889:2018 relevant. Bei der ISO/IEC 29100:2011 handelt es sich um ein Rahmenwerk zum Datenschutz, welches u.a. bezweckt, die einschlägige Terminologie zu definieren. Die ISO/IEC 20889:2018 enthält eine Beschreibung von De-Identifizierungtechniken und spezifiziert in diesem Zusammenhang die einschlägige Terminologie.

3 Personenbezogene Daten

3.1 Rechtliches Begriffsverständnis

Der Begriff der personenbezogenen Daten ist in Art. 4 Nr. 1 DSGVO als „*alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person beziehen*" definiert. Der Be-

griff besteht folglich aus den vier Elementen Informationen, Personenbezug („über"), identifiziert bzw. identifizierbar und natürliche Person.[3]

Der Begriff der Information ist weit auszulegen.[4] Er umfasst nicht nur objektive, nachprüfbare Informationen, wie das Geburtsdatum, den Personenstand oder die Steuernummer einer betroffenen Person, sondern auch subjektive Informationen, wie Werturteile, Einschätzungen und Meinungsäußerungen.[5] Bezieht sich eine solche Information insoweit auf eine Person, als sie unmittelbar Aussagen über die Person trifft oder Rückschlüsse auf sie ermöglicht, weist die Information einen Personenbezug auf.[6]

Die personenbezogenen Informationen müssen sich dabei auf eine lebende[7] identifizierte oder identifizierbare Person beziehen – die betroffene Person. Als identifizierbar gilt eine natürliche Person nach Art. 4 Nr. 1 DSGVO, wenn sie „*direkt oder indirekt, insbesondere mittels Zuordnung zu einer Kennung […] oder zu einem oder mehreren besonderen Merkmalen […] identifiziert werden kann*". Hieraus folgert die herrschende Literaturmeinung, dass eine natürliche Person identifizierbar ist, wenn die vorliegenden Informationen allein nicht ausreichen, um sie zu identifizieren, dies jedoch durch die Verknüpfung der Informationen mit weiteren Informationen gelingt.[8] Als identifiziert gilt eine natürliche Person nach der in der rechtlichen Literatur vorherrschenden Meinung hingegen, wenn deren Identität unmittelbar aus den vorliegenden Informationen hervorgeht.[9] Als Beispiele für Informationen, mit denen eine natürliche Person identifiziert ist, werden daher regelmäßig der Name[10] sowie einzigartige Informationen wie die Steuer-Identifikationsnummer,[11] die Sozialversicherungsnummer[12] oder der Fingerabdruck[13] angeführt.

3.2 Technisches Begriffsverständnis

Die ISO/IEC 29100:2011 definiert den Begriff der personenbezogenen Daten als *"any information that (a) can be used to identify the PII principal to whom such information relates, or (b) is or might be directly or indirectly linked to a PII principal."*[14]

Demnach kann grundsätzlich jede Information ein personenbezogenes Datum darstellen, sofern die Information (a) zur Identifizierung des sogenannten PII principals verwendet werden kann, oder (b) direkt oder indirekt mit dem PII principal verknüpft ist oder werden kann. Der *PII principal* wird dabei als die natürliche Person definiert, auf die sich die personenbezogenen Daten beziehen. Er ist insoweit synonym zur betroffenen Person aus der DSGVO zu verstehen.[15]

Der in Alternative (a) verwendete Begriff der Identifizierung bedeutet gemäß der ISO/IEC 29100:2011 eine Verbindung zwischen personenbezogenen Daten und der betroffenen Person zu schaffen.[16] Folglich umfasst Alternative (a) der Begriffsbestimmung die Möglichkeit der Identifizierung einer betroffenen Person. Dabei reicht es, wie aus der Formulierung *can be used* folgt, aus, dass die Informationen zur Identifizierung genutzt werden können. Es ist nicht erforderlich, ebenso wenig jedoch schädlich,[17] dass die Informationen bereits zur Identifizierung genutzt wurden.

So wie Alternative (a) adressiert auch Alternative (b) der Begriffsbestimmung der ISO/IEC 29100:2011 die Verbindung zwischen Informationen und betroffener Person. Während Alternative (a) jedoch auf die Identifizierung, also auf die Zuordnung von Informationen zur betroffenen Person, abzielt, zielt Alternative (b) auf das Vorliegen einer Verbindung zwischen der Information und der betroffenen Person ab. Dabei muss die Information (anders als bei Alternative (a)) nicht zwingend dazu geeignet sein die betroffene Person zu identifizieren. Vielmehr reicht es aus, dass die Information mit einer identifizierten oder identifizierbaren betroffenen Person verbunden ist. Zwar nimmt Alternative (b) innerhalb der Begriffsbestimmung nicht ausdrücklich Bezug darauf, dass die betroffene Person identifiziert oder identifizierbar sein muss, aus der ISO/IEC 29100:2011 geht jedoch hervor, dass solche Informationen gemeint sind, die mit der Identität einer Person verbunden sind („*is linked*") oder verbunden werden können („*might be linked*").[18] Folglich muss die betroffene Person mindestens identifizierbar sein. Alternative (b) umfasst somit auch (regelmäßig) nicht identifizierende Informationen, wie z.B. „guter Koch", „Sportart: Leichtathletik" oder „Abschluss in Wirtschaftsingenieurwesen", soweit sie mit identifizierenden Informationen, wie dem Namen, dem Portraitbild oder einer Kombination aus Geburtsdatum, Geschlecht und Postleitzahl,[19] kombiniert sind. Unerheblich ist dabei, ob die Verbindung direkt oder indirekt vorliegt.

3.3 Übereinstimmung

Wie die Analyse des rechtlichen und technischen Begriffsverständnisses zeigt, kennen grundsätzlich sowohl das Recht als auch die Technik den Begriff der personenbezogenen Daten.

Fraglich ist jedoch, inwieweit das Begriffsverständnis innerhalb der beiden Fachdisziplinen semantisch übereinstimmen. Hierfür gilt es zu prüfen, ob und inwieweit sich die aus dem rechtlichen Begriffsverständnis bekannten vier Elemente (Informationen, Personenbezug, identifiziert bzw. identifizierbar und natür-

3 Zur DSRL: *Art. 29-Datenschutzgruppe:* WP 136, S. 6; übertragen auf die DSGVO: *Klabunde* in Ehmann/Selmayr, Art. 4 DSGVO Rn. 8; *Finck/Pallas,* IDPL 2020, 11 (12); *Diel/Schreiber* in Selzer, Art. 4 DSGVO Rn. 2.

4 *EuGH,* NJW 2018, 767 (767 f.); *Ernst* in Paal/Pauly, Art. 4 DSGVO Rn. 3; *Klar/Kühling* in Kühling/Buchner, Art. 4 Nr. 1 DSGVO Rn. 8.

5 *Gola* in Gola/Heckmann, Art. 4 DSGVO Rn. 6; *Klabunde* in Ehmann/Selmayr, Art. 4 DSGVO Rn. 9.

6 *Spindler/Dalby* in Spindler/Schuster, Art. 4 DSGVO Rn. 6; *Diel/Schreiber* in Selzer, Art. 4 DSGVO Rn. 2.

7 Erwägungsgrund 27 DSGVO.

8 *Arning/Rothkegel* in Taeger/Gabel, Art. 4 DSGVO Rn. 30; *Klar/Kühling* in Kühling/Buchner, Art. 4 DSGVO Rn. 19; *Ernst* in Paal/Pauly, Art. 4 DSGVO Rn. 8; *Karg* in Simitis/Hornung/Spiecker gen. Döhmann, Art. 4 DSGVO Rn. 57.

9 *Ziebarth* in Sydow/Marsch, Art. 4 DSGVO Rn. 14; *Klar/Kühling* in Kühling/Buchner, Art. 4 Nr. 1 DSGVO Rn. 18; *Borges* in Borges/Hilber, Art. 4 DSGVO Rn. 11; *Arning/Rothkegel* in Taeger/Gabel, Art. 4 DSGVO Rn. 25; *Karg* in Simitis/Hornung/Spiecker gen. Döhmann, Art. 4 Nr. 1 DSGVO Rn. 49.

10 *Art. 29-Datenschutzgruppe*, WP 136, S. 15; *Schild* in Wolff/Brink, Art. 4 DSGVO Rdnr, 16.

11 *Borges* in Borges/Hilber, Art. 4 DSGVO Rn. 11; *Schild* in Wolff/Brink, Art. 4 DSGVO Rn. 16.

12 *Karg* in Simitis/Hornung/Spiecker gen. Döhmann, Art. 4 DSGVO Rn. 57; *Schild* in Wolff/Brink, Art. 4 DSGVO Rn. 16.

13 *EuGH*, NVwZ 2014, 435 (437); *Borges* in Borges/Hilber, Art. 4 DSGVO Rn. 11; *Karg* in Simitis/Hornung/Spiecker gen. Döhmann, Art. 4 DSGVO Rn. 56.

14 *ISO/IEC*, ISO/IEC 29100:2011, Kap. 2.9.

15 *ISO/IEC*, ISO/IEC 29100:2011, Kap. 2.11.

16 *ISO/IEC*, ISO/IEC 29100:2011, Kap. 2.6.

17 Dies folgt auch aus *ISO/IEC*, ISO/IEC 29100:2011, Kap. 4.4.3.

18 Vgl. *ISO/IEC*, ISO/IEC 29100:2011, Kap. 4.4.3.

19 Dabei handelt es sich um einen sog. Quasi-Identifikator, welcher eine Person regelmäßig identifizierbar macht, vgl. *Sweeney*, Simple Demographics Often Identify People Uniquely, Carnegie Mellon University, Data Privacy Working Paper 3 2000, S. 7; *Stummer*, INFORMATIK 2022, S. 179-194.

liche Person) in der technischen Begriffsbestimmung aus der ISO/IEC 29100:2011 wiederfinden.

Grundsätzlich enthalten beide Begriffsbestimmungen – die des Art. 4 Nr. 1 DSGVO sowie die der ISO/IEC 29100:2011 – die vier Elemente von personenbezogenen Daten. So bestimmen beide Begriffsbestimmungen, dass jede Information ein personenbezogenes Datum darstellen kann, sofern sie sich auf eine betroffene Person bezieht und diese betroffene Person identifiziert oder identifizierbar ist. Dabei zählen sowohl solche Informationen als personenbezogen, die selbst zur Identifizierung der betroffenen Person genutzt werden können, als auch solche Informationen, die zwar selbst keine Identifizierung ermöglichen, jedoch mit einer identifizierten oder identifizierbaren betroffenen Person verbunden sind.

Vor dem Hintergrund dieser Analyse entsteht zunächst der Eindruck, dass im Recht und in der Technik das gleiche Begriffsverständnis von personenbezogenen Daten herrscht. Bei einer näheren Betrachtung dessen, was unter den Begriffen identifiziert und identifizierbar im Recht und in der Technik verstanden wird, werden jedoch Unterschiede deutlich.

So gilt eine natürliche Person nach der in der rechtlichen Literatur vorherrschenden Sichtweise als identifiziert, wenn deren Identität unmittelbar aus den vorliegenden Informationen hervorgeht. Als identifizierbar gilt sie, wenn die vorliegenden Informationen allein nicht ausreichen, um die Person zu identifizieren, dies jedoch durch die Verknüpfung mit weiteren Informationen gelingt. In der Technik meint *identifiziert* hingegen, dass Informationen einer spezifischen Person zugeordnet wurden – unerheblich davon, ob dies aufgrund der vorliegenden Informationen unmittelbar möglich war oder ob dies zunächst eine Verknüpfung mit weiteren Informationen erforderte. Identifizierbar ist eine Person nach dem technischen Begriffsverständnis dagegen, wenn die Informationen noch nicht mit der spezifischen Person verbunden sind, dies jedoch möglich ist.

3.4 Stellungnahme und Empfehlung

Diese Unstimmigkeit zwischen dem rechtlichen und technischen Verständnis von identifiziert bzw. identifizierbar sowie dem damit zusammenhängenden Verständnis von personenbezogenen Daten ist nach der hier vertretenen Auffassung dadurch bedingt, dass sich die in der rechtlichen Literatur vorherrschende Sichtweise, dass eine Person als identifiziert gilt, wenn deren Identität unmittelbar aus den vorliegenden Informationen hervorgeht,[20] und als identifizierbar gilt, wenn die vorliegenden Informationen allein nicht ausreichen, um die Person zu identifizieren,[21] nur schwer mit den Begriffsbestimmungen des Art. 4 Nr. 1, 5 DSGVO vereinbaren lässt.

Zwar definiert Art. 4 Nr. 1 DSGVO weder den Begriff „identifiziert" noch den Begriff "identifizierbar", er bestimmt jedoch, dass eine natürliche Person als identifizierbar anzusehen ist, wenn sie „*direkt oder indirekt […] identifiziert werden kann*". Insoweit sieht Art. 4 Nr. 1 DSGVO – wie auch das technische Begriffsverständnis durch direkte und indirekte Identifikatoren[22] – zwei Möglichkeiten zur Identifizierung einer natürlichen Person vor: eine direkte und eine indirekte. Aus dem in der Literatur vorherrschenden rechtlichen Verständnis von identifizierbar folgt dagegen nur eine Möglichkeit zur Identifizierung einer natürlichen Person: die Verknüpfung von verfügbaren Informationen mit weiteren Informationen. Es könnte zwar angenommen werden, dass die Verknüpfung von Informationen in direkter und indirekter Weise erfolgen könnte. Bei näherer Betrachtung der Bedeutung von *direkt* und *indirekt* wird jedoch deutlich, dass die Verknüpfung von Informationen stets einen indirekten Charakter hat. So bedeutet der Begriff *direkt* ohne Umweg und kann insoweit synonym zu unmittelbar verstanden werden.[23] *Indirekt* bedeutet hingegen über einen Umweg.[24] Die Identifizierung einer Person durch Verknüpfung von Informationen erfordert, dass aktiv Informationen aus unterschiedlichen Quellen zusammengeführt und miteinander verknüpft werden, um eine Identifizierung herbeizuführen. Sie erfolgt also gerade nicht unmittelbar, vielmehr handelt es sich immer – unabhängig von Aufwand oder Quelle – um einen zusätzlichen Schritt, der für die Identifizierung vorgenommen werden muss. Die Verknüpfung von Informationen ist folglich immer ein Umweg und somit eine indirekte Identifizierung. Ein solches Verständnis entspräche neben der Begriffsbestimmung des Art. 4 Nr. 1 DSGVO auch dem technischen Verständnis der Identifizierung mittels indirekter Identifikatoren.

Des Weiteren folgt aus der Begriffsbestimmung zu Pseudonymisierung aus Art. 4 Nr. 5 DSGVO, dass Daten auch nach ihrer Pseudonymisierung zwar noch als personenbezogen gelten, jedoch ohne Hinzuziehung zusätzlicher Informationen keiner spezifischen betroffenen Person mehr zugeordnet werden können. Die infolge der Pseudonymisierung entstehenden Daten machen die betroffene Person also noch identifizierbar (durch die Möglichkeit der Hinzuziehung zusätzlicher Informationen), identifizieren sie jedoch nicht (da sie ohne die Hinzuziehung zusätzlicher Informationen keiner spezifischen Person zugeordnet werden können). Die im Zusammenhang mit dem herrschenden Begriffsverständnis von *identifiziert* aufgeführten Beispielinformationen (Steuer-Identifikationsnummer, Fingerabdruck etc.) implizieren hingegen, dass eine Person bereits dann als identifiziert gilt, wenn eine Information so einzigartig ist, dass sie nur einer einzigen Person zugeordnet werden kann – unabhängig davon, ob die Zuordnung zu dieser Person ggf. noch die Hinzuziehung zusätzlicher Informationen erfordert oder nicht. Bei einem solchen Verständnis gälten die hinter den pseudonymisierten Daten stehenden Personen immer als identifiziert. Des Weiteren gälte ein Straftatverdächtiger bereits in dem Moment als identifiziert, in dem ein Fingerabdruck gefunden wird und nicht erst dann, wenn der Fingerabdruck mit der Datenbank abgeglichen wurde, weitere Informationen über den Verdächtigen, insbesondere dessen Name, bekannt sind und somit klar ist, von welcher Person die vorgefundenen Fingerabdrücke stammen. Vor dem Hintergrund dieses Beispiels sowie der Begriffsbestimmung des Art. 4 Nr. 5 DSGVO kann – wie auch nach dem technischen Begriffsverständnis – konsequenterweise erst dann von einer *identifizier-*

20 So etwa *Arning/Rothkegel* in Taeger/Gabel, Art. 4 DSGVO Rn. 30; *Klar/Kühling* in Kühling/Buchner, Art. 4 DSGVO Rn. 19; *Ernst* in Paal/Pauly, Art. 4 DSGVO Rn. 8; *Karg* in Simitis/Hornung/Spiecker gen. Döhmann, Art. 4 DSGVO Rn. 57.

21 So etwa *Ziebarth* in Sydow/Marsch, Art. 4 DSGVO Rn. 14; *Klar/Kühling* in Kühling/Buchner, Art. 4 Nr. 1 DSGVO Rn. 18; *Borges* in Borges/Hilber, Art. 4 DSGVO Rn. 11; *Arning/Rothkegel* in Taeger/Gabel, Art. 4 DSGVO Rn. 25; *Karg* in Simitis/Hornung/Spiecker gen. Döhmann, Art. 4 Nr. 1 DSGVO Rn. 49.

22 Gemäß der ISO/IEC 20889 handelt es sich bei direkten Identifikatoren um Attribute, die alleine die Identifizierung einer betroffenen Person ermöglichen. Ein indirekten Identifikator ermöglicht die Identifizierung einer betroffenen Person hingegen erst durch eine Kombination mit anderen Attributen.

23 https://www.duden.de/rechtschreibung/direkt_Adjektiv.

24 https://www.duden.de/rechtschreibung/indirekt,.

ten betroffenen Person die Rede sein, wenn eine Zuordnung zur Person erfolgt ist,[25] denn erst dann ist für die verarbeitende Stelle klar, um welche spezifische Person es sich handelt.[26]

Zur Harmonisierung des in der rechtlichen Literatur bestehenden Begriffsverständnisses mit den rechtlichen Begriffsbestimmungen des Art. 4 Nr. 1, 5 DSGVO sowie dem aus der ISO/IEC 29100:2011 folgenden technischen Begriffsverständnis von personenbezogenen Daten, und somit zur Bildung eines interdisziplinär vereinheitlichten Begriffsverständnisses, sollten die Begriffe vor dem Hintergrund des folgend dargestellten Gesamtkontextes betrachtet und verstanden werden:

Die Identifizierung einer natürlichen Person kann grundsätzlich direkt oder indirekt erfolgen. Direkt erfolgt sie, wenn die vorliegenden Informationen allein ausreichen, um sie einer spezifischen natürlichen Person zuzuordnen und somit eine Identifizierung herbeizuführen.[27] Dies ist insbesondere durch direkte Identifikatoren, wie den Namen, und damit verbundene Informationen gegeben.[28] Ob eine Information direkt identifizierend ist, hängt dabei grundsätzlich vom jeweiligen Kontext ab. So sind Namen bezogen auf die Weltbevölkerung regelmäßig nicht einzigartig und ermöglichen somit keine direkte Identifizierung. Innerhalb einer Fußballmannschaft, eines Unternehmens oder einer Schulklasse können Namen eine direkte Identifizierung hingegen ermöglichen.[29] Neben der direkten Identifizierung kann eine Identifizierung auch indirekt durch indirekte Identifikatoren erfolgen. Indirekt erfolgt sie, wenn die vorliegenden Informationen allein nicht ausreichen, um eine Identifizierung herbeizuführen, eine Identifizierung jedoch durch die Kombination der Informationen mit anderen Informationen ermöglicht wird.[30] Informationen, die eine indirekte Identifizierung ermöglichen sind z.B. Fingerabdrücke, Sozialversicherungsnummern, Steueridentifikationsnummern, Autokennzeichen, Telefonnummern[31] oder IP-Adressen.[32] Wird von der Möglichkeit insofern Gebrauch gemacht, als die Daten einer spezifischen natürlichen Person zugeordnet werden, ist die natürliche Person identifiziert. Wird von der Möglichkeit (noch) kein Gebrauch gemacht – erfolgt also keine direkte oder indirekte Identifizierung, obwohl eine solche theoretisch möglich wäre – ist die natürliche Person identifizierbar[33] – und zwar so lange bis eine Identifizierung erfolgt.

4 Pseudonyme Daten

4.1 Rechtliches Begriffsverständnis

Im Zusammenhang mit pseudonymen Daten findet sich in der DSGVO lediglich eine Legaldefinition von Pseudonymisierung, aus welcher sich auch ein rechtliches Begriffsverständnis von Daten ableiten lässt, die pseudonymisiert wurden (folgend *pseudonymisierte Daten*[34] genannt).

Bei der Pseudonymisierung handelt es sich gemäß Art. 4 Nr. 5 DSGVO um die Verarbeitung personenbezogener Daten in einer Weise, dass die personenbezogenen Daten ausschließlich durch Hinzuziehung zusätzlicher Informationen einer spezifischen betroffenen Person zugeordnet werden können, wobei diese zusätzlichen Informationen gesondert aufbewahrt werden und technischen und organisatorischen Maßnahmen unterliegen müssen, welche gewährleisten, dass die personenbezogenen Daten keiner natürlichen Person zugewiesen werden. Aus der Begriffsbestimmung des Art. 4 Nr. 5 DSGVO folgen damit drei Anforderungen, welche kumulativ erfüllt werden müssen, um Daten i.S.d. DSGVO zu pseudonymisieren:[35]

- Zunächst darf eine Zuordnung der Daten zu einer natürlichen Person nicht ohne Hinzuziehung der zusätzlichen Informationen möglich sein.
- Zudem müssen die zusätzlichen Informationen getrennt aufbewahrt werden.
- Schließlich müssen die zusätzlichen Informationen Gegenstand technischer und organisatorischer Schutzmaßnahmen sein, die gewährleisten, dass keine Identifizierung erfolgt.

Die Grundvoraussetzung für die Pseudonymisierung i.S.d. DSGVO ist dabei, dass der identifizierende Datensatz aufgeteilt wird in einen pseudonymisierten Datensatz einerseits und zusätzliche Informationen andererseits.[36] Dies erfolgt regelmäßig, indem identifizierende Informationen innerhalb eines Datensatzes (z.B. der Name) durch ein Pseudonym (z.B. eine Zahlenkombination) ersetzt werden (hier *konventionelle Pseudonymisierung*[37] genannt).[38] Bei der Zuordnungsregel vom Pseudonym zum identifizierenden Merkmal handelt es sich sodann um die zur Identifizierung erforderliche zusätzliche Information. Zu beachten ist jedoch, dass die Pseudonymisierung i.S.d. Art. 4 Nr. 5 DSGVO zusätzlich zur Separierung von pseudonymisierten Daten und zusätzlichen Informationen fordert, dass die Identifizierung ohne Hinzuziehung der zusätzlichen Informationen ausgeschlossen ist. Dies ist insbesondere dadurch zu gewährleisten, dass die Zuordnungsregel (technisch oder räumlich)[39] getrennt und durch technisch-organisatorische Maßnahmen geschützt aufbewahrt wird.[40] Insoweit erfüllt die Pseudonymisierung i.S.d. DSGVO zusätzliche Anforderungen und stellt somit ein „Mehr" gegenüber der konventionellen Pseudonymisierung dar.

Auf Basis der Begriffsbestimmung des Art. 4 Nr. 5 DSGVO und dem anerkannten Begriffsverständnis von konventioneller Pseudonymisierung lässt sich auch ein rechtliches Begriffsver-

25 So auch *ULD*, Towards a Better Understanding of Identification, Pseudonymization, and Anonymization, S. 24.

26 So auch *Ziebarth* in Sydow/Marsch, Art. 4 DSGO Rn. 17.

27 *ULD*, Towards a Better Understanding of Identification, Pseudonymization, and Anonymization, S. 23.

28 *Art. 29-Datenschutzgruppe*, WP 136, S. 15; *Schild* in Wolff/Brink, BeckOK Datenschutzrecht, Art. 4 DSGVO Rdnr, 16.

29 Art. 29-Datenschutzgruppe, WP 136, S. 15.

30 *Art. 29-Datenschutzgruppe*, WP 136, S. 15; *Schild* in Wolff/Brink, BeckOK Datenschutzrecht, Art. 4 DSGVO Rn. 17; *ULD*, Towards a Better Understanding of Identification, Pseudonymization, and Anonymization, S. 23.

31 *Schild* in Wolff/Brink, BeckOK Datenschutzrecht, Art. 4 DSGVO Rn. 17.

32 *EuGH*, ZD 2017, 24 (24).

33 So auch *Art. 29-Datenschutzgruppe*, WP 136, S. 14.

34 Zum Verhältnis zwischen den Begriffen „pseudonyme Daten" und „pseudonymisierten Daten" siehe Fußnote 47.

35 *Spindler/Dalby* in Spindler/Schuster, Art. 4 DSGVO Rn. 13; *Hansen* in Simitis/Hornung/Spiecker gen. Döhmann, Art. 4 Nr. 5 DSGVO Rn. 31 ff.; *Ernst* in Paal/Pauly, Art. 4 DSGVO Rn. 40 ff.; *Schwartmann/Weiß*, Whitepaper zur Pseudonymisierung der Fokusgruppe Datenschutz, S. 10 f.

36 *ULD*, Towards a Better Understanding of Identification, Pseudonymization, and Anonymization, S. 55 f.

37 Mourby et al., CLSR 2018, 222 (224).

38 Vgl. z.B. *ICO*, Anonymisation: managing data protection risk code of practice, S. 49; *An Coimisiún um Chosaint Sonraí*, Guidance Note: Guidance on Anonymisation and Pseudonymisation, S. 3; Art. 29-Datenschutzgruppe: WP 216, S. 24; *Hansen* in Simitis/Hornung/Spiecker gen. Döhmann, Art. 4 Nr. 5 DSGVO Rn. 4; *Diel/Schreiber* in Selzer, Art. 4 DSGVO Rn. 17.

39 *Ernst* in *Paal/Pauly*, Art. 4 DSGVO Rn. 43.

40 So auch *ULD*, Towards a Better Understanding of Identification, Pseudonymization, and Anonymization, S. 50.

ständnis für pseudonymisierte Daten ableiten. Zu unterscheiden ist dabei zwischen konventionell pseudonymisierten Daten und pseudonymisierten Daten i.S.d. DSGVO: Um konventionell pseudonymisierte Daten handelt es sich bei solchen personenbezogenen Daten, bei denen identifizierende Informationen durch ein Pseudonym ausgetauscht wurden. Um pseudonymisierte Daten i.S.d. DSGVO handelt es sich dagegen erst, wenn die personenbezogenen Daten nach Art. 4 Nr. 5 DSGVO so pseudonymisiert wurden, dass sie ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können, wobei die für die Zuordnung zur natürlichen Person erforderlichen zusätzlichen Informationen getrennt aufbewahrt werden und Gegenstand technischer und organisatorischer Maßnahmen sind, die gewährleisten, dass keine Identifizierung erfolgt.[41]

4.2 Technisches Begriffsverständnis

Auch die ISO 29100:2011 und ISO/IEC 20889:2018 enthalten keine Definitionen von pseudonymen oder pseudonymisierten Daten.[42] Daher gilt es – wie auch beim rechtlichen Begriffsverständnis – bei der Pseudonymisierung anzusetzen, um ein technisches Verständnis von pseudonymisierten Daten abzuleiten.

Nach der ISO 29100:2011 handelt sich bei der Pseudonymisierung um einen *„process applied to personally identifiable information (PII) which replaces identifying information with an alias."*[43] Die ISO/IEC 20889:2018 definiert Pseudonymisierung als *„de-identification technique (3.7) that replaces an identifier (or identifiers) for a data principal (3.4) with a pseudonym (3.26) in order to hide the identity of that data principal."*[44]

Diese beiden Begriffsbestimmungen drücken, wenn auch sie anderslautend sind, das gleiche aus. So handelt es sich bei der Pseudonymisierung gemäß den beiden Definitionen um einen Austausch von identifizierenden Informationen (sog. Identifikatoren) durch einen eindeutigen Identifikator, welcher für eine betroffene Person erstellt wurde, um ihn gegen den oder die üblicherweise genutzten Identifikator(en) auszutauschen (sog. Pseudonym bzw. Alias)[45].[46]

Daraus folgend handelt es sich bei pseudonymisierten Daten um solche, bei denen identifizierende Informationen durch ein Pseudonym ausgetauscht wurden.

4.3 Übereinstimmung

Wie die Analyse des rechtlichen und technischen Begriffsverständnisses von Pseudonymisierung und pseudonymisierten Daten zeigt, bestehen zwischen dem rechtlichen und technischen Begriffsverständnis sowohl Unterschiede als auch Gemeinsamkeiten.

Zwar definieren sowohl das Recht als auch die Technik den Begriff der Pseudonymisierung, woraus sich auch ein rechtliches und technisches Begriffsverständnis von pseudonymisierten Daten[47] ableiten lässt, dieses Begriffsverständnis stimmt inhaltlich jedoch nicht überein. So meint die Pseudonymisierung nach dem technischen Begriffsverständnis den Austausch von identifizierenden Informationen durch ein Pseudonym. Im Recht wird dieses Begriffsverständnis zwar anerkannt (hier konventionelle Pseudonymisierung genannt), die Pseudonymisierung i.S.d. rechtlichen Legaldefinition des Art. 4 Nr. 5 DSGVO erfüllt jedoch zusätzliche Anforderungen und stellt somit regelmäßig ein „Mehr" gegenüber der konventionellen bzw. technischen Pseudonymisierung dar.

4.4 Stellungnahme und Empfehlung

Da die Unterschiede zwischen dem rechtlichen und technischen Begriffsverständnis rund um pseudonymisierte Daten folglich primär auf eine uneinheitlich verwendete Terminologie zurückzuführen sind, wird für ein interdisziplinär vereinheitlichtes Begriffsverständnis vorgeschlagen, zwischen verschiedenen Arten von Pseudonymisierung und pseudonymisierten Daten zu unterscheiden und eine interdisziplinär vereinheitlichte Terminologie einzuführen. Im Rahmen dieser interdisziplinär vereinheitlichten Terminologie sollte unter konventioneller Pseudonymisierung der Austausch von identifizierenden Informationen durch ein Pseudonym verstanden werden. Durch die konventionelle Pseudonymisierung entstehen konventionell pseudonymisierte Daten. Entstehen die pseudonymisierten Daten dagegen durch eine Pseudonymisierung i.S.d. DSGVO, werden personenbezogene Daten also in einer Weise verarbeitet, dass sie ausschließlich durch Hinzuziehung zusätzlicher Informationen einer spezifischen betroffenen Person zugeordnet werden können, wobei diese zusätzlichen Informationen gesondert aufbewahrt werden und technischen und organisatorischen Maßnahmen unterliegen, handelt es sich um i.S.d. DSGVO pseudonymisierte Daten.

5 Anonyme Daten

5.1 Rechtliches Begriffsverständnis

Anders als bei personenbezogenen und pseudonymisierten Daten enthält der Begriffskatalog des Art. 4 DSGVO keinerlei Hinweise darauf, was unter anonymen Daten zu verstehen ist. Jedoch findet sich in den Erwägungsgründen eine Erläuterung zu anonymen Informationen, welche darauf hinweist, was anonyme Daten aus Sicht des Verordnungsgebers sind.

Gemäß Erwägungsgrund 26 der DSGVO handelt es sich bei anonymen Informationen um solche *„Informationen, die sich nicht auf eine identifizierte oder identifizierbare natürliche Person beziehen, oder personenbezogene Daten, die in einer Weise anonymisiert worden sind, dass die betroffene Person nicht oder nicht mehr*

41 *Arning/Rothkegel* in Taeger/Gabel, DSGVO, Art. 4 DSGVO Rn. 126; *ULD*, Towards a Better Understanding of Identification, Pseudonymization, and Anonymization, S. 60 – hier „strictly pseudonymous data" genannt.

42 Zum Verhältnis zwischen den Begriffen „pseudonyme Daten" und „pseudonymisierten Daten" siehe Fußnote 47.

43 *ISO/IEC*, ISO/IEC 29100:2011, Kap. 2.24.

44 *ISO/IEC*, ISO/IEC 20889:2018, Kap. 3.27.

45 *ISO/IEC*, ISO/IEC 20889:2018, Kap. 3.26.

46 So auch NIST, NISTIR 8053, S. 2; Elliot et al,, The Anonymisation Decision Making Framework, S. 15; ENISA, Recommendations on shaping technology according to GDPR provisions, S. 10.

47 In Literatur und Praxis wird vielfach auch von „pseudonymen Daten" gesprochen. Es folgt jedoch weder aus den betrachteten ISO-Normen noch aus der DSGVO ein eigenständiges Begriffsverständnis. Daher wird vertreten, dass pseudonyme und pseudonymisierte Daten synonym zu verstehen sind. Der Einheitlichkeit und Eindeutigkeit wegen wird im Rahmen des vorliegenden Beitrags dennoch vorgeschlagen für Daten, die infolge der Pseudonymisierung entstehen, durchgängig den Begriff „pseudonymisierte Daten" zu verwenden.

identifiziert werden kann." Sie stellen somit den Konterpart zu personenbezogenen Daten dar[48] und unterliegen nicht den Vorgaben der DSGVO.[49]

Anonymität beschreibt demnach einen Zustand von Daten, in dem natürliche Personen nicht identifiziert werden können. Anonymität kann von Beginn an vorliegen sowie durch Anonymisierung entstehen. Der Begriff der Anonymisierung wird in der DSGVO nicht näher konkretisiert. Aus der Erläuterung des Begriffs von anonymen Informationen folgt jedoch, dass die Anonymisierung das Verändern personenbezogener Daten derart ist, dass eine Identifizierung der hinter den Datensätzen stehenden natürlichen Personen nicht mehr möglich oder nach allgemeinem Ermessen unwahrscheinlich ist.[50] Dies folgt insbesondere daraus, dass gemäß Erwägungsgrund 26 DSGVO bei der Feststellung, ob Daten eine Identifizierung der natürlichen Person ermöglichen, „*alle Mittel berücksichtigt werden [sollen], die von dem Verantwortlichen oder einer anderen Person nach allgemeinem Ermessen wahrscheinlich genutzt werden, um die natürliche Person direkt oder indirekt zu identifizieren.*" Maßgeblich sind dabei alle objektiven Faktoren, wie die Kosten der Identifizierung und der dafür erforderliche Zeitaufwand, wobei die zum Zeitpunkt der Verarbeitung verfügbaren Technologien und technologische Entwicklungen zu berücksichtigen sind.

5.2 Technisches Begriffsverständnis

In der ISO/IEC 29100:2011 finden sich im Zusammenhang mit anonymen Daten Begriffsbestimmungen zu Anonymität, Anonymisierung und anonymisierten Daten. Den Begriff der anonymen Daten kennt die ISO/IEC 29100:2011 hingegen nicht.

Der Begriff der anonymisierten Daten wird definiert als „*data that has been produced as the output of a personally identifiable information anonymization process.*"[51]

Ein *personally identifiable information anonymization process* ist dabei ein Anonymisierungsprozess, durch den personenbezogene Daten unwiderruflich so verändert werden, dass die betroffene Person weder durch den Verantwortlichen alleine noch in Zusammenarbeit mit anderen Parteien direkt oder indirekt identifiziert werden kann. Durch Anwendung eines solchen Anonymisierungsprozesses entstehen anonymisierte Daten. Bei diesen handelt es sich folglich um Daten, die in Folge der Anonymisierung so verändert wurden, dass die betroffene Person nicht mehr identifiziert werden kann.

Anonymität wird schließlich definiert als „*characteristic of information that does not permit a personally identifiable information principal to be identified directly or indirectly.*"[52] Anonymität liegt folglich vor, wenn es eine Information nicht erlaubt eine betroffene Person direkt oder indirekt zu identifizieren.

5.3 Übereinstimmung

Wie auch bei pseudonymisierten Daten, hat die Analyse des rechtlichen und technischen Begriffsverständnisses rund um anonyme Daten gezeigt, dass zwischen dem rechtlichen und technischen Begriffsverständnis sowohl Unterschiede als auch Gemeinsamkeiten bestehen. Im Wesentlichen herrscht zwar das gleiche Grundverständnis rund um Anonymität, für den gleichen Gegenstand werden jedoch teilweise unterschiedliche Begriffe verwendet.

So besteht ein Unterschied zwischen dem rechtlichen und technischen Begriffsverständnis zunächst darin, dass das Datenschutzrecht lediglich den Begriff der anonymen Informationen kennt, nicht jedoch den Begriff der anonymisierten Daten, wohingegen die Technik lediglich den Begriff der anonymisierten Daten, nicht jedoch den Begriff der anonymen Daten kennt. Unter anonymen Informationen werden im rechtlichen Sinne sowohl solche verstanden, die sich von Beginn an auf keine identifizierte oder identifizierbare natürliche Person beziehen, als auch solche, die in einer Weise anonymisiert worden sind, dass die betroffene Person nicht oder nicht mehr identifiziert werden kann. Letztere Alternative des rechtlichen Begriffsverständnisses von anonymen Informationen entspricht dem technischen Begriffsverständnis von anonymisierten Daten. So handelt es sich bei anonymisierten Daten im technischen Sinne um Daten, die in Folge der Anonymisierung so verändert wurden, dass die betroffene Person nicht mehr identifiziert werden kann. Unter Anonymität wird schließlich sowohl im Recht als auch in der Technik ein Zustand verstanden, welcher es nicht mehr ermöglicht, eine betroffene Person zu identifizieren.

In den Grundzügen entsprechen sich auch das rechtliche und technische Begriffsverständnis von Anonymisierung. So handelt es sich in beiden Fachdisziplinen um eine Veränderung personenbezogener Daten derart, dass keine natürliche Person mehr identifiziert werden kann. Fraglich ist jedoch, ob das rechtliche und technische Begriffsverständnis auch im Detail – das heißt hinsichtlich der konkreten Anforderungen – übereinstimmen. So ist in Erwägungsgrund 26 der DSGVO die Rede davon, dass eine Identifizierung nicht mehr möglich oder nach allgemeinem Ermessen unwahrscheinlich ist. In der ISO/IEC 29100:2011 ist hingegen die Rede von einer „unwiderruflichen" Veränderung der Daten. Unklar ist dabei, ob die Anonymisierung nach dem technischen Verständnis tatsächlich unwiderruflich sein muss, oder ob es ausreicht, dass die Anonymisierung praktisch, etwa aufgrund zu großen Aufwands, unwiderruflich ist.

5.4 Stellungnahme und Empfehlung

Einen Hinweis darauf, wie die Begriffsbestimmung der ISO/IEC 29100:2011 zu verstehen ist, bietet die ISO/IEC 29889:2018. Dieser Standard nutzt zwar weder den Begriff der anonymen Daten noch den Begriff der Anonymisierung, sondern stattdessen die Begriffe *de-identified dataset* und *de-identification process*. Der Begriff *de-identification* wird dabei jedoch so genutzt wie anonymization in der ISO 29100:2011[53] und meint die Transformation eines Datensatzes mit dem Ziel, das Ausmaß, in dem Informationen mit betroffenen Personen in Verbindung gebracht werden, zu reduzie-

48 *Karg*, DuD 2015, 520 (523); *Klar/Kühling* in Kühling/Buchner, Art. 4 Nr. 1 DSGVO Rn. 31; *Roßnagel*, ZD 2021, 188 (189).

49 *BfDI*, Positionspapier zur Anonymisierung unter der DSGVO unter besonderer Berücksichtigung der TK-Branche, S. 3; *Arning/Rothkegel* in Taeger/Gabel, Art. 4 DSGVO Rn. 47.

50 Ähnlich *Meyer*, ZD 2021, 669 (673 f.); abgeleitet aus den Kriterien des Erwgr. 26 DSGVO, der nationalen Gesetzgebung (z.B. § 3 Abs. 6 BDSG a.F.; § 3 Abs. 3 Nr. 1 BbgDSG; § 2 Abs. 4 BrmDSG; § 2 Abs. 7 DSG LSA; § 24 Nr. 18 NDSG; § 28 Abs. 3 ThürDSG) sowie dem national herrschenden Begriffsverständnis, vgl. etwa *Roßnagel/Geminn*, ZD 2021, 487 (487); *Schwartmann/Hermann* in Leupold/Wiebe/Glossner, IT-Recht, Teil 10.5 Rn. 21 f.; *Ernst* in Paal/Pauly, DSGVO, Art. 4 DSGVO Rn. 48; *Roßnagel*, ZD 2021, 188 (189).

51 *ISO/IEC*, ISO/IEC 29100:2011, Kap. 2.3.

52 *ISO/IEC*, ISO/IEC 29100:2011, Kap. 2.1.

53 *ISO/IEC*, ISO/IEC 20889:2018, Annex B.

ren.[54] Dabei erkennt der Standard an, dass trotz Anwendung einer De-Identifizierungstechnik (wie der Anonymisierung) ein Rest-Re-Identifizierungsrisiko bleibt.[55] Vor diesem Hintergrund wird vertreten, dass das technische Begriffsverständnis von Anonymisierung auch im Detail so zu verstehen ist wie das rechtliche Begriffsverständnis und dass eine praktische Unwiderrufbarkeit der Anonymisierung ausreicht.[56]

Folglich herrscht im Recht und in der Technik im Wesentlichen das gleiche Grundverständnis rund um Anonymität. Dabei werden jedoch teilweise für den gleichen Gegenstand unterschiedliche Begriffe und Definitionen verwendet. Um diese Differenzen zu beseitigen und zu einem interdisziplinären Verständnis zu gelangen, sollten anonyme Daten als Informationen verstanden werden, die eine natürliche Person nicht mehr identifizierbar machen. Anonyme Daten können bereits von Beginn an vorliegen oder durch Veränderung personenbezogener Daten derart, dass eine Identifizierung der hinter den Datensätzen stehenden natürlichen Personen nicht möglich oder nach allgemeinem Ermessen unwahrscheinlich ist (praktische Anonymisierung) entstehen. Sofern anonyme Daten durch Anonymisierung entstehen, handelt es sich um anonymisierte Daten.

54 *ISO/IEC*, ISO/IEC 20889:2018, Kap. 3.7.
55 *ISO/IEC*, ISO/IEC 20889:2018, Kap. 7.1.
56 Ein ähnliches Verständnis des technischen Begriffsverständnisses aus der ISO/IEC 29100:2011 scheint auch die Art. 29-Datenschutzgruppe zu vertreten: *Art. 29-Datenschutzgruppe*, WP 216, S. 6.

6 Zusammenfassung

Aufgrund der rechtlichen und technischen Natur des Themenkomplexes der Anonymisierung (und Pseudonymisierung) bedarf es bei dessen Bearbeitung einer interdisziplinären Betrachtung. Hierbei gilt es bereits bei den grundlegenden Begriffen anzusetzen und interdisziplinär vereinheitlichte Begriffsbestimmungen zu nutzen.

Dieser Beitrag hat durch eine Analyse des rechtlichen und technischen Begriffsverständnisses rund um Personenbezogenheit, Pseudonymität und Anonymität Unterschiede und Gemeinsamkeiten herausgearbeitet. Auf Basis dessen konnte festgestellt werden, dass das aktuell bestehende rechtliche und technische Begriffsverständnis noch nicht vollumfänglich harmonisiert ist, eine Harmonisierung jedoch durch eine leichte Anpassung des aktuell vorherrschenden rechtlichen Verständnisses von *identifiziert* und *identifizierbar* sowie durch eine differenzierte Nutzung der Terminologie möglich ist. Insbesondere gilt es zwischen konventioneller Pseudonymisierung und Pseudonymisierung i.S.d. DSGVO sowie den daraus jeweils entstehenden pseudonymisierten Daten zu differenzieren. Zudem gilt es die Unterschiede zwischen anonymen und anonymisierten Daten zu würdigen und sprachlich anzuerkennen: So handelt es sich zwar bei allen anonymisierten Daten auch um anonyme Daten, nicht jedoch bei allen anonymen Daten um anonymisierte Daten.