INFORMATIONSEXTRAKTION UND DIE DS-GVO

Datenschutzrechtliche Grenzen der Nutzung von Open Source Intelligence

Stephanie von Maltzan

Wissenschaftliche Mitarbeiterin, Karlsruher Institut für Technologie, Zentrum für Angewandte Rechtswissenschaften
Vincenz-Prießnitz-Str. 3, 76131 Karlsruhe, DE
Stephanie.maltzan@kit.edu; https://www.zar.kit.edu/

Schlagworte: ***Informationsextraktion, OSINT, allgemein zugängliche Daten, IT-Sicherheit, DS-GVO***

Abstract: *Der Beitrag widmet sich der praktisch relevanten Frage, in welchen datenschutzrechtlichen Grenzen allgemein zugängliche Daten erhoben und verarbeitet werden dürfen. Die Aufbereitung von OSINT bietet verbesserte Möglichkeiten IT-Sicherheitsvorfälle zu detektieren sowie den Angreifer zu identifizieren. Einhergehend damit ist nicht nur die massive Verarbeitung von personenbezogenen Daten, sondern auch der Transparenzverlust. Dieser Konflikt zwischen der IT-Sicherheit auf der einen und die Verletzung von Persönlichkeitsrechten auf der anderen Seite muss mittels wirksamen Zusammenwirkens verschiedener technischer und organisatorischer Schutzmechanismen gelöst werden.*

1. Einführung

In Anbetracht der verstärkt auftretenden Bedrohungen[1] durch sowohl externe als auch interne Angriffe auf IT-Systeme werden unter Berücksichtigung der Sicherheitsanforderungen vermehrt Maßnahmen zur Sicherheit der IT erprobt und ausgelotet. Unternehmen, die aktuellen Angriffen widerstehen wollen, müssen in Echtzeit einschätzen können, welche ihrer Systeme von einem Cyber-Angriff oder einer Datenpanne betroffen sind bzw. sein könnten. Solche Schwachstellen und Sicherheitsvorfälle möglichst schnell zu erkennen und zu behandeln, gehört zu den Grundprinzipien der Informationssicherheit und damit zum Datenschutzmanagement. Aufgrund dessen ist eine zuverlässige Situationsbeschreibung und Prognose auftretender Gefahren von immenser Bedeutung. Im Rahmen einer Risikoanalyse mittels der Etablierung eines sogenannten Incident Response[2] Managements werden regelmäßig sicherheitsrelevante Lagebilder erstellt, ausgewertet und weitergeleitet, um größtmögliche Sicherheit gewährleisten zu können. Infolge der Vernetzung der IT-Systeme genügt es jedoch unter Umständen nicht, lediglich die eigene Sicherheitslage zu betrachten. Vor dem Hintergrund vielfältiger Bedrohungsszenarien ist es für Unternehmen notwendig, ein ganzheitliches IT-Sicherheitskonzept zu erstellen und umzusetzen. Daher hat sich das Verbundprojekt ITS.Overview[3] zum Ziel gesetzt, Lagebildinformationen zu erstellen, diese mit Lagebildern von Unternehmen gleicher Branche zu korrelieren und auszutauschen. Lagebilder werden dabei aus gesammelten Daten gewonnen, die mittels einer Risikoanalyse als eine wahrscheinliche Gefahr identifiziert wurden und den Unternehmen sowohl von außen als auch von innen drohen könnten. Basierend auf dieser umfangreichen Datenanalyse können die Unternehmen schnell geeignete Schritte einleiten. Mit zunehmender Technisierung und ubiquitärer Datenverarbeitung gibt es vielfälti-

1 Ponemon Institute LLC and IBM Security, 2017 Cost of Data Breach Study, 06.2017; BSI, Die Lage der IT-Sicherheit in Deutschland 2017, 2017.

2 Incident Response ist ein organisierter Ansatz der unmittelbaren Reaktion auf erkannte oder auch vermutete IT-Sicherheitsvorfälle inklusive hierzu vorbereitender Maßnahmen und Prozesse.

3 https://itsec.cs.uni-bonn.de/overview/.

ge Möglichkeiten Sicherheitsvorfälle zu detektieren und den Angreifer bzw. dessen Methodik zu identifizieren. Mittels Informationsextraktion[4] lassen sich aus verschiedenen allgemein zugänglichen Datensätzen Informationen generieren, die in ihrem Aussagegehalt weit über den Informationsgehalt einzelner Daten hinausgehen. Das dabei genutzte Verfahren wird als Open Source Intelligence (OSINT) oder auch Social Media Intelligence (SOCMINT) bezeichnet. Auf Grundlage zuvor erkannter Muster sollen mittels statistischer Verfahren durch Daten von sozialen Netzwerken Auffälligkeiten erkannt, gesammelt und verknüpft werden. Deren schnelle und akkurate Aufbereitung von unstrukturierten zu strukturierten Informationen hat sich durch die Menge an allgemein zugänglichen Daten qualitativ wie auch quantitativ gewandelt. Dies hat im Rahmen von Gefährdungsanalysen konkrete Auswirkungen auf die Risikoprävention durch exaktere datenbasierte Vorhersagen. Aus den aufbereiteten Informationen lassen sich Trends und Muster ableiten sowie Anomalien entdecken. Um effektiv gegen die zunehmend aufwändigeren Angriffe auf die IT-Sicherheit gerüstet zu sein, genügt ein reaktives[5] Sicherheitskonzept nicht mehr den heutigen Anforderungen. Die Aufbereitung von OSINT bietet verbesserte Möglichkeiten Sicherheitsvorfälle im Bereich der IT-Struktur zu detektieren und zu vermindern sowie den Angreifer durch die umfassende Auswertung von OSINT zu identifizieren. Als weiteres Anwendungsfeld ist die Gefahrenabwehr zu benennen. Mittels Auswertung der sozialen Netzwerke, die ebenfalls eine maßgebliche Rolle bei der Verbreitung von extremistischer Propaganda spielen, sollen über die IT-Sicherheit hinaus auch systematisch Terrorpropaganda und Mobilisierungsvideos aufgespürt und Beweismittels für die Strafverfolgung[6] beigebracht werden. Dies stellt aufgrund der massenhaften Erfassung auch von Daten, die nicht durch den Betroffenen allgemein zugänglich gemacht wurden sowie der Gefahr des Erstellens umfassender Persönlichkeitsprofile durch Korrelation der Daten einen besonders intensiven Eingriff dar. Kritisch zu beäugen ist die Erklärung der Bundesregierung – unter Berufung auf das Urteil des Bundesverfassungsgerichts zur Onlinedurchsuchung 1 BvR 370/07 – dass Daten, die Nutzer von sozialen Medien öffentlich ins Netz stellen, frei zur Überwachung sind.[7] Einer Ermächtigungsgrundlage bedarf es nach diesem Urteil jedoch, soweit Informationen, die durch Sichtung allgemein zugänglicher Inhalte erhoben, gespeichert und gegebenenfalls korreliert wurden und sich daraus eine besondere Gefahrenlage für die Persönlichkeit des Betroffenen ergibt. Ein absoluter Schutz würde dem Umstand des bewussten Teilens dennoch nicht gerecht. In seiner Entwicklung des Rechts auf informationelle Selbstbestimmung hat das BVerfG betont, dass der Einzelne als soziales Wesen zwangsläufig in kommunikative Vorgänge integriert ist und insoweit ein absoluter Schutz seiner Daten weder möglich noch vom Grundgesetz gewollt ist.[8] Dazu gehört in der Konsequenz, dass ein geringerer Schutz für Daten besteht, die für jedermann ohne weiteres erkennbar sind.[9] Diese sind Abbild der sozialen Realität, in der sich der Einzelne bewegt. Informationen, die allgemein zugänglich sind, genießen daher insgesamt einen geringeren Schutz, ohne aber gänzlich schutzlos zu sein.

2. Die Erhebung und Verarbeitung sicherheitsrelevanter Informationen

Der konkrete Einsatz von OSINT wirft verschiedene rechtliche Fragen auf. Dieser Beitrag will aufgrund der systemimmanenten Datenansammlung ein Augenmerk insbesondere auf eine datenschutzgerechte Implementierung der Systeme richten.

---

4 Hierbei versteht man die Anwendung von Verfahren mit dem Ziel der automatischen maschinellen Verarbeitung von unstrukturierten Informationen.

5 Die aktuelle IT-Sicherheitslage wird im unternehmerischen Umfeld zumeist mittels passender Zusatzinformationen reaktiv unterstützt. Da erst bei Eintreten eines Vorfalls, also dem festgestellten Entstehen einer Sicherheitslage, der Prozess der Informationsanreicherung stattfindet, ist der Ansatz als reaktiv zu bezeichnen.

6 In Bezug auf die öffentlich-rechtliche Fragestellung ist zu verweisen auf Rückert, Zeitschrift für die gesamte Strafrechtswissenschaft 129 (2017).

7 Deutscher Bundestag, Antwort der Bundesregierung auf die Kleine Anfrage des Abgeordneten Andrej Hunko, weiterer Abgeordneter, der Fraktion DIE LINKE – Drucksache 18/540 – sowie der schriftlichen Nachfrage, 5. März 2014.

8 So bereits BVerfGE 65, 1 (43 f.).

9 BVerfGE 120, 378 (404) – KFZ-Kennzeichenerfassung.

2.1. Allgemein zugängliche Daten

Gewonnen werden die unstrukturierten Informationen unter anderem aus sozialen Medien. Zu sozialen Medien können auch Blogs, Foren und Plattformen zählen.[10] Foren, wie Twitter nehmen eine gewisse Hybridstellung ein, da es den Nutzern zusätzlich auch nicht öffentliche Kommunikationskanäle bietet. Ausgerichtet ist Twitter dennoch auf die Darbietung öffentlicher Kommunikation. Eine Vielzahl der in offenen sozialen Medien verfügbaren Daten weisen einen Personenbezug auf. Unabhängig von dem Umstand des teilweise willentlichen Einstellens durch die Betroffenen sowie der Möglichkeit, diese Informationen ohne nennenswerten Zugangsbarrieren online abzurufen, sind sie als personenbezogene Daten grundsätzlich von der DS-GVO und BDSG neu umfänglich geschützt. Bei diesen Medien bestehen für die Wahrnehmung bestimmter Inhalte – ebenso auch personenbezogene Daten – allerdings geringere Beschränkungen, soweit diese «offen und frei zugänglich»[11] sind. Der Begriff des «Öffentlichmachens» ist gesetzlich nicht definiert. Die DS-GVO stellt jedoch vor allem auf den Verarbeitungszweck und nicht die Herkunft der Daten ab. Damit unterliegt auch die Verarbeitung frei zugänglicher Daten infolge des Personenbezugs Beschränkungen. Sie sind grundsätzlich erst einmal umfänglich von der DS-GVO und BDSG neu geschützt. Als allgemein zugängliche Quellen zählen alle Formen der Bereitstellung von Informationsquellen, die Informationen an eine unbestimmte Anzahl von Personen vermitteln können und sollen.[12] Wichtig ist nur, dass die Inhalte von jedem gesucht und abgerufen werden können, sie also gerade nicht durch Privatsphäreneinstellungen nur einem bestimmten Personenkreis zugänglich gemacht werden.[13] Maßgeblich zur Bestimmung der Öffentlichkeit ist demnach, ob die Daten der Allgemeinheit oder nur innerhalb einer abgeschlossener Gruppen bzw. eines Kreises zur Verfügung gestellt wurden.[14] Unproblematisch ist dies, wenn netzwerkgeteilte Informationen ohne Verwendung von Zugangsbarrieren frei abrufbar und damit geeignet sind, obige Kriterien zu erfüllen.[15] Soziale Medien stellen unterschiedliche Schwellen für Anmeldungen. Es stellt sich damit die Frage, ab welcher Verwendung von Zugangsbarrieren – beispielsweise durch Anmeldung – Informationen nicht mehr als öffentlich angesehen werden können. Soweit keine oder nur eine unerhebliche faktische Beschränkung der Wahrnehmbarkeit vorliegt und damit der Zugang ohne besonderen Aufwand für jedermann möglich ist, spricht dies für Daten, die als öffentlich anzusehen sind. Dies gilt auch bei technischen Zugangsbarrieren, die Bots und Crawlern den Zugang erschweren bzw. das Hinterlegen von der E-Mail-Adresse, wenn zusätzlich keine individuellen Anforderungen an den Zugang gestellt werden.[16] Eine individuelle Zugangssperre liegt vor, wenn die Anmeldung spezifisch zur Überprüfung der Zugehörigkeit zum Adressatenkreis des Contents dient. Beispielsweise sind bei Twitter keine besonderen technischen Hürden für eine Anmeldung vorgesehen. Somit ist es für jeden ohne besondere Zugangsschwelle zugänglich. Die dort netzwerkgeteilten Informationen sind als öffentlich anzusehen.

Unterstützt werden diese Informationen mit sicherheitsrelevanten Informationen aus den unternehmenseigenen Ticketsystemen[17] und aus Events von MISP (Malware Information Sharing Platform)[18]. Als sicherheitsrelevante Information ist vor allem die IP-Adresse, E-Mail-Adresse, URL und der Hostname maßgeblich zu berücksichtigen.

---

10 BVerfGE 120, 274 (345).

11 Golla/Hofmann/Bäcker, Datenschutz und Datensicherheit – DuD 2018, 89 ff. (97).

12 Vgl. Kühling/Buchner, DS-GVO, 2017, Art. 9 Rn. 78; BVerfGE 33, 52 (65).

13 Bedeutsam ist die Abgrenzung auch hinsichtlich einer strafbaren Handlung z.B. bei einer willentlichen Übermittlung nicht allgemein zugänglicher personenbezogener Daten gem. § 42 BDSG neu.

14 Weichert, Datenschutz und Datensicherheit – DuD 38 (2014), 831 ff., 257 f.

15 Vgl. Simitis (Hrsg.), Bundesdatenschutzgesetz, 2014, 28 Rn. 151.

16 Golla/Hofmann/Bäcker, Datenschutz und Datensicherheit – DuD 2018, 89 ff. (97).

17 System zur Dokumentation und Problembehebung von möglichen Sicherheitsvorfällen mittels Eröffnung eines Tickets. Tickets bzw. Incidents werden mithilfe des Ticketsystems empfangen, klassifiziert und bearbeitet. Gängige Ticketsysteme sind beispielsweise TheHive oder Jira.

18 https://www.misp-project.org/.

2.2. Rechtsgrundlage allgemein zugänglicher Daten

Für die Annahme des Öffentlichmachens genügt jedoch nicht das «bloße Dasein im öffentlichen Raum»[19]. Beispielsweise sind Angaben auf Blogs von anderen als dem Betroffenen nicht offenkundig selbst öffentlich gemacht, wenn die Zustimmung des Betroffenen nicht aus den Umständen erkennbar ist. Die Voraussetzungen sind eng auszulegen.[20] Mit Abkehr vom strikten Grundsatz der Direkterhebung, wodurch die Daten nicht mehr beim Betroffenen erhoben werden müssen, ist grundsätzlich aber eine Erhebung bei Dritten möglich. Mangelt es offenkundig am Öffentlichmachen von Daten bei Dritten können neben der Einwilligung die gesetzlichen Erlaubnistatbestände des Art. 6 DS-GVO unter besonderer Berücksichtigung der Zweckbindung nach Art. 5 Abs. 1 lit. b i.V.m. Art. 6 Abs. 4 DS-GVO in Betracht kommen. Letztlich kommt es auf die Unterscheidung von Primär- und Sekundärdaten und dem Verarbeitungskontext an. Die bloße Kenntnisnahme, die der Verwendungsart durch Gestattung entspricht, bedarf keiner Ermächtigungsgrundlage. Das gezielte Zusammentragen und die Auswertung der Informationen gehen jedoch mit einer höheren Eingriffsintensität einher, die eine Rechtsgrundlage benötigt. Bei Primärdaten ist jedoch zu beachten, dass durch die Gestattung des Öffentlichmachens und damit partieller Verzicht auf die Vertraulichkeit ein geringerer Schutz besteht und die Abwägungskriterien überwiegen können, die für eine Datenverarbeitung sprechen.

In den meisten Fällen können sich die Verantwortlichen nicht auf die Einwilligung nach Art. 7 DS-GVO von Betroffenen stützen, da sie schwerlich praktikabel ist.[21] Dies folgt zum einen aus der Datenart und der Informationsquelle an sich und zum anderen aus dem Weg der Verbreitung bzw. Erhebung. Eine Verarbeitung ist auch nicht nach Art. 6 Abs. 1 lit. c DS-GVO i.V.m. Art. 32 DS-GVO zulässig. Die Verantwortlichen unterliegen keiner rechtlichen Verpflichtung der Vornahme einer Datenverarbeitung. Diese ist lediglich notwendiger Bestandteil der Gewährleistung geeigneter technischer und organisatorischer Maßnahmen. Zumal die vorliegende OSINT Aufbereitung lediglich dem Erstellen von sicherheitsrelevanten Lagebildern dient. Die Rechtsgrundlage kann auf Kriterien der Rechtmäßigkeit nach Art. 6 Abs. 1 lit. f DS-GVO gestützt werden. Danach ist Datenverarbeitung als Mittel für die Erfüllung eigener Geschäftszwecke zulässig, soweit sie zur Wahrung berechtigter Interessen[22] des Verantwortlichen erforderlich ist und kein Grund zur Annahme besteht, dass das schutzwürdige Interesse des Betroffenen an dem Ausschluss der Verarbeitung oder Nutzung überwiegt. Der Verpflichtung, die Interessen des Betroffenen zu berücksichtigen, muss durch eine vorausgehende Abwägung Genüge getan werden. Hierbei sind nach Erwägungsgrund 47 u.a. die vernünftigen Erwartungen der Nutzer sozialer Medien als Betroffene heranzuziehen. Es dürfen somit solche Daten nicht zu Analysezwecken verwendet werden, wenn der Betroffene ein berechtigtes Interesse daran hat, dass diese nicht verwendet werden. Es besteht infolge der Öffentlichkeit allerdings eine vereinfachte Erhebung dieser Daten. Grundsätzlich wird von der Vermutung ausgegangen, dass die Verwendung den Belangen des Betroffenen grundsätzlich nicht widerspricht, soweit die Interessen des Betroffenen nicht offensichtlich[23] überwiegen.[24] Der Betroffene, der seine Daten ausdrücklich der Öffentlichkeit bereitstellt, verzichtet auf den spezifischen Schutz der DS-GVO bei Erhebung seiner Daten.[25] Eine solche vereinfachte Verarbeitung und damit einhergehenden Abwägung gilt allerdings nur für Primärdaten. Die Sorgfaltspflicht des Verantwortlichen muss infolge des leichteren Zugangs sowie der möglichen Verwendung von Sekundärdaten entsprechend restriktiver ausfallen. Grundsätzlich dürfen die Angaben so übernommen werden, wie sie in der Informationsquelle vorhanden sind, es hat jedoch eine Überprüfung der Korrektheit und Vollständigkeit zu erfolgen. Zumindest in den Fällen, in denen – bei-

19 Kühling/Buchner, DS-GVO, 2017, Art. 9 Rn. 82.
20 Däubler/Klebe/Wedde u.a., Bundesdatenschutzgesetz, Aufl. 5, 2016 § 28 Rn. 257.
21 Gola/Klug/Körffer u.a., BDSG Bundesdatenschutzgesetz, Aufl. 10, 2015; Plath/Becker, BDSG/DSGVO, Aufl. 2, 2016, § 4a Rn. 70.
22 Nach Erwägungsgrund 47 sind diese weit zu verstehen.
23 Hiervon ist bei Erhebung von Sekundärdaten bzw. intimen Daten auszugehen.
24 Simitis (Hrsg.), Bundesdatenschutzgesetz, 2014, § 28 Rn. 162.
25 Kühling/Buchner, DS-GVO, 2017, Art. 9 Rn. 77.

spielsweise bei der Angabe von Sicherheitsvorfällen – damit gerechnet werden kann, dass diese Daten für die Öffentlichkeit von Interesse sind und nicht nur von einem beschränkten Personenkreis wahrgenommen werden, überwiegt das Interesse der Verarbeitung.

2.3. Automatisierte Entscheidungen und deren Reichweite

Einhergehend mit der Auswertungsmethodik und der Möglichkeit große Datenmengen zu analysieren, können umfassend Profile erstellt werden, die die Rechte und Freiheiten der Betroffenen erheblich beeinträchtigen können. Eine solche automatisierte Auswertungsmethode und damit Nutzung bestimmter Ergebnisse einer Datenverarbeitung ist dogmatisch nicht grundsätzlich verboten,[26] sondern nach Art. 22 DS-GVO in gewissem Umfang gestattet. Dem Konzept der informationellen Selbstbestimmung folgend, soll der Betroffene nicht zum «bloßem Objekt von Computeroperationen degradiert»[27] und die Verantwortung der Entscheidung nicht Computersystemen zugeschrieben werden. Anhand der Wortwahl der benannten Tatbestandsalternativen «rechtliche Wirkung» bzw. «erhebliche Beeinträchtigung» wird deutlich, dass Art. 22 DS-GVO lediglich schwerwiegende Auswirkungen abdeckt. Erwägungsgrund 71 enthält für letztere Tatbestandsalternative folgende typische Beispiele: «automatische Ablehnung eines Online-Kreditantrags» oder «Online-Einstellungsverfahren ohne jegliches menschliche Eingreifen». Hierbei ist zu fragen, ob eine reine Gefährdungsanalyse mit pseudonymisierten Daten als auch die Weitergabe von sicherheitsrelevanten IP-Adressen (nach Auflösung der Zuordnungsregel) eine nach Art. 22 DS-GVO geforderte schwerwiegende Auswirkung hat. Unabhängig hiervon sind geeignete technische und organisatorische Maßnahmen zu ergreifen sowie dem Betroffenen durch geeignete Maßnahmen die Wahrung seiner berechtigten Interessen zu ermöglichen. Hierzu gehört auch das Erteilen von Informationen über die Datenverarbeitung gemäß Art. 12 Abs. 1 ff. DS-GVO. Einhergehend mit der Auswertungsmethodik werden diese durch die fehlende Transparenz sowie Kenntnis der Datenverarbeitung und dem Risiko der Fehlprogrammierung konterkariert. Sowohl Unternehmen als auch staatliche Einrichtungen betrachten ihre Analysealgorithmen zumeist als Betriebs- und Geschäftsgeheimnis. Dies hat zur Folge, dass eine unabhängige Überprüfung der Analyseverfahren und auch -ergebnisse faktisch kaum möglich ist. Deshalb sind Konzepte notwendig, bei denen eine hinreichende Sicherheit der personenbezogenen Daten gewährt werden können. Derartige Verfahren bedürfen regelmäßig eines wirksamen Zusammenwirkens verschiedener technisch organisatorischer Schutzmechanismen.[28]

2.4. Gewährleistung geeigneter Sicherheitsmaßnahmen

Unabdingbar sind daher geeignete Sicherheitsmaßnahmen,[29] die nicht nur die Einhaltung der wesentlichen IT-Sicherheitsprinzipien[30] gewährleisten, sondern auch Datenkonformität sicherstellen und damit unangemessene Folgen für die Betroffenen – ohne sich über das berechtigte Interesse des für die Verarbeitung Verantwortlichen hinwegzusetzen – abmildern können. Diese Schutzziele sollten dabei – dem Ansatz einer Layered

---

[26] Mit weiteren Anmerkungen zur Auslegung: *Article 29 Data Protection Working Party*, WP251, 2018 21.

[27] Kühling/Buchner, DS-GVO, 2017, Art. 22 Rn. 11.

[28] Siehe Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein, Verkettung digitaler Identitäten, 2007.

[29] Hierzu zählen unter anderem strikte Einschränkungen der Anzahl der erhobenen Daten, die Löschung dieser unmittelbar nach deren Verwendung, technische und organisatorische Maßnahmen zur Sicherstellung der funktionellen Trennung, ein angemessener Einsatz von Anonymisierungstechniken, Datenaggregation und Technologien zur Stärkung der Privatsphäre, aber auch mehr Transparenz, verstärkte Rechenschaftspflicht und die Möglichkeit, die Verarbeitung zu verweigern.

[30] Die klassischen Schutzziele der IT-Sicherheit sind die Vertraulichkeit, Integrität und Verfügbarkeit.

Defense[31] folgend – in möglichst vielen Schichten der IT-Umgebung umgesetzt werden. Die Sicherheit von IT-Systemen[32] bedingt immer auch die Sicherheit der in diesen Systemen kursierenden Daten.[33]

Anonymität ermöglicht die Verwendung von Daten einer Person ohne Personenbezug und damit eine optimale Umsetzung des datenschutzrechtlichen Ziels der Datensparsamkeit. Sie ist per definitionem nicht umkehrbar. Zur Vermeidung des Zielkonflikts zwischen einer notwendigen Identifizierung des Betroffenen und dem Aspekt der Anonymität, kann auf das Konzept des Pseudonymisierung nach Art. 4 Nr. 5 DS-GVO zurückgegriffen werden. Der Kerngehalt einer Pseudonymisierung besteht darin, den Personenbezug von Daten nicht vollständig, aber immerhin so weit aufzulösen, dass ein Rückschluss auf eine bestimmte Person nur unter Hinzuziehung zusätzlicher Informationen – der Zuordnungsregel – möglich ist[34] Die Pseudonymisierung bietet damit die Möglichkeit zwischen entgegenstehenden Interessen zu vermitteln und Anwendungskonzepte zu entwickeln, bei denen eine Verwendung von Klardaten nicht mehr erforderlich ist.[35] In einer digitalisierten Welt ubiquitärer Datenverarbeitung kann es zum Schutz der Privatheit nachhaltig beitragen, Daten schon zu einem frühen Zeitpunkt zu pseudonymisieren, um damit einen ungewollten Zugriff auf sensible Informationen zu erschweren. Der Gesetzgeber greift die Pseudonymisierung deshalb in Art. 32 Abs. 1 lit. a DS-GVO als eine zentrale Regelmaßnahme der Datenverarbeitungssicherheit heraus. Damit ist stets zu prüfen, ob die Zwecke der Verarbeitung auch durch pseudonymisierte Daten und damit ohne direkten Personenbezug realisiert werden können.[36] Diese personenbezogenen Daten können durch Anwendung unterschiedlicher Werkzeuge wie des Regulären Ausdrucks (RegEx) und Natural Language Processing (NLP) erkannt und somit vereinfacht pseudonymisiert werden. Unterschieden werden muss für die anzuwendenden Verfahren zwischen der Daten-, Datenbank- und Transportebene und damit dem Zustand der Speicherung (Data as Rest), der Übertragung (Data in Transit) und der Bearbeitung (Data in Use), in dem sich die Daten befinden.[37] Technisch kann die Pseudonymisierung auf verschiedene Weise realisiert werden.

Im Ergebnis ist festzuhalten, dass sich die anzuwendenden Pseudonymisierungsverfahren dynamisch nach dem Stand der Technik richten. Die Pseudonymisierungsmechanismen müssen in regelmäßigen Abständen überprüft und sofern notwendig angepasst werden.[38] Der Schutz von Verschlüsselungsverfahren hängt von der Generierung des eingesetzten mathematischen Verfahrens – beispielsweise des Hash-Algorithmus – sowie der Länge des dafür zum Einsatz kommenden Schlüssels ab. Das BSI veröffentlicht regelmäßig Empfehlungen, die Informationen enthalten, welche Algorithmen und Schlüssellängen für welche Anwendungen und bis zu welchem Zeitpunkt als sicher anzusehen sind.[39] Pseudonymisierungsverfahren sind bei Klartextdaten aus kleinen Wertebereichen bzw. mit geringer Streuung innerhalb des Wertebereichs anfällig für eine Aufdeckung durch

---

31 Die Bedingungen sind auf das System an sich bezogen und umfassen somit sowohl die Hard- und Softwarekomponente sowie auch die Netzwerkebene.

32 Das nicht immer konfliktfreie Verhältnis zwischen Datenschutz und Datensicherheit wird treffend als «Security-Privacy-Paradox» bezeichnet: *A Joint Report by The Information and Privacy Commissioner/Ontario and Deloitte & Touche*, The Security-Privacy Paradox: Issues, Misconceptions, and Strategies, 2003.

33 Rossnagel, Handbuch Datenschutzrecht, 2003, Kap. 4.5 Rn. 2.

34 Instruktiv *Article 29 Working Party*, Opinion 05/2014 on Anonymisation Techniques WP216, 2014 24 ff.; Kühling/Buchner, DS-GVO, 2017 Art. 4 Nr. 5 Rn. 2.

35 *Fokusgruppe Datenschutz des Digital-Gipfels*, Whitepaper zur Pseudonymisierung der Fokusgruppe Datenschutz der Plattform Sicherheit, Schutz und Vertrauen für Gesellschaft und Wirtschaft im Rahmen des Digital-Gipfels 2017, 2017, 8.

36 Vgl. Art. 5 Abs. 1 lit. e HS. 1 DS-GVO.

37 ENISA European Union Agency for Network and Information Security, Study on cryptographic protocols, 2014.

38 Plath/Becker, BDSG/DSGVO, Aufl. 2, 2016, § 9 Rn. 58.

39 https://www.bsi.bund.de/DE/Publikationen/TechnischeRichtlinien/technischerichtlinien_node.html.

Aufzählungsangriffe unter Verwendung von Rainbow-Tabellen[40]. Aufgrund dessen bietet es sich an Hashing[41] mit Salt[42] zu verbinden. Dabei wird der Suchraum durch Salts vergrößert. Aber auch bei als sicher geltenden Hash-Funktionen kann keine 100-prozentige Sicherheit erwartet werden. Damit besteht aus technischer Sicht bei Pseudonymisierungsverfahren weiterer Forschungsbedarf, um pseudonymisierte Informationen zu teilen und dennoch dem Datenschutz sowie der Datensicherheit Genüge zu tun. In einem zweiten Schritt bedarf es der Festlegung gewisser Standards. Weitestgehend kann damit auch die Nutzbarkeit der Daten gewährleistet werden, so dass diese zur Generierung des Lagebildes genutzt werden können. Daneben muss eine Vielzahl an Faktoren technischer und organisatorischer Art berücksichtigt werden, um hierbei ein entsprechend nutzbares und gleichzeitig sicheres System zu schaffen.

3. Zusammenfassung

Angesichts der Tatsache, dass das Erkennen eines Sicherheitsvorfalles nicht ohne Erfassung und Auswertung einer Vielzahl an Daten auskommt, muss man sich zwangsläufig mit den geltenden Grundsätzen des Datenschutzes auseinandersetzen. Das Prinzip der Datensparsamkeit gebietet es, weitgehend auf das Verarbeiten personenbezogener Daten zu verzichten bzw. möglichst gering zu halten. Sofern realisierbar, sollten personenbezogene Daten pseudonymisiert abgelegt werden. Der Stand der Technik zeigt jedoch auf, dass einen tatsächlich pseudonymen Datenbestand zu generieren und gleichzeitig alle zugrunde liegenden Informationen zu erhalten, die für eine Weiterverarbeitung notwendig sind, technisch schwer umsetzbar ist. Die Effektivität der Pseudonymisierung hängt davon ab, wie schwierig ein Zugriff auf die Zuordnungsregel ist. Zudem werden die Anforderungen an die Pseudonymisierung in unterschiedlichen Branchen und Verarbeitungssituationen verschieden zu beurteilen sein. Es wäre sinnvoll Regelungen zur Pseudonymisierung in bestimmten Fallkonstellationen in branchenspezifischen Codes of Conduct zu treffen.[43] Unabdingbar sind daher weitere technische und organisatorische Maßnahmen sowie die Kombination mehrerer Verfahren und Methoden. Auf diese Weise kann ein hohes Maß an Sicherheit erreicht werden. Im Ergebnis muss ein Ausgleich zwischen wirtschaftlichem und technischem Innovationspotenzial auf der einen und einer angemessenen Berücksichtigung des Schutzes der informationellen Selbstbestimmung auf der anderen Seite geschaffen werden. OSINT als Mittel zur Detektion und Repression von Sicherheitsvorfällen wird faktisch immer wichtiger und kann unter dem Aspekt der IT-Compliance zwingend sein. Um Vorfälle zu detektieren, werden die sicherheitsrelevanten Informationen nicht nur unternehmensintern verwendet, sondern können auch an andere weitergeleitet werden, beispielweise um andere Unternehmen über präventive Maßnahmen eines relevanten Angriffs zu informieren. Schwachstellen und Sicherheitsvorfälle sind selten spezifisch für ein einzelnes Unternehmen. Das Wissen, wie ein Angriff detektiert und behoben wurde, kann anderen Unternehmen helfen präventiv und repressiv vorzugehen. Den berechtigten Interessen der Betroffenen auf informationelle Selbstbestimmung muss dennoch Rechnung getragen werden. Deshalb sind vor Konzipierung und Einsatz Anforderungen des Datenschutzes zu berücksichtigen und bei der Umsetzung zwingend mit einzubeziehen.

---

[40] Bei diesen Angriffen wird durch einen Angreifer eine Klartext-Pseudonym-Zuordnungstabelle für alle möglichen Klartexte berechnet oder eine vorgefertigte genutzt. Mittels dieser Tabelle können dann gegebenen Pseudonymen die entsprechenden Klartexte zugeordnet werden.

[41] Bei einer Hashfunktion handelt es sich um eine Funktion, die eine beliebig große Eingabemenge auf eine bestimmte Zielmenge abbildet, wobei die Eingabemenge entweder ein einzelnes Merkmal oder eine Reihe von Merkmalen umfassen kann und nicht umkehrbar ist.

[42] Salt bezeichnet in der Kryptografie eine zufällig gewählte Zeichenfolge, die vor der Verwendung an einen gegebenen Klartext als Eingabe einer Hash-Funktion angehängt wird, um die Entropie der Eingabe zu erhöhen.

[43] Art. 40 Abs. 2 lit. d DS-GVO sieht dies ausdrücklich vor.

4. Literaturverzeichnis

A Joint Report by The Information and Privacy Commissioner/Ontario and Deloitte & Touche, The Security-Privacy Paradox: Issues, Misconceptions, and Strategies, 2003.

Article 29 Data Protection Working Party, WP251 – Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, 2018.

Article 29 Working Party, Opinion 05/2014 on Anonymisation Techniques WP216, 2014.

BSI, Die Lage der IT-Sicherheit in Deutschland 2017, 2017.

Däubler, Wolfgang/Klebe, Thomas/Wedde, Peter/Weichert, Thilo, Bundesdatenschutzgesetz – Kompaktkommentar zum BDSG, 5. Aufl., Frankfurt am Main 2016.

Deutscher Bundestag, Antwort der Bundesregierung auf die Kleine Anfrage des Abgeordneten Andrej Hunko, weiterer Abgeordneter, der Fraktion DIE LINKE – Drucksache 18/540 – sowie der schriftlichen Nachfrage, Berlin 05. März 2014.

ENISA European Union Agency for Network and Information Security, Study on cryptographic protocols, 2014.

Fokusgruppe Datenschutz des Digital-Gipfels, Whitepaper zur Pseudonymisierung der Fokusgruppe Datenschutz der Plattform Sicherheit, Schutz und Vertrauen für Gesellschaft und Wirtschaft im Rahmen des Digital-Gipfels 2017, 2017.

Gola, Peter/Klug, Christoph/Körffer, Barbara/Schomerus, Rudolf, BDSG Bundesdatenschutzgesetz – Kommentar, 10. Aufl., München 2015.

Golla, Sebastian J./Hofmann, Henning/Bäcker, Mathias, Connecting the Dots – Sozialwissenschaftliche Forschung in Sozialen Medien im Lichte von DS-GVO und BDSG neu, Datenschutz und Datensicherheit – DuD 2018, S. 89–100.

Kühling, Jürgen/Buchner, Benedikt, DS-GVO – Datenschutzgrundverordnung Kommentar 2017.

Plath, Kai-Uwe/Becker, Thomas, BDSG/DSGVO – Kommentar zum BDSG und zur DSGVO sowie den Datenschutzbestimmungen des TMG und TKG, 2. Aufl. 2016.

Ponemon Institute LLC and IBM Security, 2017 Cost of Data Breach Study, Traverse City, Michigan, USA 06.2017.

Rossnagel, Alexander, Handbuch Datenschutzrecht – Die neuen Grundlagen für Wirtschaft und Verwaltung 2003.

Rückert, Christian, Zwischen Online-Streife und Online-(Raster-)Fahndung – Ein Beitrag zur Verarbeitung öffentlich zugänglicher Daten im Ermittlungsverfahren, Zeitschrift für die gesamte Strafrechtswissenschaft 129 (2017).

Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein, Verkettung digitaler Identitäten, 2007.

Weichert, Thilo, Big Data, Gesundheit und der Datenschutz, Datenschutz und Datensicherheit – DuD 38 (2014), S. 831–838.