Forschungsprojekt EIDI

Effektive Information nach digitalem Identitätsdiebstahl

DELIVERABLE 2.4

Datenschutz-Aspekte bei Identitätsdaten-Sammlungen

ausgearbeitet von

Susan Gonscherowski

UNABHÄNGIGES LANDESZENTRUM FÜR DATENSCHUTZ SCHLESWIG-HOLSTEIN

Kiel, im Juni 2018

Projektpartner

RHEINISCHE FRIEDERICH-WILLHELMS UNIVERISÄT BONN

Informatik 4

Friedrich-Ebert-Allee 144
53113 Bonn

Prof. Dr. Michael Meier
0228 7354249
mm@cs.uni-bonn.de

UNABHÄNGIGES LANDESZENTRUM FÜR DATENSCHUTZ SCHLESWIG-HOLSTEIN

Holstenstr. 98
24103 Kiel

Harald Zwingelberg
0431 98812222
uld6@datenschutzzentrum.de

LEIBNITZ-INSTITUT FÜR INFORMATIONSINFRASTRUKTUR GMBH KARLSRUHE

Hermann-von-Helmholtz-Platz 1
76344 Eggenstein-Leopoldhafen

Prof. Dr. Franziska Boehm
07247 808555
franziska.boehm@kit.edu

UNIVERSITÄT DUISBURG-ESSEN

Allgemeine Psychologie: Kognition

Forsthausweg 2
47057 Duisburg

Prof. Dr. Matthias Brand
0203 3792541
matthias.brand@uni-due.de

XING AG
Dammtorstraße 30
20354 Hamburg

Förderhinweis

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung, und Forschung unter dem Förderkennzeichen 16KIS0697 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

Kurzfassung

In diesem Dokument werden Datenschutz-Aspekte von Identitätsdaten-Sammlungen näher betrachtet. Hierzu wird zunächst eine Typisierung von Datensenken vorgenommen. Anhand sieben generischer Gewährleistungsziele, wie sie auch im Standard-Datenschutzmodell durch die Aufsichtsbehörden vertreten werden, werden erste Anforderungen und mögliche allgemeine Risiken zusammengetragen, die sich aus dem Umgang mit Datensenken durch das EIDI-Projekt ergeben.

Betrachtet wird die Auswirkung öffentlich zugänglicher Datensenken auf die Datenminimierung, Vertraulichkeit, Verfügbarkeit, Transparenz, Intervenierbarkeit, Unverkettbarkeit und die Integrität personenbezogener Daten.

Im zweiten Teil erfolgt, dann eine erste Einschätzung des Umgangs mit Datensenken in Bezug auf das konkrete Projektziel der Benachrichtigung Betroffener und daraus resultierenden Anforderungen an die Umsetzung des Projektvorhabens.

Inhaltsverzeichnis

1 Einleitung

Der Begriff Datensenke bedarf zunächst einer genauen Definition. Unter einer Senke versteht man eine Vertiefung[1] bzw. ein Sammelbecken für beispielsweise Wasser oder Gesteinsablagerungen. Eine Datensenke ist also eine Ansammlung von Daten. Daten sind:

- durch Beobachtung, Messung oder statistische Erhebung gewonnene Werte[2]
- elektronisch gespeicherte Zeichen und Angaben sowie[3]
- vorgegebene Größen und Werte.[4]
- Elemente einer Information in Form von Zeichen oder Code[5]

Daten können unterschiedliche Formen (Ziffern, Symbole, Buchstaben)[6] annehmen. Das *Datenformat* – die Struktur der Daten – bestimmt letztendlich darüber, wie die Daten gelesen bzw. interpretiert werden können. Erst die Interpretation durch eine Maschine verleiht Daten einen Informationsgehalt.

In der Informationstechnologie ist eine Datensenke:

- „Bestimmungsort von übertragenen Daten; Es ist beispielsweise in einem Netz derjenige Teil einer Endeinrichtung, der Daten aufnimmt."[7]
- „[engl. data sink], in einem Netz oder in der Kommunikationstechnik der Empfänger (das Empfangsgerät) einer Nachricht. Auch ein Aufzeichnungsmedium, auf dem sich Daten bis zur Benutzung speichern lassen, heißt Datensenke. Das Gegenstück zur Datensenke ist die Datenquelle."[8]
- „Datensenke ist der Bestimmungsort der Daten, deren Empfangsstelle. Eine Datensenke ist Teil einer Datenendeinrichtung (DEE), die Daten von einem Übermittlungsabschnitt oder einer Übertragungsstrecke empfängt oder auch speichert. Die Übernahme erfolgt normalerweise über eine standardisierte Schnittstelle. Der Ursprungsort der Daten ist die Datenquelle, Data Source."[9]

Entsprechend lässt sich folgende Definition des Begriffs Datensenke festhalten:

> Eine Datensenke ist ein Sammelbecken für Informationen bzw. Daten in Form eines bestimmten Empfangsgeräts oder Speichermediums. Sie wird aus einer oder

[1] http://www.duden.de/rechtschreibung/Senke (zuletzt abgerufen 28.06.2017).
[2] http://www.duden.de/rechtschreibung/Daten (zuletzt abgerufen 12.07.2017).
[3] Ebenda.
[4] Ebd.
[5] http://www.itwissen.info/Daten-data.html (zuletzt abgerufen am 12.07.2017).
[6] Ebd.
[7] http://abc-recht.de/ratgeber/internet/begriffe/datensenke.php (zuletzt abgerufen 28.06.2017).
[8] http://universal_lexikon.deacademic.com/225425/Datensenke (zuletzt abgerufen 28.06.2017).
[9] http://www.itwissen.info/Datensenke-data-sink.html (zuletzt abgerufen 28.06.2017).

mehreren Datenquellen über mindestens eine Schnittstelle gespeist und kann Daten in einem maschinenlesbaren Format empfangen und speichern sowie über die Selbe oder eine andere Schnittstelle für einen Zugriff bereithalten.

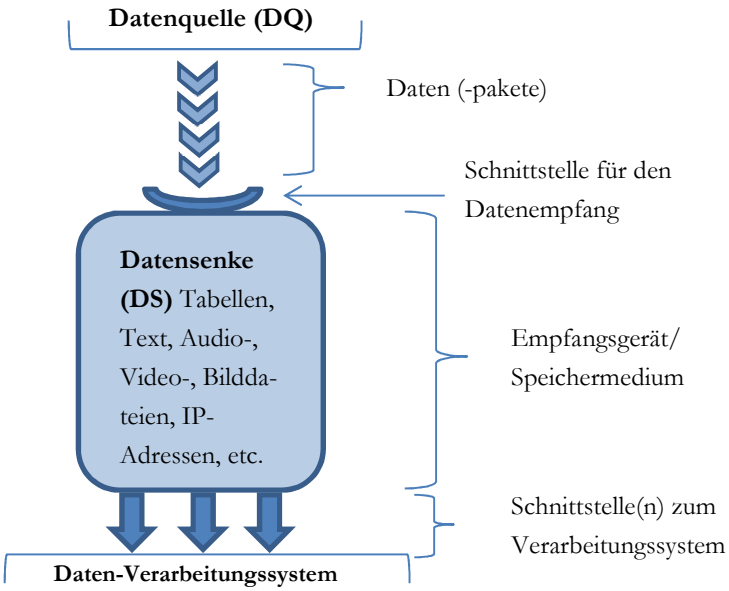


Abbildung 1 Schematische Darstellung einer Datensenke

Physisch besteht eine Datensenke aus einer Schnittstelle zum Empfang von Daten, einem Speicher und einer Schnittstelle zum Verarbeitungssystem. Beispielsweise ist bei einem Diktiergerät das Mikrofon die Empfangsschnittstelle, die Speicherkarte (oder das Tonband) ist das Speichermedium und eine mögliche Schnittstelle zu einem Verarbeitungssystem (bspw. das Ohr) ist der Lautsprecher. Die Datenquelle wäre für diesen Fall die Geräuschquelle, z. B. ein Sprecher. Die Schnittstelle für den Datenempfang und die Schnittstelle zum Verarbeitungssystem können auch identisch sein. Ein Beispiel für einen solchen Fall ist ein USB-Stick. In der Informationstechnologie existiert eine Vielzahl unterschiedlichster Speichermedien, z.B. USB-Sticks, Festplatten, Clouds oder DVDs. Entscheidend ist hier, welche Lesegeräte benötigt werden um an die gespeicherten Daten zu gelangen und wie permanent die Speicherung ist bzw. ob auf die Informationen nur Lesend oder Lesen und Schreibend zugegriffen werden kann.

Datensenken sind damit ein Kernbestandteil jeder Informationstechnologie. Ohne Datensenken gibt es keine Informationstechnologie. Das bedeutet aber auch, dass Datensenken eine besonders sensible Komponente bzw. ein lohnendes Ziel für Angriffe unterschiedlichster Art auf informationstechnische Systeme darstellen. Der Betrieb bzw. die Verwendung einer Datensenke ist daher durch verschiedene Regelungen auf nationaler und europäischer

sowie internationaler Ebene normiert. Hierzu zählen insbesondere das Datenschutzrecht sowie Regelungen zur IT-Sicherheit oder auch Best Practices, z.B. nach ISO. Welche gesetzlichen Regelungen zur Anwendung kommen, wird nicht anhand der Datensenke sondern nach den verarbeiteten Daten bestimmt. Sofern personenbezogene bzw. personenbeziehbare Daten erhoben oder verarbeitet werden, ist das Datenschutzrecht anzuwenden. Nur so kann ein zulässiger Betrieb einer Datensenke erfolgen. Eine Fehleinschätzung der Datenkategorien kann unvorhergesehene Risiken für die Rechte und Freiheiten natürlicher und juristischer Personen mit sich bringen.

1.1 Differenzierung verschiedener Typen von Datensenken

Um das Gefährdungspotential einer Datensenke beurteilen und den Umgang mit der Datensenke rechtskonform gestalten zu können, muss klar sein, ob anonyme oder personenbezogene Daten enthalten sind und ob die Daten überhaupt verarbeitet – also erhoben, gespeichert, kombiniert etc. – werden dürfen (Abb. 2).

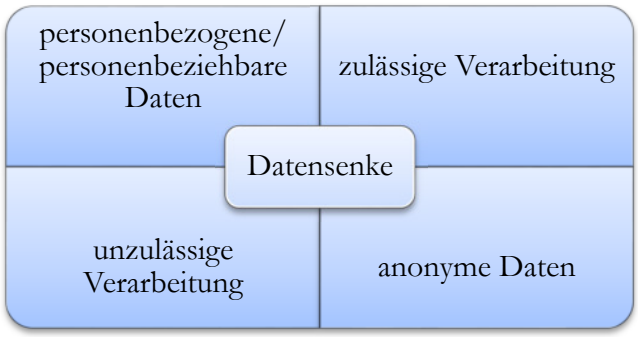
Die Unterscheidung von personenbezogenen/ personenbeziehbaren Daten und nicht personenbezogenen (anonymen) Daten spielt für den anzuwendenden Rechtsrahmen eine entscheidende Rolle. Sind personenbezogene Daten enthalten, ist automatisch Datenschutzrecht anwendbar. Hierbei ist zu beachten, dass durch die Verkettung von Datensätzen diese Unterscheidung prinzipiell aufhebbar ist. Auch das Bundesverfassungsgericht (BVerfG) hat im Volkszählungsurteil bereits 1983 festgestellt, dass es kein belangloses Datum gibt. Das Gericht argumentierte in Bezug auf den Umgang mit Daten weiter:

„Entscheidend sind ihre Nutzbarkeit und Verwendungsmöglichkeit. Diese hängen einerseits von dem Zweck, dem die Erhebung dient, und andererseits von den der Informationstechnologie eigenen Verarbeitungsmöglichkeiten und Verknüpfungsmöglichkeiten ab."[10]

Hieraus leitet sich der zweite Faktor zur Bestimmung der Datensenke ab – die Zulässigkeit der Verarbeitung innerhalb der Datensenke (Abb. 2). Für anonyme Daten gibt es vor allem bereichsspezifische Einschränkungen in Bezug auf die Zulässigkeit der Verarbeitung, z.B. im Urheber- und Strafrecht. Im Gegensatz dazu ist eine Verarbeitung im Datenschutzrecht grundsätzlich erst einmal verboten es sei denn, diese wird durch ein Gesetz oder eine Einwilligung legitimiert.

[10] Urteil vom 15. Dezember 1983 - Az. 1 BvR 209/83, Rn 176 (openjur), über: http://openjur.de/u/268440.html (zuletzt abgerufen am: 27.06.18).

Abbildung 2 Matrix Datensenken



Je nach Kombination der einzelnen Elemente der Matrix ergeben sich vier Grundtypen von Datensenken.

1.1.1 Typ 1: Zulässige Verarbeitung anonymer Daten

Anonyme Datensätze sind vom Datenschutzrecht nicht erfasst, da dieses nur auf personenbezogene bzw. personenbeziehbare Daten Anwendung findet. Das Risiko eines Identitätsdiebstahls ist hier sehr gering, vorausgesetzt die Anonymisierung der Daten wurde entsprechend sorgfältig vorgenommen. Zum Typ 1 zählen vor allem Statistiken, wie Zensusdaten, Wahlergebnisse oder Umfrageergebnisse. Beispiele für Typ 1 Datensenken sind:

- Zensus 2011 - Private Haushalte nach Wohnraumfläche[11]
- Personenbeförderungsstatistiken in Deutschland[12]
- Wahlbeteiligung und Stimmanteile bei Bundestagswahlen in Schleswig-Holstein[13]

Nichtsdestotrotz können auch anonyme Daten aus Typ 1-Datensenken durch Verkettung mit anderen Datensätzen einen Personenbezug ermöglichen oder wenigstens erleichtern.[14] In diesem Zusammenhang sei auf den stetig wachsenden Big Data-Bereich verwiesen. Neue Algorithmen und Verarbeitungsverfahren ermöglichen immer neue Auswertungen sehr großer Datenmengen, die eine sichere und dauerhafte Anonymisierung erschweren.[15]

[11] Zensus-Ergebnisse, über: https://ergebnisse.zensus2011.de/#StaticContent:00,GWZ_4_1_2,m,table (16.11.17).
[12] Personenbeförderungsstatistiken, über: https://www.destatis.de/DE/ZahlenFakten/Wirtschaftsbereiche/TransportVerkehr/Personenverkehr/Tabellen/BefoerdertePersonen.html (16.11.17).
[13] Wahlbeteiligung in Schleswig-Holstein, über: http://www.kas.de/upload/wahlen/wahlergebnisse/sh_wahlen.pdf (16.11.17).
[14] NDR: Nackt imNetz – Millionen Nutzer ausgespäht (vom 03.11.2016), über: http://www.ndr.de/nachrichten/netzwelt/Nackt-im-Netz-Millionen-Nutzer-ausgespaeht,nacktimnetz100.html (abgerufen am 16.11.17).
[15] S. Timpf: Big Data. Über: https://www.boeckler.de/pdf/timpf_bigdata_6.pdf (zuletzt abgerufen am 23.05.17).

1.1.2 Typ 2: Unzulässige Verarbeitung anonymer Daten

Datensenken bergen insbesondere dann ein hohes Gefährdungspotential, wenn sie illegal Daten verarbeiten, da hier gesetzliche Regelungen fahrlässig oder auch gezielt umgangen werden. Die Verarbeitung anonymer Daten, wie beispielsweise Musik, Filme oder Software, die ohne die Genehmigung des Rechtinhabers über Datensenken zur Verfügung gestellt oder gespeichert werden, kann diesen in seinen Rechten verletzen. Beispiele für Typ 2 Datensenken sind:

- Illegale Tauschbörsen für Filme und Musik
- Illegale Weitergabe von Firmengeheimnissen durch Angreifer (Industriespionage)

1.1.3 Typ 3: Zulässige Verarbeitung personenbezogener Daten

Typ 3-Datensenken werden durch das Datenschutzrecht reguliert. Es handelt sich hierbei um Datensenken, die zu legitimen Zwecken angelegt werden. Sie werden aus legalen Quellen gespeist und enthalten personenbezogene oder personenbeziehbare Daten. Grundlage der Datenverarbeitung ist entweder ein Gesetz oder die Einwilligung des Betroffenen. Je nach Kategorien der verarbeiteten Daten stellt der Gesetzgeber zudem mehr oder weniger hohe Anforderungen an die Sicherheit und die Systeme der Datenverarbeitung. Jede Datenverarbeitung ist dabei an die in Art. 5 Datenschutz-Grundverordnung (DSGVO)[16] dargelegten Grundsätze, wie beispielsweise Transparenz, Zweckbindung und Richtigkeit gebunden. Neben den Pflichten der Verantwortlichen sind im Gesetzestext auch die Rechte der Betroffenen festgehalten. Dreh- und Angelpunkt des deutschen Datenschutzrechts ist das 1983 vom BVerfG entwickelte Recht auf informationelle Selbstbestimmung, welches sich auch auf europäischer Ebene in Art. 8 EuCh wiederfindet. Auch eine Veröffentlichung personenbezogener Daten in frei zugänglichen Datensenken ist legitim solange die Verarbeitung (Erhebung, Speicherung, Veröffentlichung der Daten) mit Einwilligung des Betroffenen oder auf Grundlage eines Gesetzes stattfindet. Beispiele für Typ 3-Datensenken sind heute allgegenwärtig und schließen insbesondere jede Form von Nutzerkonten bei Onlinediensten ein:

- Soziale Netzwerke, wie Facebook, Twitter, Instagram
- Öffentliche Verzeichnisse, wie Telefonnummernverzeichnisse, öffentliche Schlüsselserver, Auskunfteien
- Staatliche Quellen, wie Melderegister

[16] Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG.

1.1.4 Typ 4: Unzulässige Verarbeitung personenbezogener/ personenbeziehbarer Daten

Zu diesem Typ zählen alle Datensenken, die nicht in Zusammenhang mit einer rechtmäßigen Datenverarbeitung stehen. Das Hauptmerkmal zur Unterscheidung legaler von illegalen Datensenken ist die Rechtsgrundlage der Verarbeitung. Im Datenschutzrecht gilt ein grundsätzliches Verbot mit Erlaubnisvorbehalt gem. Art. 6 DSGVO. Findet die Verarbeitung also ohne Einwilligung des Betroffenen statt und ist auch nicht durch eine Rechtsvorschrift legitimiert, handelt es sich um eine illegale Datensenke. Das Recht auf informationelle Selbstbestimmung wird hier verletzt, denn der Betroffene hat keine Möglichkeit mehr, selbst zu bestimmen wer wann was über ihn weiß. Typ 4 Datensenken sind zum Beispiel:

- Illegale Kopien von Daten nach Einbruch auf einem Server[17]
- Schadsoftware auf Endgeräten die Daten an eine anderen Stelle übermittelt[18]
- Unentdeckte Sicherheitslücken, die unerwünschte Kenntnisnahme ermöglichen[19]
- Datenschutzunfreundliche Voreinstellungen, die zu unerwünschter Datenerhebung führen[20]

Die Datensenken vom Typ 4 sind es, die durch EIDI ausgewertet werden sollen. Eine erste Übersicht über einschlägige Datensenken, die die Grundlage für die weitere Entwicklung des Projekts und insbesondere des Werkzeugs zur Benachrichtigung Betroffener darstellt, wurde durch den Projektpartner ITS:Info 4 zusammengestellt.[21]

1.2 Personenbezogene Daten in Datensenken

Datenschutzrechtlich betrachtet, ist die entscheidende Frage, ob gemäß Art. 2 Abs. 1 DSGVO personenbezogene Daten in der Datensenke enthalten sind. [22] *Personenbezogene Daten* sind entsprechend der Legaldefinition der seit Mai 2018 anzuwendenden DSGVO „alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person […] beziehen“ (Art. 4 Abs. 1 DSGVO). Im Sinne der Verordnung gilt eine Person

[17] https://www.heise.de/security/meldung/Hetzner-gehackt-Kundendaten-kopiert-1884180.html (16.11.2017).
[18] https://www.heise.de/security/meldung/Adware-Trojaner-rooten-heimlich-Android-Geraete-2878360.html (16.11.2017).
[19] https://www.heise.de/security/meldung/Ministerium-fuer-digitale-Infrastruktur-pfuscht-beim-eigenen-Web-Server-3186960.html (16.11.2017).
[20] https://www.heise.de/newsticker/meldung/Abgegriffene-Browserdaten-Mozilla-entfernt-Web-of-Trust-3455990.html (16.11.2017).
[21] Malderle/Wübbeling: Dienstekatalog gängiger Datensenken und dort öffentlich verfügbarer Identitätsdaten-Sammlungen (D 2.1), August 2017, S. 15 ff.
[22] Aus Gründen der Übersichtlichkeit wird im Folgenden auf einen Bezug zum BDSG verzichtet. Die DSGVO trat bereits 2016 in Kraft und gilt seit dem 25. Mai 2018 unmittelbar in allen Mitgliedstaaten der EU.

als identifizierbar, wenn sie „direkt oder indirekt, insbesondere mittels Zuordnung zu einer Kennung […], zu einer Kennnummer, zu Standortdaten, zu einer Online-Kennung oder zu einem oder mehreren besonderen Merkmalen […] identifiziert werden kann“ (Art. 4 Abs.1 DSGVO). Sind in einer Datensenke beispielsweise Namen (Kennung), Kreditkarten- oder Telefonnummern (Kennnummer), Adressen (Standortdaten) und/oder E-Mail-Adressen (Online-Kennung) enthalten, handelt es sich um personenbezogene Daten im Sinne der DSGVO. Die Aufzählung ist jedoch nicht abschließend. Je nach Kontext sind weitere Daten oder Kombinationen von Daten denkbar, die eine Person identifizierbar machen. Wie oben bereits erwähnt, verneint das BVerfG die Existenz belangloser Daten.[23] Endscheidend sind die Mittel und der Aufwand, die erforderlich sind um eine Verbindung zwischen den vorliegenden Daten und einer natürlichen Person herstellen zu können.[24]

Sofern festgestellt wurde, dass eine Datensenke personenbezogene Daten (Typ 3 oder Typ 4) enthält, gelten grundsätzlich die Vorschriften der DSGVO. Ausgenommen hiervon sind jedoch personenbezogene Daten, die von zuständigen Behörden zum Zwecke der Verhütung, Ermittlung, Aufdeckung oder Verfolgung von Straftaten oder der Strafvollstreckung, einschließlich des Schutzes vor und der Abwehr von Gefahren für die öffentliche Sicherheit verarbeitet werden. Für diesen Bereich regelt die JI-Richtlinie[25] die polizeiliche und justizielle Zusammenarbeit in der EU, die wiederum durch die Mitgliedstaaten auf nationaler Ebene umgesetzt wird.

Die Grenze zwischen einem anonymen und einem personenbeziehbarem Datum ist abhängig von den Mitteln und Möglichkeiten desjenigen, der diesem Bezug herzustellen versucht. In Erwägungsgrund (ErwG) 26 heißt es:

„Um festzustellen, ob eine natürliche Person identifizierbar ist, sollten alle Mittel berücksichtigt werden, die von dem Verantwortlichen oder einer anderen Person nach allgemeinem Ermessen wahrscheinlich genutzt werden, um die natürliche Person direkt oder indirekt zu identifizieren […]“

So können Kommentare auf einer Webseite, z.B. bei einem Magazin, durchaus ohne Login auf der Seite und/oder die Angabe eines (Benutzer-)Namens möglich sein. Ist ein Login erforderlich, aber der (Benutzer-)Name erscheint nicht in Zusammenhang mit dem Kommentar, ist der Bezug für den Verantwortlichen (Betreiber der Seite) herstellbar für die Leser jedoch nicht. Der Verantwortliche verfügt also über Mittel und Möglichkeiten den Personenbezug zwischen den Nutzern und dem Autor eines Kommentars herzustellen, die dem

[23] BVerfG, 1 BvR 209/83 vom 15.12.1983 – Volkszählung.
[24] Klar/Kühling, in: Kühling/Buchner, DSGVO, Art. 4 Rn. 20.
[25] Richtlinie (EU) 2016/680 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten durch die zuständigen Behörden zum Zweck der Verhütung, Ermittlung, Aufdeckung oder Verfolgung von Straftaten oder der Strafvollstreckung sowie zum freien Datenverkehr und zur Aufhebung des Rahmenbeschlusses 2008/977/JI des Rates.

Leser erst einmal nicht zur Verfügung stehen. Nichtsdestoweniger kann auch ein anderer Nutzer (Leser) den Bezug zwischen einem Kommentar und einem anderen Nutzer herstellen, wenn beispielsweise ein gesetzlicher Anspruch (z.B. Unterlassungsanspruch) auf die zusätzlichen Informationen besteht und der Verantwortliche die Daten herausgeben muss. Der Europäische Gerichtshof (EuGH) hat diesbezüglich klargestellt, dass ein personenbezogenes Datum auch dann vorliegt, wenn die „Beziehbarkeit" der Hilfe eines Dritten bedarf.[26]

Zu bedenken ist, dass diese Form der Asymmetrie zwischen Individuen und Organisationen in Bezug auf die Verarbeitung personenbezogener Daten immer besteht.[27] Organisationen verfügen in der Regel über zusätzliche Informationen, z.B. IP-Adressen, E-Mail- und Lieferadressen, da sie die Kontrolle über die genutzten Systeme und angebotenen Dienste ausüben. Abhängig vom ursprünglichen Verantwortlichen können Identitätsdatendiebe Datensenken mit sehr unterschiedlichen personenbezogenen Daten befüllen. Je sensibler diese Daten sind desto lohnenswerter ist die unrechtmäßige Veröffentlichung bzw. Übermittlung für den Identitätsdatendieb.

In Bezug auf die von EIDI herangezogenen Datensenken ist daher kaum vorhersehbar, welche Kategorien personenbezogener Daten vorhanden sind. Diese können unterschiedlich sensibel sein. Ihr Inhalt ist einerseits abhängig vom ursprünglichen Zweck der Datenverarbeitung. Andererseits spielen auch Faktoren, wie vom Verantwortlichen getroffene IT-Sicherheits- und Datenschutzmaßnahmen eine Rolle. Die vom ursprünglichen Verantwortlichen getroffenen Maßnahmen haben maßgeblich Einfluss auf die möglichen Schäden, die aus einem Datenschutzvorfall für den Betroffenen folgen können.

1.3 Kategorien personenbezogener Daten

Der Grad der Sensibilität eines personenbezogenen Datums wird subjektiv durch den Einzelnen und objektiv durch den Gesetzgeber unterschiedlich bewertet.

Die Fülle personenbezogene Daten, die beispielsweise von Individuen tagtäglich öffentlich in sozialen Netzwerken zur Verfügung gestellt werden, zeigt deutlich, dass die Einschätzung der Sensibilität eines Datums von Nutzer zu Nutzer variiert. Einige veröffentlichen über Jahre jeden Tag ein Selbstportrait, andere vermeiden es auf Fotos erkennbar zu sein.

Erste Auswertungen von Typ 4-Datensenken haben ergeben, dass in den meisten öffentlich zugänglichen Datensenken E-Mail-Adressen und Passwörter – gehasht oder im Klartext – enthalten sind (D 2.1, S. 35). Für den weiteren Projektverlauf stellt diese Kombination aus E-Mail-Adresse und Passwort den Standardanwendungsfall dar.

[26] EuGH (Breyer) C-582/14; EU:C:2016:779, Rn. 48f.
[27] Rost/ Bock: Impact Assessment im Lichte des Standard-Datenschutzmodells, in: DuD Nr. 10 (2012), S. 743.

Da eine E-Mail-Adresse eine standardisierte Struktur hat, eignet sie sich als Parameter zum Auffinden relevanter Datensätze in Datensenken. Aufgrund der großen Menge an Daten, die in Datensenken enthalten sind, muss eine Vorauswahl getroffen werden (D 2.1, S. 36). E-Mail-Adressen sind als Suchparameter für das EIDI-Projekt gut geeignet. Neben der erwähnten vorgegebenen Struktur handelt es sich bei einer E-Mail-Adresse um ein eindeutig identifizierendes Merkmal, das für sich genommen jedoch nicht zur Kategorie besonderer personenbezogener Daten nach Art. 9 DSGVO zählt. Außerdem besteht über die E-Mail-Adresse die Möglichkeit einer Benachrichtigung des betroffenen Adressinhabers. Für die Weiterentwicklung des EIDI-Frameworks ist es denkbar weitere Suchparameter, die über ähnliche Eigenschaften wie E-Mail-Adressen verfügen, aufzunehmen.

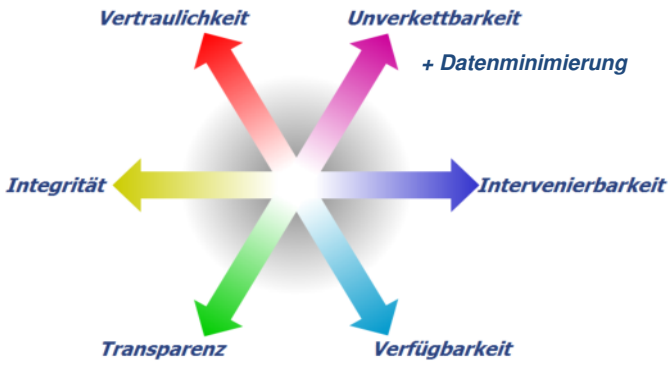
2 Datensenken und Gewährleistungsziele

Das Standard-Datenschutzmodell[28] mit seiner Systematik ermöglicht es datenschutzrelevante Aspekte von Datensenken anhand sieben generischer Schutzziele (Abb. 3) zu beurteilen. Missachtet eine Datenverarbeitung eines oder mehrere dieser Schutzziele, erhöht sich das Risiko, dass Rechte und Freiheiten natürlicher Personen verletzt werden und die Wahrscheinlichkeit, dass der Verantwortliche gegen geltendes Datenschutzrecht oder diesem übergeordnete Spezialnormen verstößt, erhöht sich entsprechend.

Im Folgenden sollen Datenschutzaspekte von Datensenken mit Hilfe der Gewährleistungsziele analysiert werden. Grundlage hierfür ist die erste Untersuchung von im Internet frei zugänglichen Datensenken (D2.1) des Projektpartners ITS:Info4.

[28] Siehe AK Technik: Das Standard-Datenschutzmodell V. 1.1 (Stand April 2018), über: https://www.datenschutzzentrum.de/uploads/sdm/SDM-Methode_V1.1.pdf, S. 11 ff.

Abbildung 3 Schutzziele



Quelle: ULD

2.1 Datensparsamkeit

Typ-4-Datensenken sind nicht auf Datensparsamkeit ausgelegt. In diesen Datensenken sind grundsätzlich Daten enthalten, die nicht für die Veröffentlichung bestimmt sind. Dieser Umstand steht dem Prinzip der Datensparsamkeit, die an die Erforderlichkeit und Zweckbestimmung personenbezogener Daten gebunden ist, entgegen. Je umfangreicher eine einzelne Datensammlung und je sensibler die Daten sind, desto lukrativer ist die Sammlung für einen kriminellen.

2.2 Vertraulichkeit

Der Standardanwendungsfall zeigt bereits, dass das Schutzziel der Vertraulichkeit in Typ-4-Datensenken nicht gewährleistet werden kann. Zwar ist eine E-Mail-Adresse als Kontaktinformation grundsätzlich dazu bestimmt anderen bekannt zu sein, die Zulässigkeit der Weitergabe bzw. Veröffentlichung ist jedoch davon Abhängig ob der Inhaber hiermit einverstanden ist.

Ein weiterer Aspekt ist, dass E-Mail-Adressen neben der Kommunikation heute vielfach auch die Funktion des Benutzernamens übernehmen. In Kombination mit dem zugehörigen Passwort dienen E-Mail-Adressen zur Authentifizierung gegenüber einer Vielzahl von Dienstanbietern. Datensenken mit entsprechendem Inhalt gefährden nicht nur die Vertraulichkeit der E-Mail-Adresse sondern mittelbar auch die Vertraulichkeit der personenbezogenen Daten, die in den jeweiligen Nutzerkonten hinterlegt sind.

Neben dem Standardanwendungsfall können in Datensenken auch direkt weitere personenbezogene Daten enthalten sein, die sensibel und damit in der Regel auch vertraulich sind. Hierzu gehören neben Kreditkartennummern bzw. Zahlungsinformationen (IBAN, Konto-

nummer, Bankleitzahl) auch Pass-, Sozialversicherungsnummern. Die Veröffentlichung besonderer Kategorien personenbezogener Daten, wie bspw. der sexuellen Orientierung oder politischer Ansichten, kann für die Betroffenen möglicherweise lebensgefährlich sein.[29] So ist Ehebruch oder Homosexualität in einigen Staaten, wie Indien[30]und dem Iran[31], eine Straftat auf die teilweise die Todesstrafe steht.

Auch das Vorhandensein weniger sensibler Daten, wie Adressen und Telefonnummern, kann eine Verletzung der Vertraulichkeit dieser Daten darstellen, auch wenn diese Daten z.T. sogar in Online-Telefonbüchern rechtmäßig veröffentlicht wurden. Anhand der Datensenke wird dies in der Regel nicht nachvollziehbar sein, denn hierzu müsste die Datenquelle bekannt und/oder selbst öffentlich zugänglich sein.

2.3 Verfügbarkeit

Die Verfügbarkeit von Datensenken schwankt. Für Kriminelle sind Datensenke der „Marktplatz“ auf dem die erbeuteten Daten gehandelt werden – sei es gegen Geld oder im Austausch gegen andere Daten. Staatsgrenzen (und damit Zollbestimmungen) existieren in diesem virtuellen Raum nicht. Erfährt eine Hoster oder der Betreiber einer Plattform davon, dass über seine Infrastruktur illegaler Datenhandel betrieben wird, bestimmt sich nach dessen Sitz, ob gesetzliche Regelungen eine Unterbindung dieser Handlungen fordern. In der EU und auch in den USA werden sind die Provider gemäß des Notice and Take Down-Prinzips daran gebunden illegale Aktivitäten mit Hilfe ihrer Systeme zu verhindern sobald sie davon Kenntnis haben, andernfalls sind sie für Folgeschäden mit haftbar. Dieser Umstand führt dazu, dass die Verfügbarkeit von Datensenken nicht dauerhaft gewährleistet werden kann.

Auch die Beschlagnahme der erforderlichen Hardware (Server) durch Behörden führt dazu, dass eine Datensenke nicht mehr erreichbar ist. Eine Möglichkeit die Betroffenen dennoch zu informieren besteht darin, dass die beschlagnahmten Daten EIDI zur Verfügung gestellt werden.

2.4 Transparenz

Das Transparenzgebot ist schon grundsätzlich schon dadurch verletzt, dass der Betroffene nicht weiß, dass Daten in die Datensenke gelangt sind und öffentlich zur Verfügung stehen.

---

[29] Vgl. Ashley Madison-Leak 2015, über: https://www.wochenblatt.de/ueberregionale-artikel/regensburg/artikel/127471/seitensprung-des-ehemanns-durch-ashley-madison-leak-entlarvt-gerechte-strafe (zuletzt abgerufen am 18.06.18).

[30] Vgl. Amnesty International Report Indien 2016 zu den Rechten von Lesben, Schwulen, Bisexuellen, Transgeschlechtlichen und Intersexuellen, über: https://www.amnesty.de/jahresbericht/2016/indien-0#section-11115

[31] Vgl. Amnesty International Report Iran 2017/2018 zur Anwendung der Todesstrafe, über: https://www.amnesty.de/jahresbericht/2018/iran#section-1723137 (zuletzt abgerufen am 18.06.18).

Die Auswertung der ersten Datensenken zeigt, dass die Leaks in Form von Textdateien veröffentlicht werden. In einigen Fällen ist die Quelle bekannt, da die Infiltration der Systeme vom Hacker als Methode der Reputation genutzt wird. In vielen Fällen ist jedoch (zunächst) nicht bekannt woher die Daten stammen.

Auch die Leaks selbst sind auf Grund ihrer Struktur undurchsichtig. Die Trennzeichen innerhalb eines Leaks variieren stark und die Auswertung von Datensenken wird dadurch erschwert.

2.5 Intervenierbarkeit

Eines der größten Datenschutzprobleme von Typ-4-Datensenken ist die fehlende Intervenierbarkeit. Weder die Rechte der ursprünglich verantwortlichen Stelle noch die der Betroffenen werden gewährleistet. Selbst wenn die unerlaubte Veröffentlichung durch eine Meldung der verantwortlichen Stelle oder durch eigene Recherchen (Leak-Checker etc.) bekannt ist, scheitert die konsequente Durchsetzung der Betroffenenrechte i.d.R. daran, dass sich die Identitätsdatendiebe bzw. Hehler dem Zugriff entziehen und datenschutzrechtlichen Verpflichtungen nicht nachkommen.

Die Intervenierbarkeit wird auch dadurch beeinträchtigt, dass das Melde- und Benachrichtigungssystem für Datenschutzvorfälle nicht alle Vorfälle abdeckt. Die in der DSGVO enthaltenen Normen greifen nur wenn die Verantwortliche Stelle (Datenquelle) bemerkt, dass es einen IT-Sicherheits- bzw. Datenschutzvorfall gab, sich für die Betroffenen ein hohes Risiko daraus ergibt und der Verantwortliche in der Lage ist zu bestimmen, wer die Betroffenen sind.

Die Tatsache, dass eine Verantwortlicher einen Vorfall bemerkt, schließt im besten Fall die Lücke bei diesem. Nichtsdestotrotz können über Stunden, Tage oder Monate Daten abgeflossen sein. Ein Schließen der Sicherheitslücke verhindert zwar ein weiteres Abfließen von Daten, holt diese jedoch nicht vom Angreifer zurück. Wurden E-Mail-Adressen und Passwörter entwendet, sind diese in jedem Fall kompromittiert. Sind personenbezogen Daten betroffen, die sich – im Gegensatz zu einem Passwort – nicht ändern lassen kompromittiert, bspw. Sozialversicherungsnummern, bedeutet dies für den Betroffenen einen nicht heilbaren Schaden.

Hat der Verantwortliche Maßnahmen ergriffen den Datenzugriff, bspw. durch Verschlüsselung zu verhindern, entfällt die Benachrichtigungspflicht an die Betroffenen ganz, da bei Verschlüsselung nach Stand der Technik kein hohes Risiko mehr besteht. Hierbei wird jedoch dem Umstand des technischen Fortschritts nicht ausreichend Rechnung getragen. Das Risiko für die Betroffenen steigt mit fortschreitender Entwicklung. Datenschutzvorfälle, die zunächst kein Risiko darstellten, können von einem Tag auf den anderen zum Risiko für die

Rechte und Freiheiten der Betroffenen werden. Eine erneute Benachrichtigungspflicht ergibt sich daraus jedoch nicht.

Ein weiteres Problem ist die Identifizierung der Betroffenen. Hier ist nicht nur die Identifizierung der Betroffen durch den Verantwortlichen ein Problem, auch der Umstand, dass es Möglicherweise gar keinen Verantwortlichen gibt, der zur Benachrichtigung verpflichtet ist, wurde vom Gesetzgeber nicht bedacht. Opfern von Phishing-Angriffen fallen regelmäßig in diese Kategorie von Datenschutzvorfällen. Aber auch die Insolvenz oder Liquidation einer verantwortlichen Stelle verhindert möglicherweise eine Benachrichtigung.

2.6 Unverkettbarkeit

Die Kombination verschiedener Leaks kann dazu führen, dass Datensätze von verschiedenen Stellen zu einer Person miteinander verknüpft werden. Da E-Mail-Adressen und auch Passwörter häufig für mehrere Dienste verwendet werden, kann dies dazu führen, dass die bei einem Dienst verschlüsselt hinterlegten Daten durch einen Leak bei einem andren Dienst doch kenntlich werden. Hinzu kommt, dass auch durch einen Angriff selbst Daten erhoben und kombiniert[32] werden können. Folgen solcher Angriffe können Deanonymisierung und Auflösung von Pseudonymen oder der Zugriff auf eigentlich verschlüsselte Daten sein[33], woraus sich wiederum Folgenschäden für die Betroffenen ergeben können.

2.7 Integrität

Eine Integritätsverletzung ist für Betroffene insbesondere dann ein Risiko, wenn ihre Reputation beeinträchtigt wird. Dies steht in engem Zusammenhang mit unzulässiger Verkettung personenbezogener Daten. An Stelle des Risikos durch die Offenlegung personenbezogener Daten eines Betroffenen, ergibt sich das Risiko hier durch eine falsche Zuordnung der Identitätsdaten zu einem Betroffenen und daraus gezogener falscher Schlussfolgerungen. So kann eine vom Betroffenen nach einem Datenschutzvorfall aufgegebene E-Mail-Adresse vom E-Mail-Provider neu vergeben werden. Der „Erbe“ der kompromittierten Adresse hat hiervon keine Kenntnis, wird möglicherweise jedoch mit den veröffentlichten Daten in Verbindung gebracht.[34] Dies kann vor allem in Verbindung mit besonderen Kategorien personenbezogener Daten, wie Gesundheitsdaten (HIV-Status), politische Überzeugungen (Oppositionelle) und sexuelle Orientierung (Homosexualität), die Rechte und Freiheiten Betroffener einschränken.

---

[32] Siehe hierzu: http://www.ndr.de/nachrichten/netzwelt/Nackt-im-Netz-Millionen-Nutzer-ausgespaeht,nacktimnetz100.html

[33] Eikenber, R.: Passwort-Zugriff. Heartbleed-Lücke mit katastrophalen Folgen (09.04.14), über: https://www.heise.de/security/meldung/Passwort-Zugriff-Heartbleed-Luecke-mit-katastrophalen-Folgen-2166861.html (zuletzt abgerufen am: 18.06.18).

[34] Spiegel Online: Yahoo vergibt ungenutzte Accounts neu (vom 16.03.2013), über: http://www.spiegel.de/netzwelt/web/e-mail-adressen-yahoo-vergibt-ungenutzte-accounts-neu-a-911335.html (zuletzt abgerufen am 25.06.18).

3 Umgang mit Datensenken im Projekt EIDI

Für das Projekt EIDI gilt es die Zielvorgaben des Projekts mit den verschiedenen Aspekten des Datenschutzrechts in Einklang zu bringen.

Zu betrachten sind 1) Auffinden der Datensenken, 2) die Bewertung der Qualität der Daten, 3) die Identifizierung der Opfer, 4) die Information der Betroffenen und 5) die Verarbeitung nachdem die Betroffenen informiert wurden. Diese Punkte sind konkrete Zielvorgaben von EIDI, die einer rechtlichen Einordnung bedürfen.

3.1 Datenschutzrecht im Umgang mit Datensenken

Die wesentlichen Prinzipien im Umgang mit personenbezogenen Daten sind in Art. 5 Abs. 1 DSGVO dargelegt. Hierzu zählen u.a. Transparenz, Zweckbindung, Datenminimierung, Integrität, Erforderlichkeit und Vertraulichkeit. Die teilweise Übereinstimmung mit den in der Schutzzielsystematik verwendeten Begriffen ist kein Zufall. Alle sieben Schutzziele lassen sich auf verschiedene Artikel und Erwägungsgründe der DSGVO übertragen. Eine datenschutzkonforme Gestaltung des EIDI-Verfahrens ist dementsprechend anhand der Einhaltung der Schutzziele nachvollziehbar.

3.1.1 Auffinden der Datensenken

Ausgehend von Standardanwendungsfall werden für EIDI öffentlich zugängliche Datensenken genutzt. Grundsätzlich steht die Information aus öffentlich zugänglichen Quellen jedermann frei.[35] Dies schließt grundsätzlich auch Quellen mit ein, die nicht veröffentlicht werden sollten, z.B. im Rahmen des investigativen Journalismus. Hierbei überwiegt dann das öffentliche Interesse an der Veröffentlichung der Informationen und die Interessen der Betroffenen sind diesem Unterzuordnen.[36]

Stoßen Behörden im Rahmen von Ermittlungen auf eine Datensenke, die personenbezogene Daten (Identitätsdaten) enthält, findet eine Verarbeitung der Daten in Ausübung hoheitlicher Aufgaben gemäß Art. 6 Abs. 1 lit. e) DSGVO statt. Danach liegt die Verarbeitung (Kenntnisnahme) der personenbezogenen Daten, die in der Datensenke enthalten sind, im öffentlichen Interesse und ist damit erforderlich. Ob die Daten an eine andere öffentliche oder

[35] Erfurter Kommentar/Dieterich, 10. Aufl. 2010, GG Art. 5 Rn. 13-14.
[36] Siehe zum Beispiel Süddeutsche Zeitung: Panama-Papers, über: http://panamapapers.sueddeutsche.de/ (zuletzt abgerufen am 25.06.18).

nicht-öffentliche Stelle übermittelt werden dürfen hängt von den Kompetenzen der übermittelnden Behörde ab.

Die Informationsfreiheit, die sich aus Art. 5 Grundgesetzt (GG) ergibt, und das Recht auf informationelle Selbstbestimmung aus Art. 2 Abs. 1 i.V.m. Art. 1 Abs. 1 GG sind bei der Betrachtung und rechtlichen Einordnung von Datensenken immer gegeneinander abzuwägen. Bei öffentlich zugänglichen personenbezogenen Daten, wie sie in Datensenken abgelegt werden, besteht kein öffentliches Interesse an den Daten. Es kann somit nicht argumentiert werden, dass der unbeschränkte Zugriff auf die Daten automatisch die Rechtmäßigkeit der Verarbeitung zur Folge hat.[37] Vielmehr ergibt sich aus dem konkreten Interesse des Einzelnen an der Löschung der unrechtmäßig veröffentlichten Daten zugleich ein – die Verarbeitung möglicherweise rechtfertigendes – öffentliches Interesse.[38] Inwiefern der Schutz eines Allgemeininteresses nicht nur im Rahmen der behördlichen Aufgabenerfüllung gewährleistet sondern auch durch eine nicht öffentliche Stelle, die ein legitimes Interesse an der Datenverarbeitung (Art. 6 Abs. 1 lit. f) DSGVO) hat, erfüllt ist, wird im Rahmen einer ausführlichen rechtlichen Begutachtung zu ermitteln sein (siehe D 5.2 Datenschutz-Folgenabschätzung). Fest steht jedoch, dass nur die Daten erhoben werden dürfen, die prinzipiell erforderlich und geeignet sind, eine spätere Benachrichtigung Betroffener zu ermöglichen.

Im Rahmen der Erhebung stehen Transparenz und Datenminimierung besonders im Fokus. Bei der Suche nach bzw. der Auswahl von geeigneten Datensenken sollte vor allem die Aktualität eines Leaks als Faktor zur Datenminimierung ausschlaggebend sein. Hintergrund ist die Annahme, dass je früher die Betroffenen von einem Datenschutzvorfall erfahren desto schneller können sie Gegenmaßnahmen ergreifen und mögliche (Folge)Schäden abwehren oder wenigstens abschwächen. Da es sich beim EIDI-Ansatz jedoch um ein neues Konzept handelt, ist es zu Forschungszwecken zunächst notwendig auch ältere Leaks heranzuziehen. Im Hinblick auf eine dem Zweck angepasste, datensparsame Verarbeitung, ist im späteren Realbetrieb nach dem Verhältnismäßigkeitsprinzip darauf zu achten, dass keine unverhältnismäßige Datenerhebung stattfindet. Dies schließt insbesondere die Löschung nicht benötigter bzw. nicht erforderlicher Daten (z.B. doppelte Datensätze) ein.

Das Transparenzgebot richtet sich, auch vor dem Hintergrund der Beweissicherung, vor allem danach, dass für eine datenschutzkonforme Information der Betroffenen Angaben zur Bezugsquelle der verarbeiteten Daten erforderlich sind. Da die Erhebung nachweislich nicht beim Betroffenen erfolgt ist, ist dieser gemäß Art. 14 Abs. 2 DSGVO umfassend zu informieren. Die Dokumentation der Fundstellen stellt zudem die Grundlage für weitere Ermittlungen bzw. die Durchsetzung gerichtlicher Ansprüche dar.

[37] Weichert: Datenschutz bei Internetveröffentlichungen, in: VuR (2009), S. 328.
[38] Robrahn/Bremert: Interessenskonflikte im Datenschutzrecht, in: ZD 7 (2018), S. 292.

An diese Anforderung schließt sich die Gewährleistung der Verfügbarkeit der Daten an. Da es keine Garantie für den Fortbestand einer konkreten Datensenke gibt, muss sichergestellt werden, dass für die spätere Benachrichtigung erforderliche Daten inklusive aller relevanten Metainformationen (Herkunft, Alter, enthaltene Kategorien) dem Betroffenen auch tatsächlich zur Verfügung gestellt werden können. Dies lässt sich nur durch das separate Abspeichern der in Datensenken gefundenen Leaks umsetzen. Die Verfügbarkeitsanforderungen sind auch innerhalb der verantwortlichen (EIDI)Stelle zu erfüllen. Die interne Verfügbarkeit beinhaltet insbesondere Schutzmaßnahmen gegen Zerstörung oder unbeabsichtigtes Löschen und ähnliches.

Auf Grund der großen Datenmenge und der Sensitivität der Daten erfordert eine Speicherung zugleich die Umsetzung hinreichender Maßnahmen zur Sicherstellung der Vertraulichkeit der Daten gegenüber unbeabsichtigtem und unbefugtem Zugriff sowie integritätssichernde Maßnahmen zum Schutz vor Verfälschungen. Die verantwortliche Stelle trägt die Verantwortung dafür jederzeit die Kontrolle über die Erhebung, Speicherung, Löschung Übermittlung und Weiterverarbeitung ausüben zu können. Dies schließt auch die Möglichkeit ein, die Verarbeitung jederzeit unterbrechen zu können, z.B. wenn es Hinweise auf einen IT-Sicherheits- oder Datenschutzvorfall gibt.

Zur Sicherung der Integrität der „Originale“ sowie zur Vermeidung unnötiger Verkettung sind auch Zusammenfassungen verschiedener Leaks in einer gemeinsamen Datenbank kritisch zu betrachten. Es ist jedoch denkbar, dass sich im weiteren Verlauf statt dem Versand verschiedener Einzelnachrichten eine Zusammenfassung für den Betroffenen anbietet um einer Warnungsmüdigkeit auf Grund ständiger Meldungen vorzubeugen.

3.1.2 Bewertung der Datenqualität

Die Bewertung der Datenqualität folgt in Bezug auf die datenschutzrechtlichen Aspekte ebenfalls der Schutzzielsystematik. Integrität, Vertraulichkeit und Nichtverkettbarkeit stehen bei der Beurteilung der Qualität der Daten aus den Datensenken im Mittelpunkt.

Die Integrität der Daten ist abhängig von der zuverlässigen Erkennung verwendeter Trennzeichen. Diese variieren von Leak zu Leak und müssen daher jedes Mal neu bestimmt werden (D 2.1, S 35). Trennzeichen, wie beispielsweise Doppelpunkte, Kommas oder Leerzeichen, grenzen die einzelnen Daten oder Datensätze innerhalb eines Leaks voneinander ab. Die Trennung der Datensätze bildet die Basis für eine spätere Identifizierung der verschiedenen Betroffenen.

Die Trennung verschiedener Datensätze voneinander führt gleichzeitig zu einer Verknüpfung einzelner Daten zu einem personenbeziehbaren Datensatz. Hier stellt sich die Anforderung an eine integre Verkettung damit die enthaltenen Identitätsdaten am Ende der richtigen Person zugeordnet werden können. Adressdaten könnten beispielsweise mehreren Personen

zugeordnet sein, dies soll jedoch nicht dazu führen, dass mit dieser Adresse verknüpfte E-Mail-Adressen als eine einzelne Identität gewertet werden. Eine solche Bewertung würde in der Folge zu einer Fehleinschätzung bezüglich des Risikos, das sich aus dem Leak ergibt führen.

Die Datenanalyse sollte sich – neben dem Standardanwendungsfall E-Mail-Adresse/Passwort und mit Blick auf eine mögliche Datenminimierung – auf Kategorien beschränken, die eine Kontaktaufnahme mit dem Betroffenen ermöglichen. Hilfreich sind hierfür beispielsweise der Name, die Adresse, Anbieterspezifische Kennnummern (Kundennummern) oder Telefonnummern. Inwieweit andere Merkmale digitaler Identitäten zur Benachrichtigung Betroffener verwendet werden können, vielleicht sogar ohne diese eindeutig identifizieren zu müssen, wird im Projektverlauf zu klären sein. Es ist jedoch davon auszugehen, dass dies eher die Ausnahme sein wird. Die Vertraulichkeit wird daher anderweitig gesichert werden müssen, z.B. durch die Vergabe von Pseudonymen. Die Qualitätsbewertung von Leaks kann anhand eines (noch zu entwickelnden) objektiven Kriterienkatalogs erfolgen, der nicht auf der Auswertung einzelner personenbezogener Daten, sondern auf anonymisierten bzw. statistischen Werten beruht.

In diesem Schritt wird eine Datensenke einem Vorauswahlprozess unterzogen. Entscheidend ist, ob a) personenbezogene Daten enthalten sind, die b) aktuell ein Risiko für die Rechte und Freiheiten der Betroffenen darstellen. Um hierüber eine Aussage treffen zu können, kann eine automatische Auswertung der Datenstrukturen erfolgen. Eine inhaltliche Auswertung der Daten ist noch nicht erforderlich. Die Muster bzw. Datenstruktur [„ZEICHEN“@“ZEICHEN“.tld] lässt bereits auf das Vorhandensein von E-Mail-Adressen schließen. Eine Kenntnisnahme ist mit Hilfe dieser Methode nicht erforderlich. Sind mindestens drei E-Mail-Adressen enthalten, wird der Leak weiter verwendet (D 2.1, S.36). Die automatische Ermittlung sich wiederholender Strukturen, wie Trennzeichen, Zeichenfolgen bestimmter Länge oder Kombinationen bestimmter Zeichen gewährleistet eine vertrauliche und dennoch akkurate Einschätzung zur Eignung einer Datensenke bzw. eines Leaks. Eine manuelle Auswertung von Datensenken ist nur während der Entwicklungsphase notwendig.

Besondere Kategorien personenbezogener Daten dürfen nur ausnahmsweise in den von der DSGVO festgelegten Fällen gemäß Art. 9 Abs. 2 DSGVO verarbeitet werden. Besondere Kategorien personenbezogener Daten sind nach abschließender Aufzählung des Art. 9 Abs. 1 DSGVO „…Daten aus denen die rassische und ethnische Herkunft, politische Meinungen, religiöse oder weltanschauliche Überzeugungen oder die Gewerkschaftszugehörigkeit hervorgehen…“ sowie genetische und Gesundheitsdaten, Daten zum Sexualleben oder zur sexuellen Orientierung sowie eindeutig identifizierende biometrische Daten. Diese Daten sind entsprechend als besonders schützenswert einzustufen. Finden sich Daten besonderer Kategorien in einer Datensenke wieder, ist davon auszugehen, dass das Risiko für die Rechte und Freiheiten des Betroffenen besonders hoch ist. Dass derartige Angaben das

Ziel einer Benachrichtigung Betroffener fördern, kann bezweifelt werden. Daher ist – auch im Sinne der Datenminimierung – von der Verarbeitung besonderer Kategorien personenbezogener Daten im EIDI-Projekt abzusehen und es sollten entsprechende Maßnahmen getroffen werden, die eine zufällige Miterhebung bzw. Verarbeitung ausschließen.

3.1.3 Identifizierung der Opfer

Die Identifizierung potentieller Opfer eines Identitätsdiebstahl bzw. eines Datenmissbrauchs ist ein geeigneter Weg Betroffene nach einem Datenschutzvorfall auch dann zu benachrichtigen, wenn es keine verantwortliche Stelle gibt oder diese nicht in der Lage ist ihren Verpflichtungen nachzukommen. Dieser Arbeitsschritt bezieht Dritte in Form von Identitätsprovidern in die Datenverarbeitung mit ein und nutzt den Umstand, dass Individuen heute häufig über eine Vielzahl digitaler Identitäten in unterschiedlichen Ausprägungen verfügen. Digitale Identitäten enthalten unterschiedliche Merkmale (z.T. personenbezogene Daten), wie E-Mail-Adressen und Passwörter. Dabei gibt es durchaus Überschneidungen der Merkmale zwischen verschiedenen Identitäten. Diese können genutzt werden um Betroffene als Träger einer Identität verschiedenen Anbietern zuzuordnen. Ein Datensatz in einem Leak der E-Mail-Adressen und Passwörter enthält, kann von einem anderen Identitätsprovider über die E-Mail-Adresse einem Nutzer zugeordnet werden. Der Identitätsprovider ist dann auf Grund der Verfügbarkeit dieser Daten in der Lage den Betroffenen über einen eigenen Kanal zu informieren. Dies liegt insbesondere dann sowohl im Interesse des Betroffenen wie auch des Identitätsproviders, wenn die öffentlich verfügbaren (Login)Daten für mehrere Nutzerkonten verwendet werden können.

Eine Verkettung mehrerer Identitäten und der darin enthaltenen personenbezogenen Daten zu einem Profil darf und soll jedoch nicht das Ziel oder die Konsequenz dieser Identifizierungsmöglichkeit sein. Die beteiligten Identitätsprovider sollen daher nach der Identifizierung nicht über mehr Informationen zum Träger der Identität verfügen als ihnen vorher bekannt waren, um dem Prinzip der Nichtverkettung gerecht zu werden. Das bedeutet auch, dass einem Anbieter nicht bekannt sein darf, welches Identitätsmerkmal bei welchen anderen Anbietern verwendet wurde. Eine direkte Rückmeldung zur Vermeidung von Mehrfachbenachrichtigung wird entsprechend nicht möglich sein, da hierdurch ersichtlich wird, welcher Betroffene bei welchen Anbietern digitale Identitäten angelegt hat.

Sind unter Umständen doch zusätzliche Informationen unverzichtbar, sollte im Sinne der Datenminimierung sichergestellt werden, dass die Daten nur für den Zweck der Benachrichtigung verwendet werden dürfen und anschließend zu löschen sind.

Die zentrale Rolle der Identitätsprovider als Datenempfänger zieht die Frage nach sich, wie einem zusätzlichen Datenmissbrauch auf deren Seite vorgebeugt werden kann. Vorab ist daher zu prüfen ob eine Übermittlung überhaupt zulässig ist. Dies gilt insbesondere, wenn dies die Datenübermittlung in ein Drittland mit einschließt. Dies ist nur möglich sofern die

Anforderungen der Art. 44 ff DSGVO in vollem Umfang erfüllt werden. Generell sollte – im Rahmen der Intervenierbarkeit - gewährleistet sein, dass Akteure, die sich nicht regelkonform verhalten von der Nutzung ausgeschlossen werden können. Grundsätzlich ist der Zugriff auf die Daten so zu gestalten, dass die Datenverarbeitung jederzeit gestoppt werden kann. Denkbar wäre auch, dass vor der Teilnahme an EIDI eine Zertifizierung des Identitätsanbieters gemäß Art. 42 DSGVO erforderlich ist.

Der mit der Identifizierung verbundene Datenaustausch soll zudem nicht dazu führen, dass eine weitere unbefugte Verarbeitung personenbezogener Daten erfolgen kann und muss durch Maßnahmen zur Sicherung der Vertraulichkeit, Integrität und Verfügbarkeit geschützt werden.

Sollte es dennoch zu einem IT-Sicherheits- und/oder Datenschutzvorfall kommen, muss nachvollziehbar sein, welcher Akteur wann auf welche Daten zugreifen konnte. Es müssen daher Bereits im Vorfeld Dokumentationspflichten festgelegt werden mit deren Hilfe Transparenz für das EIDI-Übermittlungsverfahren gewährleistet werden kann.

Im Projektverlauf ist zudem zu klären, wie mit Leak-Daten umzugehen ist, die keinem Betroffenen zugordnet werden können bzw. für die sich keine Benachrichtigungstelle findet. Im Sinne des Datenminimierungsprinzips sind die Grenzen der Datenverarbeitung durch EIDI festzusetzen. Ziel ist es die Verhältnismäßigkeit (Erforderlichkeit und Zweckbindung) der Datenverarbeitung zu gewährleisten. Eine unberechtigt veröffentlichte E-Mail-Adresse ist zwar ein Datenschutzvorfall, trägt jedoch nur ein geringes Risiko für die Rechte und Freiheiten des Betroffenen in sich. Je nach Datum und Kombinationen sind Bagatellgrenzen festzulegen, damit nicht unnötig Daten an Dritte übermittelt werden, die möglicherweise ein höheres Risiko für die Betroffen darstellen als der ursprüngliche Datenschutzvorfall. Auf diese Weise wird verhindert, dass EIDI selbst zum Datenhehler wird, der ein hohes Risiko für die Betroffenen darstellt.

3.1.4 Information der Betroffenen

Die Information der Betroffenen beinhaltet nicht nur die Benachrichtigung über den Datenschutzvorfall (Verfügbarkeit) sondern stellt auch die Wahrung der Informationspflichten gemäß Art 14 DSGVO sicher (Transparenz). Der Betroffene hat demnach auch gegenüber EIDI einen Anspruch auf die Wahrung der Betroffenenrechte (Intervenierbarkeit) nach Art. 15 ff DSGVO.

Aus Sicht der Betroffenen muss die Benachrichtigung vor allem integer, vertraulich und transparent sein. Die Integrität der Benachrichtigung – Echtheit des Absenders, Hintergrund der Benachrichtigung - sollte für den Betroffenen erkennbar und nachvollziehbar sein (Siegel, Kontaktmöglichkeiten, Link etc.). Außerdem sollte der Inhalt der Nachricht noch aktuell sein. Falsch-positive Benachrichtigungen müssen unterbunden werden, da hierdurch nicht

nur der Empfänger unnötig verunsichert, sondern auch die Vertraulichkeit gegenüber dem tatsächlich Betroffenen verletzt wird. Die für den Betroffenen wahrscheinlich entscheidenden Informationen - neben der Mitteilung einer Datenschutzverletzung – sind sehr wahrscheinlich die zum weiteren Vorgehen.

Die Art und Weise der Benachrichtigung, z.B. der verwendete Kommunikationskanal, die Gestaltung und der Inhalt, sollte sowohl dem Risiko der Datenschutzverletzung wie auch dem Empfänger(kreis) angepasst sein. Die wirksamste Form und Aufmachung der Nachricht wird im Projektverlauf evaluiert werden müssen. Im Spannungsfeld von Verfügbarkeit und Vertraulichkeit muss unteranderem die Frage geklärt werden, wann neben dem Betroffenen auch Dritte, wie beispielsweise Admins oder der Arbeitsgeber informiert werden müssen und wem diese Verpflichtung obliegt.

In Zuge dieses Erstkontakts muss dem Betroffenen zudem die Möglichkeit zum Widerspruch gegen die Datenverarbeitung durch EIDI eingeräumt werden. Legt der Betroffene dar, dass die besonderen Umstände des Einzelfalls einer Datenverarbeitung entgegenstehen und es überwiegen keine zwingend schutzwürdigen Interessen auf Seiten des Verantwortlichen, muss die verantwortlichen Stelle entsprechende Maßnahmen einleiten um die Einhaltung von Art. 21 DSGVO zu gewährleisten.

3.1.5 Weitere Verarbeitung

Meldepflichtig ist nach Art. 33 DSGVO der für die Datenverarbeitung Verantwortliche. Meldepflichtig ist dabei die Verletzung des Schutzes personenbezogener Daten. Entsprechend der Legaldefinition aus Art. 4 Nr. 12 DSGVO ist hiermit „eine Verletzung der Sicherheit, die, ob unbeabsichtigt oder unrechtmäßig, zur Vernichtung, zum Verlust, zur Veränderung, oder zur unbefugten Offenlegung von beziehungsweise zum unbefugten Zugang zu personenbezogenen Daten führt, die übermittelt, gespeichert oder auf sonstige Weise verarbeitet wurden“ gemeint. Die Meldung eines Datenschutzvorfalls im Sinne von Art. 33 DSGVO erfolgt an die zuständige Aufsichtsbehörde.

Im Sinne einer den Datenschutz fördernden Meldepflicht kann bei einer unbekannten verantwortlichen Stelle die Meldung an eine Aufsichtsbehörde möglicherweise auch durch EIDI erfolgen. Mit Hilfe zusätzlicher Analysen der Datensenken könnte es auch möglich sein den Verantwortlichen zu bestimmen und über einen bisher unbekannten Datenschutzvorfall zu informieren, damit dieser seinen Verpflichtungen gemäß DSGVO nachkommen kann. Ein Datenaustausch mit anderen Identitätsprovidern ist dann nicht mehr notwendig. Die Benachrichtigung der Betroffenen würde durch die verantwortliche Stelle erfolgen. Die Datenverarbeitung zu diesem Leak durch EIDI kann dann beendet und die Daten können gelöscht werden (Datenminimierung).

Wurde die Benachrichtigung gemäß des EIDI-Frameworks eingeleitet, sind nicht mehr erforderliche Daten ebenfalls zu löschen. Sollen bestimmte Informationen, bspw. zur Verbesserung der Verarbeitung genutzt werden, ist die Einwilligung der Betroffenen einzuholen oder die Daten sind so zu anonymisieren, dass keine Rückschlüsse auf den Betroffenen mehr möglich sind.

Eine weitere Möglichkeit der Nachbetreuung ist die Zusammenarbeit mit Strafverfolgungsbehörden. Hier ist zu prüfen unter welchen Umständen diese einbezogen werden müssen oder sollten. Da die Täter häufig über Ländergrenzen hinweg arbeiten, ist auch eine Zusammenarbeit auf internationaler bzw. europäischer Ebene, z. B. mit Europol, für konkrete Einzelfälle auch aus Sicht der Betroffenen wünschenswert.

Die Betroffenen können zudem zivilrechtliche Schritte gegen die Identitätsdatendiebe, die Identitätsdatenhehler, die Plattformbetreiber oder die ursprüngliche verantwortliche Stelle einleiten und benötigen hierfür die entsprechenden belastbaren Ergebnisse der Datenanalyse durch EIDI.

Kommen Verantwortliche ihren Pflichten nach DSGVO nicht nach, haben auch die Datenschutzaufsichtsbehörden ein Interesse an diesen Informationen um einen bußgeldbewährten Verstoß nachweisen zu können.

3.2 Schlussfolgerungen

Der Umgang mit Datensenken richtet sich nach dem Schutzbedarf der darin enthaltenen Daten und dem daraus resultierenden Risiko für den Betroffenen. Das Datenschutzrecht differenziert zwischen Daten mit normalem oder mit hohem bzw. sehr hohem Schutzbedarf.

Für jede Datensenke ist im Zuge der Erhebung und Qualitätsbewertung der Datensenke das konkrete Risiko zu bestimmen. Ausgehend von den verschiedenen Risikostufen und dem daraus resultierenden Schutzbedarf sind entsprechende Maßnahmen für den Umgang mit den personenbezogenen Daten zu treffen.

Die Bestimmung des Risikos richtet sich dabei mindestens nach den aktuell vom Gesetzgeber und den Aufsichtsbehörden angewendeten objektiven Kriterien[39]. Im Sinnen des Datenschutz-by-Design- Prinzips sollte jedoch auch der aktuelle Forschungsstand mit berücksichtigt werden[40].

[39] DSK: Kurzpapier Nr. 18 – Risiko (Stand 26.04.2018) , über: https://www.datenschutzzentrum.de/uploads/dsgvo/kurzpapiere/DSK_KPNr_18_Risiko.pdf

[40] Vgl. Bieker: Die Risikoanalyse nach dem neuen EU-Datenschutzrecht und dem Standard-Datenschutzmodell, in: DuD, Nr. 1 (2018), S. 27-31.

4 Literaturverzeichnis

AK Technik der Konferenz der unabhängigen Datenschutzbehörden des Bundes und der Länder (Hrsg.): Das Standard-Datenschutzmodell (V 1.1), Düsseldorf 2018.

Bieker: Die Risikoanalyse nach dem neuen EU-Datenschutzrecht und dem Standard-Datenschutzmodell, in: DuD, Nr. 1 (2018), S. 27-31.

Koreng: Wer hat Angst vor Social Bots? In: NJW 21/2017.

Kühling/Buchner: Datenschutz-Grundverordnung - Kommentar, 1. Aufl. 2017.

Malderle/Wübbeling: Dienstekatalog gängiger Datensenken und dort öffentlich verfügbarer Identitätsdaten-Sammlungen (D 2.1), August 2017.

Robrahn/Bremert: Interessenskonflikte im Datenschutzrecht, in: ZD 7 (2018), S. 291-297.

Rost/ Bock: Impact Assessment im Lichte des Standard-Datenschutzmodells, in: DuD Nr. 10 (2012), S. 743-747.

Weichert, Thilo: Datenschutz bei Internetveröffentlichungen, in: VuR (2009), S. 323-330.