Bundesministerium des Innern

Digital Gipfel
Ludwigshafen 2017
Rhein-Neckar

Schwartmann / Weiß (Hrsg.)

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

– Leitlinien für die rechtssichere Nutzung von Pseudonymisierungslösungen unter Berücksichtigung der Datenschutz-Grundverordnung –

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

– Leitlinien für die rechtssichere Nutzung von Pseudonymisierungslösungen unter Berücksichtigung der Datenschutz-Grundverordnung –

Digital Gipfel
Ludwigshafen 2017
Rhein-Neckar

Version 1.0, 2017

Herausgeber:
Fokusgruppe Datenschutz des Digital-Gipfels

Kontakt Leitung:
Prof. Dr. Rolf Schwartmann
(TH Köln/GDD)

Kölner Forschungsstelle für Medienrecht

Technology
Arts Sciences
TH Köln

Kontakt Sherpa:

Steffen Weiß

Gesellschaft für Datenschutz
und Datensicherheit e.V.
Heinrich-Böll-Ring 10
53119 Bonn
Tel.: +49 228 96 96 75 00
E-Mail: info@gdd.de · www.gdd.de

GDD
Gesellschaft für Datenschutz
und Datensicherheit e.V.

Leitung: **Prof. Dr. Rolf Schwartmann**
Kölner Forschungsstelle für Medienrecht – TH Köln

Sherpa: **Steffen Weiß, LL.M.**
Gesellschaft für Datenschutz und Datensicherheit e.V.

Mitglieder: **Prof. Dr. Christoph Bauer**
ePrivacy GmbH

Patrick von Braunmühl
Bundesdruckerei GmbH

Susanne Dehmel
Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V.

Mitglieder: **Walter Ernestus**
Die Bundesbeauftragte für den Datenschutz und die Informationsfreiheit

Nicolas Goß
eco – Verband der Internetwirtschaft e.V.

Michael Herfert
Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.

Serena Holm
SCHUFA Holding AG

Dr. Detlef Houdeau
Infineon Technologies AG

Mitglieder: **Annette Karstedt-Meierrieks**
Deutscher Industrie- und Handelskammertag e.V.

Johannes Landvogt
Die Bundesbeauftragte für den Datenschutz und die Informationsfreiheit

Prof. Dr. Michael Meier
Universität Bonn/Gesellschaft für Informatik e.V.

Jonas Postneek
Bundesamt für Sicherheit in der Informationstechnik

Frederick Richter, LL.M.
Stiftung Datenschutz

Mitglieder: **Dr. Sachiko Scheuing**
Acxiom Deutschland GmbH

Irene Schlünder
Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V.

Dr. Claus D. Ulmer
Deutsche Telekom AG

Dr. Winfried Veil
Bundesministerium des Innern

Dr. Martina Vomhof
Gesamtverband der Deutschen Versicherungswirtschaft e.V.

Vorwort

Daten sind der Rohstoff der Zukunft. Er kann aus Informationen über unsere Lebensführung bestehen, die für Medizinprodukte genutzt werden können, oder aus Informationen über unser Fahrverhalten, aus dem Wettervorhersagen genauso generiert werden können, wie Fahrfehler. Er entsteht durch das Nutzen vernetzter Geräte, vom Fernseher bis zum Toaster.

Weil der Rohstoff aus menschlicher Persönlichkeit besteht, darf man ihn nicht abbauen wie Erz, das in der Esse zu Stahl geschmolzen wird. Dieser Rohstoff aus digitalisierter Persönlichkeit benötigt Schutz, bevor er genutzt werden darf.

Das europäische Datenschutzrecht der Datenschutz-Grundverordnung gibt der Wirtschaft die Verantwortung für den Schutz der personenbezogenen Daten, indem sie von den verantwortlichen Unternehmen verlangt, Datenschutz durch Technikgestaltung und durch datenschutzfreundliche Voreinstellungen zu gewährleisten. Ein wesentliches Mittel dazu stellt nach der Konzeption des neuen Rechts die Pseudonymisierung von personenbezogenen Daten dar. Durch entsprechende Verfahren sollen Daten von ihrem Personenbezug befreit werden, um so datenschutzkonform wirtschaftlich genutzt werden zu können.

Wegen der besonderen Bedeutung der Pseudonymisierung hat sich die Fokusgruppe Datenschutz des Digital-Gipfels der Bundesregierung dieses wirtschaftlich bedeutsamen Themas angenommen und unter Mitwirkung von interdisziplinär ausgewählten Vertretern aus Wirtschaft, Ministerialverwaltung, Wissenschaft und Aufsicht dieses Whitepaper erstellt.

Allen Mitwirkenden gilt herzlicher Dank für die intensive, konstruktive und effiziente Zusammenarbeit an diesem Projekt. Besonderer Dank gebührt Herrn Assessor Steffen Weiß für die fachkundige und umsichtige Koordination der Arbeit der Fokusgruppe.

Köln, im Juni 2017

Professor Dr. Rolf Schwartmann
(Leiter der Fokusgruppe Datenschutz der Plattform Sicherheit, Schutz und Vertrauen für Gesellschaft und Wirtschaft im Rahmen der Digitalen Agenda der Bundesregierung)

Inhalt

1. Einleitung

Die Digitalisierung und Vernetzung der Welt schreitet weiter voran und ist mit einer signifikanten Steigerung von Datenmengen und dem Bedarf an Informationen verbunden. Dies betrifft auch personenbezogene Daten von Bürgerinnen und Bürgern. Der Schutz dieser Daten sollte ein zentrales Anliegen sein, um zu verhindern, dass Betroffene einer Datenverarbeitung ausgesetzt werden, die sie nicht mehr beherrschen oder überblicken können. Gleichzeitig bietet der technische Fortschritt auch neue Möglichkeiten für den Betroffenen selbst, was mit einer Preisgabe seiner persönlichen Daten einhergehen kann. Es bedarf daher eines angemessenen Ausgleichs zwischen dem Interesse eines Betroffenen an einem möglichen Ausschluss der Verarbeitung seiner Daten und einer erforderlichen Verarbeitung von persönlichen Daten durch einen Datenverarbeiter. Ein solcher Ausgleich sollte sich grundsätzlich daran orientieren, ob eine personenbezogene Datenverwendung notwendig ist oder nicht.

Die Pseudonymisierung personenbezogener Daten bietet eine Möglichkeit, zwischen den entgegenstehenden Interessen von Betroffenen und Datenverarbeitern zu vermitteln und Szenarien zu gestalten, bei denen eine Verwendung von Klardaten nicht mehr erforderlich ist. Sie zeichnet sich dadurch aus, dass identifizierende Merkmale einer Person im Zuge einer Verarbeitung entfernt werden. Die Pseudonymisierung erfüllt eine Vielzahl von für den Datenschutz nützlichen Funktionen. Zum Beispiel sorgt sie dafür, dass selbst im Falle eines ungewollten Datenverlustes einer verantwortlichen Stelle Persönlichkeitsrechte von Betroffenen gewahrt werden können, da eine Zuordnung von Daten wesentlich erschwert ist. Pseudonyme Daten schaffen aber auch Vertrauen, indem in möglichst transparenter Weise mithilfe technischer Aufwände eine „Codierung“ von persönlichen Informationen zugunsten vom Betroffenen erfolgt und er hierdurch das Interesse einer verantwortlichen Stelle am Schutz dieser Informationen erkennen kann.

Dieses Whitepaper möchte einen Überblick darüber geben, wie eine Pseudonymisierung aus Sicht des Datenschutzes einzuordnen ist und welche Funktionen sie einnehmen kann. Darüber hinaus werden technisch-organisatorische Möglichkeiten aufgezeigt, um eine Pseudonymisierung faktisch umzusetzen. Abschließend werden konkrete Anwendungsszenarien beleuchtet, wie ein Umgang mit pseudonymisierten Daten in der Praxis bereits heute erfolgt.

Die Inhalte dieses Papiers werden an den neuen rechtlichen Vorgaben der EU-Datenschutz-Grundverordnung (DS-GVO) ausgerichtet, die ab dem 25.05.2018 europaweit Anwendung finden wird und sich an verschiedenen Stellen ausdrücklich zur Pseudonymisierung positioniert. Hierbei kann der Digital-Gipfel einen wesentlichen Beitrag dadurch leisten, dass jahrelange Erfahrungen zur Pseudonymisierung auf Basis des noch gültigen Bundesdatenschutzgesetzes (BDSG) in einen europäischen Diskurs eingebracht werden, um diesbezüglich zunächst ein gemeinsames Verständnis zu entwickeln.

2. Auftrag für den Digital-Gipfel / Ziele der Fokusgruppe

Die Erarbeitung des Whitepapers erfolgt im Kontext des Digital-Gipfels, der auf Initiative der Bundesregierung mit insgesamt 9 Plattformen auf den digitalen Wandel reagiert. Die Plattform 8 „Sicherheit, Schutz und Vertrauen für Gesellschaft und Wirtschaft“ des Gipfels unter dem Vorsitz von BM Dr. Thomas de Maizière hat es sich zur Aufgabe gemacht, Sicherheit und Schutz im Netz so herzustellen, dass die Digitalisierung ihr volles Potenzial für Gesellschaft und Wirtschaft in Deutschland entfalten kann. Ein moderner Datenschutz auf hohem Niveau kann dabei die Freiheit und Persönlichkeitsrechte der Bürgerinnen und Bürger gewährleisten und gleichzeitig die Chancen der Digitalisierung nutzbar machen. Hierzu wurde im Zuge der Ausrichtung der Plattform 8 die Fokusgruppe Datenschutz unter Leitung von Prof. Dr. Schwartmann gegründet.

Neben der erwähnten Durchleuchtung der Pseudonymisierung und ihrer konkreten Anwendung möchte die Fokusgruppe Datenschutz erreichen, dass sich aus diesem Papier Leitlinien für einen rechtssicheren Umgang mit Pseudonymisierungslösungen gewinnen lassen, die sowohl in privatrechtlichen wie auch öffentlichen Organisationen und Einrichtungen eingesetzt werden können. Mittels Leitlinien kann ein wesentlicher Beitrag für einen flächendeckenden und einheitlichen Einsatz von Pseudonymisierungen geleistet werden.

Die Ziele dieser Fokusgruppe gehen jedoch über das Schaffen von Leitlinien hinaus. Es wäre im Sinne eines europaweit harmonisierten Vorgehens erstrebenswert, die hier entwickelnden Leitlinien in einen Code of Conduct zur Pseudonymisierung zu überführen, um hier einen anerkannten und verbindlichen Standard zu schaffen. Die DS-GVO ermutigt ausdrücklich Verbände und Vereinigungen zur Schaffung von Verhaltensregeln um Datenverarbeitungen in der DS-GVO – so auch eine Pseudonymisierung – zu konkretisieren. Diese Verhaltensregeln sollen dann von der zuständigen Datenschutzaufsichtsbehörde genehmigt werden. Bei besonderer Bedeutung für den europäischen Rechtsraum ist ferner eine Allgemeinverbindlichkeitserklärung durch die EU-Kommission möglich.

3. Rahmenbedingungen der Pseudonymisierung

3.1. Definition

Im deutschen Datenschutzrecht spielt die Pseudonymisierung schon lange eine Rolle und ist per Definition über § 3 Abs. 6a im BDSG verankert. Sie wird sowohl als technisch-organisatorische Maßnahme zur sicheren Verarbeitung als auch als datenminimierende Maßnahme eingesetzt. Im Telemediengesetz (TMG) befindet sich mit § 15 Abs. 3 sogar ein Erlaubnistatbestand, der die Pseudonymisierung als Tatbestandsvoraussetzung enthält. Mit der DS-GVO wird nun eine europaweit einheitliche Definition der Pseudonymisierung eingeführt und an mehreren Stellen auf dieses Instrument Bezug genommen.

„Pseudonymisierung“ wird in Art. 4 Nr. 5 DS-GVO definiert als „die Verarbeitung personenbezogener Daten in einer Weise, dass die personenbezogenen Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können, sofern diese zusätzlichen Informationen gesondert aufbewahrt werden und technischen und organisatorischen Maßnahmen unterliegen, die gewährleisten, dass die personenbezogenen Daten nicht einer identifizierten oder identifizierbaren natürlichen Person zugewiesen werden.“

Daraus ergeben sich drei Anforderungen, die eine Pseudonymisierung erfüllen muss:

a) Ohne Hinzuziehung zusätzlicher Informationen keine Zuordnung der Daten zu einer spezifischen Person

Können die vorhandenen Daten **ohne Weiteres** einer identifizierbaren Person zugeordnet werden (z.B. über einen Namen, eine Anschrift oder eine Personalnummer), liegt ohnehin keine Pseudonymisierung vor. In diesem Fall spielen die weiteren Anforderungen an eine Pseudonymisierung keine Rolle mehr.

b) Getrennte Aufbewahrung der zusätzlichen Informationen

Die Daten, mit denen die Zuordnung zu einer Person möglich wäre, müssen derart getrennt aufbewahrt werden, dass sie nicht ohne Weiteres zusammengeführt werden können. Dies kann z.B. durch eine logische Trennung mit unterschiedlichen Zugriffsberechtigungen erfolgen. Eine technische oder organisatorische Trennung, die nicht geeignet ist, den Zugriff auf die eine Zuordnung ermöglichenden Daten zu verhindern, genügt nicht. Der Aufwand der Trennung kann sich an der Schutzbedürftigkeit der Daten orientieren.

Beim Einsatz von Pseudonymisierungsverfahren ist stets im Vorhinein zu klären, wer über Zuordnungstabellen bzw. Verschlüsselungsverfahren verfügen soll, wer das Pseudonym generiert, ob ein De-Pseudonymisierungsrisiko ausgeschlossen werden kann und unter welchen Voraussetzungen eine Zusammenführung mit den Identifikationsdaten gestattet ist.[1] Sollen zu den pseudonymisierten Daten weitere Daten hinzugespeichert werden, so ist zu prüfen, ob durch die hinzugefügten Daten die Pseudonymisierung aufgehoben werden könnte, weil der Datensatz über die weiteren Daten wieder eindeutig einer Person zuzuordnen ist.

c) Gewährleistung technischer und organisatorischer Maßnahmen zur Nichtzuordnung

Erwägungsgrund 26 der DS-GVO stellt klar, dass einer Pseudonymisierung unterzogene personenbezogene Daten, die durch Heranziehung zusätzlicher Informationen einer natürlichen Person zugeordnet werden könnten, als Informationen über eine identifizierbare natürliche Person betrachtet werden sollen. Als **personenbeziehbare Daten** fallen sie weiterhin in den Anwendungsbereich der DS-GVO.

Die Verordnung versteht die Pseudonymisierung in erster Linie als **risikomindernde technisch-organisatorische Maßnahme** (vgl. ErwG 28). Aus dem Erwägungsgrund geht auch hervor, dass durch die ausdrückliche Einführung der Pseudonymisierung in der Verordnung nicht beabsichtigt sei, andere Datenschutzmaßnahmen auszuschließen.

Um einen Anreiz für den Einsatz von Pseudonymisierung zu schaffen, wird in ErwG 29 klargestellt, dass Pseudonymisierungsmaßnahmen, auch wenn sie eine allgemeine Analyse zulassen, bei **demselben Verantwortlichen** möglich sein sollen, wenn dieser die erforderlichen technischen und organisatorischen Maßnahmen getroffen hat, um die unberechtigte Wiederherstellung des Personenbezugs zu verhindern. D.h. die Einbindung eines Dritten, so z.B. einen Datentreuhänder, ist nicht zwingend erforderlich. Im Einzelfall sollte geprüft werden, welcher Variante aus Sicht des Datenschutzes der Vorzug gegeben werden soll.

Zu berücksichtigen ist die Pseudonymisierung außerdem als ein Kriterium bei der Prüfung der Frage, ob eine Datenverarbeitung zu einem anderen Zweck als dem, zu dem die Daten ursprünglich erhoben wurden, mit dem ursprünglichen Zweck vereinbar ist (Art. 6 Abs. 4 Buchst.e).[2] Da pseudonymisierte Daten als personenbeziehbar zu behandeln sind, sind sie im Übrigen nach Wegfall des Zweckes ihrer Verarbeitung zu löschen.[3]

---

1 Paal/Pauly, Datenschutz-Grundverordnung, DS-GVO Art. 4 Rn. 40-47.
2 Siehe hierzu Ziff. 3.3.2.
3 Siehe hierzu Ziff. 4.4.

3.2. Abgrenzung Pseudonymisierung und Anonymisierung

Im Verlauf der Diskussionen um die DS-GVO kam immer wieder die Frage auf, ob die Unterscheidung zwischen Anonymisierung und Pseudonymisierung überall in Europa gleich getroffen wurde. Die bisher geltende EU-Datenschutzrichtlinie 95/46/EG (DSRL) erwähnte den Begriff der Pseudonymisierung nicht und enthielt nur in ErwG 26 die Klarstellung, dass die Schutzprinzipien der DSRL „keine Anwendung auf Daten finden, die derart anonymisiert sind, dass die betroffene Person nicht mehr identifizierbar ist". Die Grundverordnung schafft hier klare Verhältnisse, in dem sie in Art. 4 Nr. 5 eine Definition der Pseudonymisierung enthält und auch in den Erwägungsgründen (insbesondere ErwG 26) die Abgrenzung von Pseudonymisierung und Anonymisierung vornimmt. Eine eigene Definition von Anonymisierung enthält der Art. 4 der Verordnung nicht – sie ergibt sich aber im Umkehrschluss aus der Definition der „personenbezogenen Daten" in Art. 4 Nr. 1 DS-GVO und wird in ErwG 26 so erklärt:

„Die Grundsätze des Datenschutzes sollten daher nicht für anonyme Informationen gelten, d. h. für 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 identifiziert werden kann****. Diese Verordnung betrifft somit nicht die Verarbeitung solcher anonymer Daten, auch für statistische oder für Forschungszwecke."*

Der Unterschied zwischen pseudonymisierten und anonymisierten Daten lässt sich also kurz so darstellen:

Pseudonymisierte Datenrung		Anonymisierte Daten
Betroffene Person kann unter Hinzuziehung der gesondert aufbewahrten oder ggf. öffentlich zugänglicher Informationen wieder identifiziert werden.	⟷	Betroffene Person kann nicht oder nur mit unverhältnismäßigem Aufwand wieder identifiziert werden.

Wann der Zustand erreicht ist, dass die betroffene Person nicht mehr identifiziert werden kann, war schon in der Vergangenheit immer wieder umstritten und ist relevant für die Abgrenzung zwischen Pseudonymisierung und Anonymisierung. Im BDSG war der Maßstab hierfür in § 3 Abs. 6 ausgeführt und besagte, dass die Anonymisierung dann erreicht war, wenn die Zuordnung von Angaben zu einer betroffenen Person nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitsaufwand zu erreichen war.

Auch die DS-GVO führt in ErwG 26 Näheres zur Frage aus, wann Informationen in einer Weise anonymisiert worden sind, dass die betroffene Person nicht mehr identifiziert werden kann:

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

Entscheidend sind also:

- Kosten der Identifizierung
- Erforderlicher Zeitaufwand
- Verfügbare Technologien zum Zeitpunkt der Verarbeitung
- Technologische Entwicklung (d.h. was ist an technischen Entwicklungen zukünftig absehbar)
- Sonstige objektive Faktoren

Die Grundverordnung folgt, ebenso wie das BDSG, einem relativen Ansatz hinsichtlich der Frage nach der Bestimmbarkeit einer Person. Gefordert ist also nicht die absolute Irreversibilität der Anonymisierung für alle Zeiten, sondern ein **Zustand**, in dem aller Wahrscheinlichkeit nach niemand die De-Anonymisierung vornehmen kann oder würde, weil sie viel zu aufwändig und schwierig bis unmöglich wäre. Dabei ist grundsätzlich die Beurteilung zum Zeitpunkt der Verarbeitung ausschlaggebend, aber es müssen auch in die Zukunft gerichtete technologische Entwicklungen mitberücksichtigt werden.[4] Bedeutsam ist diese Bewertung auch für die Frage, ob Daten, die bei einem Verantwortlichen pseudonymisiert wurden und für die nur er die notwendigen Informationen zur Wiederherstellung des Per-

4 Zu den rechtlichen Mitteln eines Webseitenanbieters zur Herstellung eines Personenbezugs von dynamischen IP-Adressen seiner Besucher, vgl. EuGH, Urteil vom 19-10.2016, Az. C-582/14.

sonenbezugs besitzt, bei Übermittlung an einen anderen Verantwortlichen für diesen nunmehr anonymisierte Daten sind.[5]

Eine Abgrenzung zwischen Pseudonymisierung und Anonymisierung wird daher mit Hilfe der oben genannten Kriterien aus der DS-GVO jeweils im Einzelfall vorzunehmen sein.

3.3. Funktionen

3.3.1. Schutzfunktion

Die Pseudonymisierung erfüllt in der DS-GVO unterschiedlichste Funktionen. Hierbei nimmt die **Schutzfunktion** zugunsten des Betroffenen einer Datenverarbeitung eine wesentliche Rolle ein, die an verschiedenen Stellen im Gesetz zum Ausdruck gebracht wird. Grundsätzlich schützt eine Pseudonymisierung vor einer direkten Identifikation einer Person. In diesem Sinne weist die DS-GVO in Art. 4 Abs. 5 darauf hin, dass Pseudonyme ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können. Ohne Zusatzwissen, das im Herrschaftsbereich des Verarbeitenden liegt, kann ein Betroffener möglicherweise nur mit einem unverhältnismäßigen Aufwand bestimmt werden. Das Pseudonym fungiert als seine „Maskierung". Der Schutz vor einer unmittelbaren Identifikation wird hierbei dadurch konkretisiert, dass diese zusätzlichen Informationen gesondert aufzubewahren sind und technischen und organisatorischen Schutzmaßnahmen unterliegen müssen.

Das Prinzip der **Datenminimierung** gem. Art. 5 Abs. 1 Buchst. c, d.h. personenbezogene Daten müssen dem Zweck angemessen und erheblich sowie auf das für die Zwecke der Verarbeitung notwendige Maß beschränkt sein, adressiert zwar nicht unmittelbar eine Pseudonymisierung von personenbezogenen Daten, eine Solche kommt jedoch dann zum Tragen und dem Betroffenen zugute, wenn ein Personenbezug von Informationen nicht mehr notwendig ist, um festgelegte Zwecke einer Verarbeitung zu realisieren. Die Pseudonymisierung ist damit eine konkrete Umsetzungsmaßnahme des gesetzlichen Gebots und Ausdruck eines sparsamen Umgangs mit personenbezogenen Daten. In Fortführung dieses Appells fordert das Datenschutzprinzip **„Privacy by Design"** gem. Art. 25[6], dass sowohl zum Zeitpunkt der Festlegung der Mittel für eine Verarbeitung personenbezogener Daten als auch zum Zeitpunkt der eigentlichen Verarbeitung durch technischorganisatorische Maßnahmen eine Datenminimierung wirksam umgesetzt wird, so unter anderem durch eine Pseudonymisierung. Im Sinne des Leitgedankens von „Privacy by Design" sorgt eine Pseudonymisierung dafür, dass bereits in einem frühen Stadium eine Entkoppelung persönlicher Informationen von anderen Daten erfolgen kann, was zu einem wirksamen und durchdringenden Schutz für die Betroffenen führen kann.

Die **Sicherheit** einer Verarbeitung personenbezogener Daten, die sich gem. ErwG 83 durch Maßnahmen zum Schutz vor unbeabsichtigter oder unrechtmäßiger Vernichtung von Daten, deren Verlust, Veränderung, unbefugter Offenlegung oder eines unbefugten Zugangs auszeichnet, kann, nach dem Willen der DS-GVO, ausdrücklich auch eine Pseudonymisierung beinhalten. D.h. eine Pseudonymisierung wäre dann Teil einer Datensicherheitsstrategie[7], die durch einen technisch-organisatorischen Maßnahmenkatalog umgesetzt werden kann. Hierdurch soll ein dem Risiko angemessenes Schutzniveau für die Daten des Betroffenen gewährleistet werden (vgl. Art. 32 Satz 1 1. Hs.).

Die Schutzfunktion einer Pseudonymisierung wirkt sich gem. ErwG 28 auch durch das **Absenken von Risiken** für die betroffenen Personen aus, die einer Verarbeitung ihrer Daten ausgesetzt sind. Eine Spezifizierung dieser „Risiken" erfolgt im Zuge der Erläuterungen der DS-GVO in ErwG 75 zu den Verletzungen des Schutzes personenbezogener Daten, die auch als sog. „Datenpanne" bezeichnet werden können. Risiken können sich beispielsweise in einem physischen, materiellen oder immateriellen Schaden manifestieren, so z.B. in Gestalt eines Identitätsdiebstahls oder -betrugs, eines finanziellen Verlusts oder einer Rufschädigung. Pseudonyme Daten können diese **Risiken mindern**, indem sie im Falle eines Abhandenkommens von Daten keinen direkten Rückschluss zum Betroffenen zulassen oder einen solchen erschweren. Entsprechend sieht der europäische Gesetzgeber über ErwG 85 besondere **Informationspflichten** vor, sollte eine solche Pseudonymisierung aufgehoben werden. Ein Unterwandern dieser Schutzfunktion soll daher zu einem unmittelbaren Handeln des Verantwortlichen führen.

3.3.2. Ermöglichungs- und Erleichterungsfunktion

Sofern nach dem Verarbeitungszweck keine direkte Identifizierung nötig ist, jedoch eine Anonymisierung der Daten ausscheidet, kann die Pseudonymisierung der Daten dem Schutz der Betroffenen Rechnung tragen. Der Rückbezug auf einen konkreten Menschen ist den mit der Verarbeitung befassten Personen, die keinen Zugang zu dem Schlüssel haben ebenso wenig möglich wie bei anonymisierten Daten, da zusätzliche Informationen, mit denen die

5 Siehe hierzu ergänzend Ziff. 5.2.3.
6 Vgl. Art. 25.

7 ErwG 28 Satz 2 führt diesbezüglich aus: „Pseudonymisierung in dieser Verordnung ist nicht beabsichtigt, andere Datenschutzmaßnahmen auszuschließen."

personenbezogenen Daten einer speziellen betroffenen Person zugeordnet werden können, gesondert aufbewahrt werden.

Im Rahmen des risikobasierten Ansatzes der DS-GVO ist es daher gerechtfertigt, dass sich eine Pseudonymisierung auch zugunsten des Verantwortlichen auswirkt. Die Pseudonymisierung kann Verarbeitungen zulässig machen, die ansonsten nicht zulässig wären, was insbesondere heute, im Zeitalter von **Big Data** und **Internet of Things**, von wesentlicher Bedeutung ist.

Ein wichtiges Beispiel hierfür ist Art. 6 Abs. 4, der für zweckändernde Datenverarbeitungen gilt. Ob ein neuer Verarbeitungszweck mit dem ursprünglichen Zweck vereinbar ist und ob die Weiterverarbeitung daher auf die ursprüngliche Rechtsgrundlage gestützt werden kann, ist das Ergebnis einer Abwägung. Der Gesetzgeber hat in Art. 6 Abs. 4 verschiedene Kriterien genannt, die dabei zu berücksichtigen sind. Ein Kriterium, das für eine Kompatibilität der Zwecke spricht, ist das Vorhandensein geeigneter Garantien[8], wozu auch die Pseudonymisierung gehören kann. Die Pseudonymisierung ist in der Regel nicht allein entscheidend, sie kann aber den Ausschlag für eine Zulässigkeit der Weiterverarbeitung geben.

Ein Sonderfall der Vereinbarkeit einer Verarbeitung mit dem ursprünglichen Zweck ist nach Art. 5 Abs. 1 Buchst. b die Verarbeitung für im öffentlichen Interesse liegende Archivzwecke, für wissenschaftliche oder historische Forschungszwecke oder für **statistische Zwecke**. Wenn hierfür die Anforderungen des Art. 89 Abs. 1 eingehalten werden, sind diese Verarbeitungen mit dem ursprünglichen Zweck vereinbar. Das setzt geeignete Maßnahmen zum Schutz der Rechte und Freiheiten der betroffenen Person voraus. Zu diesen Maßnahmen kann die Pseudonymisierung gehören, sofern es möglich ist, die oben genannten Zwecke auf diese Weise zu erfüllen (Art. 89 Abs. 1 Satz 2). Daher kann die Pseudonymisierung dazu beitragen, dass Statistiken und Forschungsvorhaben zulässig werden.

Über die in der Verordnung ausdrücklich genannten Fälle hinaus kann die Pseudonymisierung von Daten im Rahmen von Normen eine Rolle spielen, die eine **Interessenabwägung** vorsehen. Ein Beispiel ist Art. 6 Abs. 1 Buchst. f. Die Norm erlaubt eine Datenverarbeitung, die zur Wahrung berechtigter Interessen des Verarbeiters erforderlich ist. Voraussetzung ist, dass die Interessen oder Grundrechte und Grundfreiheiten der betroffenen Person, die den Schutz der personenbezogenen Daten erfordern, nicht überwiegen. In die Interessenabwägung kann zugunsten des Verarbeiters einfließen, dass er die Daten pseudonymisiert hat.[9]

8 Art. 6 Abs. 4 Buchst. e, dazu auch Art. 29-Datenschutzgruppe, WP203, Seite 25 ff. sowie Monreal, ZD 2016, 507, 511, nach dessen Ansicht durch geeignete Garantien Defizite ausgeglichen werden können, die sich bei Anwendung der anderen Abwägungskriterien ergeben haben.

9 So z.B. Buchner/Petri, in: Kühling/Buchner, DS-GVO, Art. 6 Rn. 154.

Schließlich kann die Pseudonymisierung von Daten den für die Datenverarbeitung Verantwortlichen von der Erfüllung bestimmter **datenschutzrechtlicher Pflichten befreien oder diese Pflichten reduzieren**. Ein Beispiel dafür ist Art. 11 Abs. 1, der eine Einschränkung der Pflicht zur **Erfüllung der Betroffenenrechte** enthält. Ist bei der Verarbeitung durch einen Verantwortlichen die Identifizierung der betroffenen Person nicht oder nicht mehr erforderlich, so ist der Verantwortliche nicht verpflichtet, zur bloßen Einhaltung der Pflichten der Datenschutz-Grundverordnung zusätzliche Informationen aufzubewahren, einzuholen oder zu verarbeiten, um die betroffene Person zu identifizieren. Damit wird dem Gebot der Datensparsamkeit nach Art. 5 Abs. 1 Buchst. c entsprochen. Die Anforderungen an die Erfüllung der Betroffenenrechte verändern sich damit nach Art. 11 Abs. 2.

Inwiefern die Pseudonymisierung von Daten die Ermöglichungs- oder Erleichterungsfunktion entfalten kann, wird in unterschiedlichen Branchen und Verarbeitungssituationen verschieden zu beurteilen sein. Daher kann es sinnvoll sein, Regelungen zur Pseudonymisierung in bestimmten Fallkonstellationen in branchenspezifischen Codes of Conduct zu treffen. Art. 40 Abs. 2 Buchst. d sieht dies ausdrücklich vor.

4. Verfahren und technisch-organisatorische Anforderungen

Zur Umsetzung einer Pseudonymisierung können unterschiedliche Verfahren eingesetzt werden. Beispielsweise kann eine **Zuordnungstabelle** verwendet werden, in der jedem Klartextdatum ein oder mehrere Pseudonyme zugeordnet werden. Sofern Zugriff auf die Zuordnungstabelle besteht, kann durch Auslesen der entsprechenden Einträge einem Pseudonym das zugehörige Klartextdatum zugeordnet werden. Entsprechend ist hier der Zugriff auf die Zuordnungstabelle einzuschränken. Alternativ können zur Pseudonymisierung verschiedene **kryptographische Verfahren** eingesetzt werden, die jeweils ein Klartextdatum in ein oder mehrere Pseudonyme überführen. Über den Zugriff auf die verwendeten kryptographischen Schlüssel und ggf. weitere Parameter kann hier die Umkehrbarkeit der Pseudonymisierung gesteuert/eingeschränkt werden. Die nachfolgenden Abschnitte betrachten insbesondere kryptographischen Verfahren und deren Verwendung zur Pseudonymisierung.

4.1. Kryptographische Grundlagen und Verfahren

Im Folgenden wird ein Überblick zu wesentlichen Begrifflichkeiten im Umfeld kryptographischer Verfahren und deren Nutzung zur Pseudonymisierung gegeben:

Eine **Einwegfunktion** ist eine „leicht" zu berechnende mathematische Funktion, die nur „schwer" umzukehren ist. Das heißt, man kann „leicht" den Funktionswert einer Eingabe berechnen, aber die Umkehrung, nämlich aus diesem Funktionswert wieder die Eingabe zu bestimmen, ist „schwer". Die Begriffe „leicht" und „schwer" sind hier komplexitätstheoretisch zu verstehen. Den Begriff „schwer" kann man sehr vereinfachend verstehen als „ist in angemessener Zeit praktisch nicht durchführbar".

Eine **kryptographische Hashfunktion** ist eine Einwegfunktion, die kollisionsresistent ist. Sie ordnet einer Eingabe beliebiger Länge einen Hashwert fester Länge zu. Da die Eingabemenge größer als die Ausgabemenge ist, existieren Kollisionen. Durch die Eigenschaft „kollisionsresistent" ist die Berechnung dieser Kollisionen, d.h. verschiedene Eingaben mit gleichem Hashwert, jedoch praktisch unmöglich. Ein Beispiel einer kryptographischen Hashfunktion ist SHA-256, die Hash-Werte mit einer Länge von 32 Byte berechnet.

Ein **Verschlüsselungsverfahren** transformiert unter Verwendung eines Schlüssels einen Klartext in einen Chiffretext. Die Entschlüsselung ist der umgekehrte Vorgang; hier wird der Chiffretext wieder in den ursprünglichen Klartext transformiert. Bei **symmetrischen Verschlüsselungsverfahren** wird der gleiche Schlüssel zur Ver- und Entschlüsselung verwendet. **Asymmetrische Verschlüsselungsverfahren** verwenden ein Schlüsselpaar bestehend aus einem öffentlichen und einem privaten Schlüssel. Der öffentliche Schlüssel wird zur Verschlüsselung, und der private Schlüssel wird zur Entschlüsselung verwendet.

Normalerweise sind Verschlüsselungsverfahren **deterministisch**, d.h. der gleiche Klartext wird (mit dem gleichen Schlüssel) immer in den gleichen Chiffretext transformiert. Erzeugt ein Verschlüsselungsverfahren für einen Klartext bei Verwendung des gleichen Schlüssels bei jeder Verschlüsselung unterschiedliche Chiffretexte, so wird es **probabilistisch** genannt. Jedes deterministische Verschlüsselungsverfahren kann zu einem probabilistischen Verfahren überführt werden, in dem vor der deterministischen Verschlüsselung ein neu generierter Zufallswert an den Klartext z.B. angehängt wird und der Zufallswert nach der deterministischen Entschlüsselung wieder entfernt wird.

4.2. Anforderungen an Pseudonyme

Für eine Verarbeitbarkeit der pseudonymisierten Daten kann es erforderlich sein, dass die erzeugten Pseudonyme bestimmte Eigenschaften der zugrundeliegenden Klartexte enthalten. Dazu werden diese zuvor festgelegten Eigenschaften der ansonsten verdeckten Klartexte verfügbar gemacht. Diese werden als **Verfügbarkeitsanforderungen** an eine Pseudonymisierung bezeichnet. Pseudonyme, die bestimmten Verfügbarkeitsanforderungen gerecht werden,

werden hier im Folgenden Pseudonyme mit Verfügbarkeitsoptionen genannt. Mit einer Pseudonymisierung wird die Vertraulichkeit der zu schützenden Daten gewährleistet und somit die Privatsphäre ihrer Besitzer (Privacy) geschützt. Verfügbarkeitsoptionen von Pseudonymen machen Teile der in den zugrundeliegenden Daten enthaltenen Informationen verfügbar und gewähren somit trotz Vertraulichkeit ein gewisses Maß an Nutzbarkeit (Utility). Sie stellen somit „Utility trotz Privacy" bereit.

4.2.1. Verfügbarkeitsanforderungen

Eine mögliche Verfügbarkeitsoption ist die **Aufdeckbarkeit** des dem Pseudonym zugrundeliegenden Klartextes unter bestimmten Voraussetzungen. Dies kann erreicht werden, indem das Pseudonym durch Verschlüsselung des Klartextes generiert wird. So kann bei Kenntnis des Verfahrens und des verwendeten Schlüssels das Pseudonym entschlüsselt und somit der zugrundeliegende Klartext aufgedeckt werden. Ggf. ist die Aufdeckbarkeit an die Erfüllung eines spezifischen Zwecks bzw. an die Verwendung durch eine bestimmte Rolle gebunden.

Eine weitere Verfügbarkeitsoption ist die **Verkettbarkeit** bzgl. einer Relation r. Wenn Verkettbarkeit hinsichtlich eines spezifischen Zusammenhangs gegeben ist, kann für je zwei Pseudonyme bestimmt werden, ob die zugrundeliegenden Klartexte in dem spezifischen Zusammenhang zueinanderstehen. Ein einfaches Beispiel sei die Verkettbarkeit hinsichtlich der Gleichheit. Pseudonyme, die diese Verfügbarkeitsoption erfüllen, bieten die Möglichkeit zu überprüfen, ob bei je zwei gegebenen Pseudonymen die zugrundeliegenden Klartexte gleich sind. Auch die Verkettbarkeit kann an bestimmte Rollen oder Zwecke gebunden sein.

4.2.2. Rollenbindung

Verfügbarkeitsoptionen können an bestimmte Rollen gebunden sein. Sind in einem System verschiedene Rollen definiert, so können einzelnen Rollen bestimmte Verfügbarkeitsoptionen zugeordnet sein. So wird ermöglicht, dass ausschließlich diese Rollen die durch die Verfügbarkeitsoptionen bereitgestellten Informationen einsehen können.

4.2.3. Zweckbindung

Verfügbarkeitsoptionen können an die Verwendung für bestimmte Zwecke gebunden sein. So kann ein System den Zugriff auf ein Pseudonym, welches einer bestimmten Verfügbarkeitsoption genügt, an den Nachweis des Vorhandenseins einer bestimmten, einen Zweck beschreibenden Situation binden. Die Zweckbindung von Verfügbarkeitsoptionen kann technisch oder organisatorisch durchgesetzt werden.

Bei der **technischen Zweckbindung** wird die Zweckbindung einer Verfügbarkeitsoption ohne das Eingreifen einer natürlichen

Person mit rein technischen Mitteln durchgesetzt. Ein Beispiel für die technischen Mittel ist die Überprüfung von Systemwerten (die den Zweck charakterisieren), wie z.B. die Systemzeit vor Freigabe einer Verfügbarkeitsoption. Ein weiteres Beispiel ist die Verwendung von Schwellwertverfahren. Treten beispielsweise bestimmte in Datensätzen dokumentierte Ereignisse mit großer Häufigkeit auf, so wird die betroffene Verfügbarkeitsoption nur bei Überschreiten eines definierten Häufigkeits-Schwellwertes freigegeben.

Bei der **organisatorischen Zweckbindung** werden eine oder mehrere natürliche Personen mit der Möglichkeit ausgestattet, bestimmte Verfügbarkeitsoptionen freizugeben. D.h. eine natürliche Person tritt in Interaktion mit dem System, um das Zutreffen des Zwecks zu überprüfen. Um die organisatorische Zweckbindung sicherzustellen, können die betroffenen Pseudonyme zusätzlich mit einem ausschließlich der natürlichen Person bekannten Schlüssel verschlüsselt sein.

Abschließend sind die Vorteile der technischen gegenüber der organisatorischen Zweckbindung von Verfügbarkeitsoptionen hervorzuheben. Im Gegensatz zur organisatorischen Zweckbindung ist zur Durchsetzung der technischen Zweckbindung keine direkte Interaktion einer natürlichen Person mit dem System erforderlich. Eine Ausnutzung von Machtverhältnissen wird somit eingeschränkt. Das zugrundeliegende Vertrauensmodell kann direkt aus der Sicherheit der implementierten Techniken abgeleitet werden. Durch die rein technische Implementierung kann die automatisierte Durchsetzung der Zweckbindung in Echtzeit geschehen.

4.3. Beispiele für Pseudonymisierungsverfahren zur Umsetzung von Verfügbarkeitsanforderungen

Im Folgenden werden Beispielverfahren zur Konstruktion von Pseudonymen mit spezifischen Verfügbarkeitsoptionen dargestellt. Hinsichtlich einer Verkettbarkeit der Daten wird die Gleichheit der einer Pseudonymisierung zugrunde liegenden Klartexte betrachtet.

4.3.1. Verkettbare aufdeckbare Pseudonyme

Zur Konstruktion verkettbarer aufdeckbarer Pseudonyme können deterministische Verschlüsselungsverfahren verwendet werden. Da durch diese Verfahren gleiche Klartexte auf gleiche Chiffretexte (Pseudonyme) abgebildet werden, ist Verkettbarkeit (auch ohne Kenntnis des Ver- bzw. Entschlüsselungsschlüssels) gegeben. Bei Kenntnis des Entschlüsselungsschlüssels ist Aufdeckbarkeit gegeben.

4.3.2. Nicht-verkettbare aufdeckbare Pseudonyme

Unter Verwendung probabilistischer Verschlüsselungverfahren können nicht-verkettbare aber aufdeckbare Pseudonyme konstruiert werden. Da probabilistische Verschlüsselungsverfahren auch gleiche Klartexte auf unterschiedliche Chiffrate (Pseudonyme) abbilden, ist keine Verkettbarkeit von Pseudonymen gegeben. Bei Kenntnis des Entschlüsselungsschlüssels ist die Aufdeckbarkeit von Pseudonymen gewährleistet.

4.3.3. Verkettbare nicht-aufdeckbare Pseudonyme

Verkettbare aber nicht-aufdeckbare Pseudonyme können unter Verwendung deterministischer Einwegfunktion (z.B. kryptographische Hash-Funktionen) erstellt werden. Da deterministische Verfahren gleiche Klartexte auf gleiche Ergebniswerte (Pseudonyme) abbilden, ist Verkettbarkeit sichergestellt. Durch die Nutzung einer Einwegfunktion soll die Umkehrung der Pseudonymisierung, also Aufdeckung, verhindert werden.

4.3.4. Rollenbindung

Eine Rollenbindung von Verfügbarkeitsoptionen kann durch eine ggf. zusätzliche probabilistische Verschlüsselung des Pseudonyms, mit einem nur der definierten Rolle bekannten Entschlüsselungsschlüssel, erfolgen. Dadurch ist das ggf. verkettbare oder aufdeckbare Pseudonym nur durch die definierte Rolle zugreifbar. Die alleinige Rollenbindung der Aufdeckbarkeit von Pseudonymen kann erreicht werden, indem der zur Aufdeckung erforderliche Entschlüsselungsschlüssel ausschließlich der definierten Rolle zur Kenntnis gegeben wird.

4.3.5. Organisatorische Zweckbindung

Eine organisatorische Zweckbindung von Verfügbarkeitsoptionen kann durch Rollenbindung an die zweckprüfende Person erreicht werden. Im Falle einer zweckgebundenen Aufdeckbarkeit kann aufgrund einer gegebener Verkettbarkeit von Pseudonymen das Vorhandensein einer Ausnahmesituation erkannt und beispielsweise Abhilfemaßnahmen auf Basis von Klardaten eingeleitet werden. Entsprechend kann eine zweckprüfende Person eine auf Basis einer zuvor definierten Ausnahmesituation zweckbeschränkte Aufdeckbarkeit umsetzen, in dem sie den nur ihr bekannten Entschlüsselungsschlüssel einsetzt, und das durch Verschlüsselung erstellte Pseudonym entschlüsselt.

4.3.6. Technische Zweckbindung

Auch im Zuge einer technischen Zweckbindung kann beispielsweise eine Zweckbindung auf Ausnahmesituationen beschränkt sein, die sich z.B. durch das gehäufte Auftreten von Datensätzen mit Pseudonymen, die den gleichen Klartext repräsentieren, manifestiert und mit einem Häufigkeitsschwellwert verknüpft ist. Hier kann ein kryptographisches Secret-Sharing-Verfahren[10] genutzt werden, das für jeden Klar-

10 Siehe hierzu A. Shamir. How to share a secret. In: Communications of the ACM Bd. 22, ACM, 1979, Seiten 612–613.

textwert einen geheimen Gesamtschlüssel verwaltet und dem Pseudonym bei jeder Pseudonymisierung eines Klartextwertes einen einzigartigen Teilschlüssel des zugehörigen Gesamtschlüssels dem Pseudonym beilegt. Liegt eine über einem definierten Schwellwert liegende Anzahl von Pseudonymen zu einem Klartext vor, kann aus den beigefügten Teilschlüsseln der Gesamtschlüssel bestimmt und zur Entschlüsselung verwendet werden.

4.4. Technisch-organisatorische Anforderungen

Die in der DS-GVO geforderten erforderlichen technisch-organisatorischen Maßnahmen bei der Pseudonymisierung können wie folgt konkretisiert werden:

- Für die Pseudonymisierung müssen **State-of-the-Art**-Transformationsverfahren (vgl. z.B. BSI- oder ENISA-Richtlinie zu Kryptoverfahren) verwendet und – insbesondere bei langfristig verwendeten pseudonymisierten Daten – durch jeweils aktuelle Verfahren ausgetauscht werden. Dies soll ein Höchstmaß an Sicherheit bieten.
- Die Pseudonymisierung ist im Verarbeitungsprozess so früh wie möglich durchzuführen.
- Bei Klartextdaten aus kleinen Wertebereichen bzw. mit geringer Streuung innerhalb des Wertebereichs sind Pseudonymisierungsverfahren anfällig für eine Pseudonymaufdeckung durch Aufzählungsangriffe bspw. unter Verwendung von Rainbow-Tabellen. Bei diesen Angriffen wird durch einen Angreifer eine Klartext-Pseudonym-Zuordnungstabelle für alle (wenigen) möglichen Klartexte berechnet oder eine vorgefertigte genutzt. Mittels dieser Tabelle können dann gegebenen Pseudonymen die entsprechenden Klartexte zugeordnet werden. Durch die Verwendung von sogenannten **Salt-Werten** kann der Aufwand für den Angreifer erhöht und damit das Risiko reduziert werden. Hierzu wird vor Anwendung des Pseudonymisierungsverfahrens abhängig vom Kontext (z.B. pro Datensatz) ein Salt-Wert gewählt und mit dem Klartextwert kombiniert. Dadurch können der Wertebereich und die Streuung der Werte vergrößert werden und vorausberechnete bzw. vorgefertigte Zuordnungstabellen nicht mehr verwendet werden, weil sie den mit dem Klartext kombinierten Salt-Wert nicht berücksichtigen. Sofern zu gleichen Klartexten die gleichen Pseudonyme erstellt werden sollen, muss die Verwendung der gleichen Salt-Werte sichergestellt und eine geeignete Speicherung der Salt-Werte umgesetzt werden.
- Die Erzeugung und Verwaltung (u.a. Verteilung, Speicherung, Verwendung, Löschung) geheimer Parameter (Schlüssel und Salt-Werte) sind durch nach **Stand der Technik** geeignete technische und organisatorische Maßnahmen zu realisieren.
- Abhängig vom Anwendungsfall sind – zeit- oder datenvolumenabhängig – geeignete Intervalle zu definieren, in denen ein Wechsel verwendeter geheimer Parameter (Salt-Wert und Schlüssel) erfolgt.
- Der Zugriff auf Salt-Werte und Schlüssel muss auf ein absolutes Minimum an vertrauenswürdigen Nutzern eingeschränkt werden (Need-to-Know-Prinzip).
- Einbindung des Pseudonymisierungskonzepts in ein IT-Sicherheitsmanagement (z.B. nach ISO/IEC 27001), um einen unbefugten Zugriff auf pseudonymisierte Daten zu verhindern.
- Pseudonymisierte Daten sind nach Wegfall des Verarbeitungszwecks datenschutzgerecht zu löschen.

5. Transparenz und Betroffenenrechte bei pseudonymisierten Daten

5.1. Generelle Anforderungen an Transparenz

Transparenz muss grundsätzlich über die Tatsache der Datenverarbeitung, die verarbeiteten Daten, die Zwecke und Funktionsweise der Datenverarbeitung und zahlreiche weitere die Datenverarbeitung betreffende Gesichtspunkte hergestellt werden (siehe hierzu insbesondere die Katalogtatbestände der Art. 13 Abs. 1 und 2, 14 Abs. 1 und 2 sowie 15 Abs. 1 und 2).

Ziel der Herstellung von Transparenz ist es, die Ungewissheit der betroffenen Person über die Datenverarbeitung zu beseitigen, sie zum Selbstdatenschutz zu befähigen und es ihr zu ermöglichen, ihre Rechte im Zusammenhang mit der Verarbeitung personenbezogener Daten auszuüben. Dies muss in **präziser, transparenter, verständlicher und leicht zugänglicher Form** in einer **klaren und einfachen Sprache** erfolgen (Art. 12 Abs. 1 Satz 1). Die Informationspflichten der Art. 13 und 14 können auch durch geeignete, für betroffene Personen verständliche Icons unterstützt werden (Art. 12 Abs. 7).

5.2. Besonderheiten bei pseudonymisierten Daten

5.2.1. Informationspflichten (Art. 13 und 14 DS-GVO)

5.2.1.1. Ausgangslage

Für die Grundkonstellationen der Informationspflichten der Art. 13 und 14 ergeben sich kaum Besonderheiten bei der Verarbeitung pseudonymisierter Daten. Die Informationen gemäß Art. 13 Abs. 1 und 2 sind **„zum Zeitpunkt der Erhebung"** der Daten beim Betroffenen zu erteilen. Die Informationen gemäß Art. 14 Abs. 1 und 2 sind „innerhalb einer **angemessenen Frist** nach Erlangung der personenbezogenen Daten" (vgl. Art. 14 Abs. 3 Buchst. a) zu erteilen. Zum Zeitpunkt der Datenerhebung oder kurz nach Datenerlangung dürften die Daten in der Regel noch nicht in pseud-

onymisierter Form vorliegen, so dass eine Information des Betroffenen unproblematisch möglich ist. Auch in der Konstellation des Art. 14 Abs. 3 Buchst. b, in der die personenbezogenen Daten zur Kommunikation mit dem Betroffenen verwendet werden sollen (z.B. beim Direktmarketing), ergeben sich keine Besonderheiten, denn die Kommunikation „mit" dem Betroffenen schließt ja gerade aus, dass die Zuordnung der Daten zum Betroffenen beseitigt wurde.

5.2.1.2. Sonderfall der Offenlegung pseudonymisierter Daten an Dritte

Problematisch können hingegen die Fälle des Art. 14 Abs. 3 Buchst. c und insbesondere die Fälle der Weiterverarbeitung für einen anderen Zweck gemäß Art. 13 Abs. 3 und Art. 14 Abs. 4 sein. Hat der Verantwortliche vor der Offenlegung an einen anderen Empfänger oder vor der beabsichtigten Weiterverarbeitung eine Pseudonymisierung vorgenommen, kann der Empfänger den Betroffenen nicht mehr ohne Weiteres informieren, denn ohne Hinzuziehung weiterer Informationen kann er den Betroffenen nicht mehr identifizieren. Es wird in diesen Fällen empfohlen, dass der Dritte, der pseudonyme Daten verarbeitet, **allgemeine Informationen** über die eigene Webseite bereitstellt, dass pseudonyme Daten verarbeitet werden.

Die Information gemäß Art. 13 und 14 DS-GVO muss nicht individuell erfolgen, sondern kann generell z.B. über eine Webseite, Vertragsgrundlagen oder Aushänge vorab vorgenommen werden. Dies ergibt sich insbesondere aus ErwG 58 S. 1 bis 3. Gegen die individuell zu adressierende Information spricht aber auch, dass individuelle Benachrichtigungen (etwa via E-Mail) zu einem Informationsüberfluss beim Betroffenen führen würden – mit der Folge, dass die einzelne Information von den meisten Betroffenen nicht mehr wahrgenommen würde. Auch im Rahmen von Art. 13 Abs. 3 und Art. 14 Abs. 4, wenn erst später über eine weitere Verwendung der Daten entschieden wird und damit nicht bereits bei Erhebung oder Erstverwendung informiert werden kann, kann die Information über öffentlich zugängliche Quellen (z.B. im Rahmen von Kampagnen auch auf der Webseite oder in anderen Medien) erfolgen.

5.2.2. Betroffenenrechte gemäß Art. 15 bis 22 und Art. 34

Für die Erfüllung der Auskunftspflicht des Art. 15 und der Betroffenenrechte der Art. 16 bis 22, nach Vornahme einer Pseudonymisierung, gilt Folgendes:

Ist eine Identifizierung des Betroffenen durch den Verantwortlichen nicht oder nicht mehr erforderlich (Art. 11), was nach Vornahme einer Pseudonymisierung häufig der Fall sein wird, gelten besondere Regeln:

a) Nach Art. 11 Abs. 1 ist der Verantwortliche zur bloßen Einhaltung von Pflichten der DS-GVO nicht verpflichtet, zur Identifizierung der betroffenen Person zusätzliche Informationen aufzubewahren, einzuholen oder zu verarbeiten.

b) Kann der Verantwortliche nachweisen, dass er nicht in der Lage ist, die betroffene Person zu identifizieren, muss er die betroffene Person darüber – sofern möglich (also in der Regel bei den antragsabhängigen Betroffenenrechten) – unterrichten (Art. 11 Abs. 2).

In den Fällen **a)** und **b)** finden die Betroffenenrechte gemäß Art. 15 bis 20 DS-GVO keine Anwendung. Das betrifft die Rechte auf Auskunft, Berichtigung, Vervollständigung, Löschung (außer bei Wegfall des Verarbeitungszwecks, dann gilt eine Löschverpflichtung), Einschränkung der Verarbeitung, Mitteilung und Datenübertagung. Dies gilt nur dann nicht, wenn es der betroffenen Person gelingt, durch **Bereitstellung zusätzlicher Informationen** die Identifizierung doch noch zu ermöglichen (Art. 11 Abs. 2 Satz 2 2. Hs.).

Anwendung finden dagegen die Betroffenenrechte gemäß Art. 21 und 22.[11] Der Betroffene kann demnach weiterhin Widerspruch gegen die folgenden Datenverarbeitungen[12] einlegen (Art. 21):

- Verarbeitung personenbezogener Daten aufgrund eines berechtigten Interesses (gem. Art. 6 Abs. 1 Buchst. f), wozu auch ein auf das berechtigte Interesse gestütztes Profiling gehört,
- Verarbeitung, die für die Wahrnehmung einer Aufgabe erforderlich ist, die im öffentlichen Interesse liegt, oder die in Ausübung öffentlicher Gewalt, die dem Verantwortlichen übertragen wurde, erfolgt (Art. 6 Abs. 1 Buchst. e),
- Verarbeitung zu wissenschaftlichen oder historischen Forschungszwecken oder zu statistischen Zwecken gemäß Art. 89 Abs. 1 (Art. 21 Abs. 6),
- Verarbeitung zu Zwecken der Direktwerbung (Art. 21 Abs. 2).

Fraglich ist, welche Rolle das **Widerspruchsrecht gegen Direktwerbung** im Zusammenhang mit der Pseudonymisierung spielt. Versteht man unter Direktwerbung die individuelle Ansprache eines konkreten Betroffenen (via personalisiertem Brief oder personalisierter E-Mail), kommt eine solche Ansprache bei pseudonymisierten Daten ohnehin nicht in Betracht. Werbung, die durch

11 Art. 12 Abs. 2 verweist allerdings abweichend von Art. 11 Abs. 2 nicht nur auf die Betroffenenrechte der Art. 15 bis 20, sondern auch auf die Betroffenenrechte der Art. 21 und 22. Nach Art. 12 Abs. 2 S. 1 scheint der Verantwortliche somit ein Weigerungsrecht auch in den Fällen des Widerspruchsrechts (Art. 21) und der automatisierten Einzelentscheidung (Art. 22) zu haben. Art. 11 Abs. 2 und Art. 12 Abs. 2 sind offensichtlich widersprüchlich.

12 Dies gilt sowohl bei personenbezogenen Datenverarbeitungen als auch bei pseudonymisierten Datenverarbeitungen.

Tracker, wie z.B. Cookies, Mobile Identifier, Fingerprinting etc. ermöglicht wird, würde daher nicht dem Widerspruchsrecht des Art. 21 Abs. 2 unterliegen, da diese Tracker keine direkte Identifikation einer Person ermöglichen. Ein Widerspruch gemäß Art. 21 Abs. 1 bei Werbung unter Verwendung pseudonymisierter Daten bliebe weiterhin möglich. Im Unterschied zum Widerspruch nach Art. 21 Abs. 2, der keine weiteren Voraussetzungen hat, müssen beim Widerspruch nach Art. 21 Abs. 1 besondere Gründe des Betroffenen vorliegen und es findet eine Interessenabwägung statt.

Ferner hat eine betroffene Person das Recht, nicht einer ausschließlich auf automatisierten Verarbeitung – einschließlich Profiling – beruhenden Entscheidung unterworfen zu werden, die ihr gegenüber rechtliche Wirkung entfaltet oder sie in ähnlicher Weise erheblich beeinträchtigt (Art. 22).

Wenn der Betroffene erfolgreich Widerspruch eingelegt hat, wird der Original-Datensatz des Betroffenen für die Datenverarbeitungszwecke, gegen die der Betroffene Widerspruch eingelegt hat, **gesperrt**. Erfolgt der Widerspruch vor Vornahme einer Pseudonymisierung, ist die Sperrung unproblematisch. Erfolgt der Widerspruch nach Vornahme einer Pseudonymisierung, muss der Verantwortliche zunächst eine **Re-Identifizierung** des Betroffenen vornehmen, bevor er die erforderliche Sperrung vornehmen kann. In der Regel kann der Datensatz aber aus den schon einmal pseudonymisierten Daten nicht mehr herausgenommen werden, da es sich meist um automatisierte Datenverarbeitung in großem Umfang handelt.

Zusammenfassend: Den verantwortlichen Stellen, die die pseudonyme Weiterverarbeitung von personenbezogenen Informationen beabsichtigen, wird empfohlen, dies den Betroffenen rechtzeitig vorher transparent zu machen. Jedenfalls sollte ein Widerspruchsrecht derart eingeräumt werden, dass die Daten des Betroffenen mit Wirkung für die Zukunft nach Ausübung des Widerspruchsrechts nicht mehr in die pseudonyme Weiterverarbeitung einfließen.

Als weiteres Betroffenenrecht gilt der Art. 34, nach dem der Verantwortliche die betroffene Person unverzüglich zu benachrichtigen hat, wenn die **Verletzung des Schutzes personenbezogener Daten (sog. „Datenpanne")** voraussichtlich ein hohes Risiko für die persönlichen Rechte und Freiheiten natürlicher Personen zur Folge hat. Hier gilt: Vorbehaltlich einer Prüfung des konkreten Einzelfalls besteht eine gewisse Wahrscheinlichkeit, dass eine Schutzverletzung dann nicht zu einem hohen Risiko für den Betroffenen führt, wenn der Verantwortliche zuvor bereits eine Pseudonymisierung vorgenommen hatte. Wenn z.B. ein Datendiebstahl stattgefunden hat, dann dürfte es in den meisten Fällen für den Täter erst recht nicht mehr möglich sein, den Betroffenen zu identifizieren. Es wird empfohlen, dass die Betroffenen über definierte Kanäle **benachrichtigt werden**. Eine individuelle Ansprache der Betroffenen wird als nicht notwendig erachtet.

5.2.3. Übermittlung pseudonymisierter Daten an Dritte

Hat der Verantwortliche die pseudonymisierten Daten an Dritte weitergegeben, muss der Empfänger der Daten prüfen, ob die Daten für ihn als zuordenbare Daten im Sinne von Art. 11 anzusehen sind. Der Empfänger der Daten hat noch mehr als der Erstverantwortliche das Problem, die Betroffenenrechte nicht erfüllen zu können, weil er die Zuordnung zu einer Person, die ggf. von ihren Rechten Gebrauch machen will, nicht herstellen kann.

5.3. Rückbeziehung der Ergebnisse auf eine Person

Eine Rückbeziehung der Pseudonyme, die auf einer höherwertigen Pseudonymisierung (z.B. mittels eines anerkannten Hashing-Algorithmus) beruhen, auf eine natürliche Person ist nur für denjenigen möglich, der den Schlüssel besitzt, mit dem die Daten pseudonymisiert wurden.

Sollen die Daten auf eine konkrete Person rückbezogen werden und dient dieser Rückbezug nicht der Erfüllung von Verpflichtungen gegenüber der betroffenen Person, ist dafür die **Einwilligung** der betroffenen Person erforderlich. Der Verantwortliche für die Datenverarbeitung besitzt diesen Schlüssel und muss ihn einsetzen, wenn der Betroffene seine Wahlrechte ausübt. Dann muss er nämlich, je nach Wahl des Betroffenen, über die Daten Auskunft geben, die Daten berichtigen oder löschen. Hierbei muss er seinen eventuell eingesetzten Dienstleister, der für ihn die Daten verarbeitet, einbeziehen.

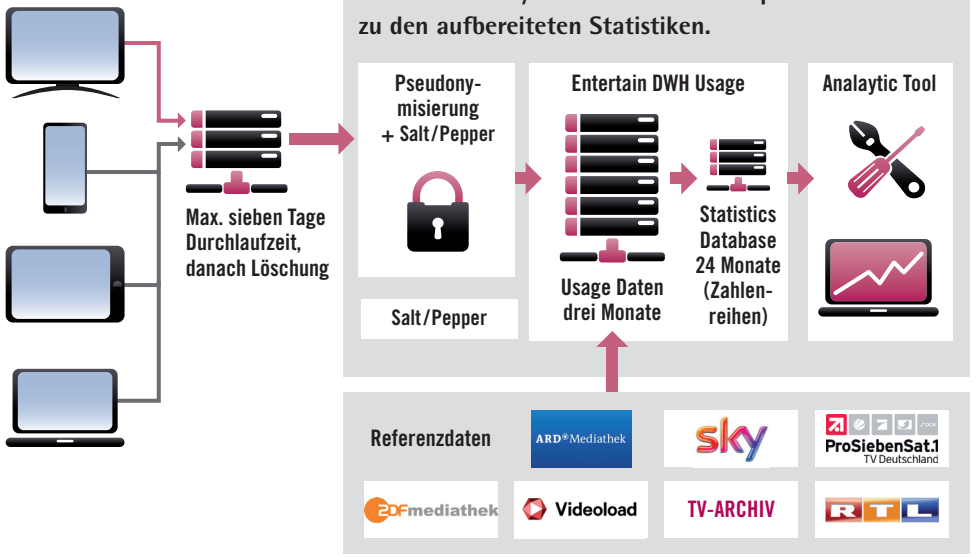
6. Anwendungsszenarien

6.1. Pseudonymisierung Entertain TV (DTAG)

Unter dem Produktnamen Entertain TV vermarktet die Deutsche Telekom den Bezug von Fernsehprogrammen und Filmen über das Internet. Zur Nutzung des Produkts wird den Kunden eine Set-Top-Box zur Verfügung gestellt. Unter anderem aufgrund bestehender Verpflichtungen gegenüber den Fernsehanstalten werden Statistiken über das Nutzungsverhalten der Fernsehzuschauer erstellt.

6.1.1. Übersicht

Das folgende Diagramm zeigt den Datenfluss durch die einzelnen Systeme von der Set-Top-Box bis hin zu den aufbereiteten Statistiken.



6.1.2. Datenerzeugung

Bei der Benutzung einer Entertain Set-Top-Box, d.h. beim Betätigen der zugehörigen Fernbedienung durch den Benutzer, werden unterschiedliche Ereignisse generiert, je nachdem welche Tasten gedrückt wurden und in welchem Kontext sich der Benutzer befindet. Diese Ereignisse der Set-Top-Box stellen die Basis der Auswertungen dar. Durch diese werden beispielsweise Ein-/Ausschaltvorgänge, Kanalumschaltungen, Informationen zu den gesehenen Sendern oder Informationen zu Aktivitäten rund um das Aufnehmen bzw. Anschauen von Aufnahmen dokumentiert. Entsprechende Ereignis-Datensätze enthalten z.B. Informationen über die Set-Top-Box (Device-ID), die Account-ID des Kunden, Datum/Uhrzeit sowie weitere spezifische Datenfelder.

6.1.3. Pseudonymisierung

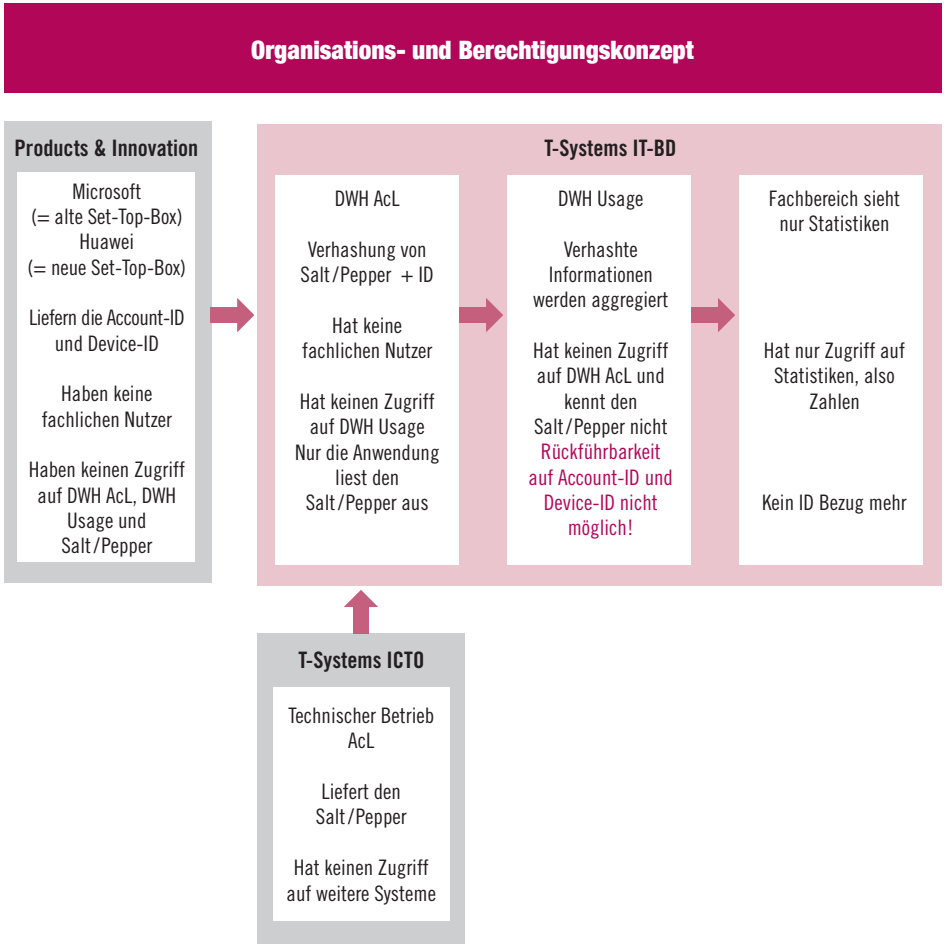
Die Account-ID ist ein Pseudonym für den Kunden und die Device-ID ist ein Pseudonym für die jeweilige Set-Top-Box. In den zur Auswertung notwendigen Ereignis-Datensätzen sind keine Attribute enthalten, welche direkt personenbezogene Daten enthalten. Organisatorisch getrennt werden Zuordnungstabellen verwaltet, die eine Zuordnung zwischen den Pseudonymen (Account-ID und Device-ID) und Kunden bzw. Set-Top-Boxen erlauben. Durch zusätzlichen Zugriff auf diese Zuordnungstabellen würden die Kennzeichen Device-ID und Account-ID letzten Endes einen Rückschluss auf den Kunden ermöglichen. Diese Rückführung auf den einzelnen Kunden ist z.B. notwendig für die Rechnungserstellung über die vertraglich bereit gestellte Leistung.

Da die Rückführung auf Klardaten hier im Zuge der Statistikerstellung jedoch nicht gewollt ist, werden die Kennzeichen Account-ID und Device-ID vor der Verarbeitung zusätzlich (nochmals) pseudonymisiert. Dabei erfolgt die Pseudonymisierung innerhalb des Organisationsbereichs Data Warehouse Acquisition Layer (DWH ACL) und die Verarbeitung (Statistikerstellung) wird in dem hiervon getrennten Organisationsbereich Data Warehouse Usage (DWH Usage) vorgenommen, vgl. Abbildung unten.

Das zu Grunde liegende Pseudonymisierungsverfahren führt dazu, dass die Statistikerstellung mittels Pseudonymen erfolgt, die für den verarbeitenden Organisationsbereich DWH Usage **verkettbar** aber **nicht-aufdeckbar** sind. Die Pseudonyme werden unter Verwendung eines deterministischen kryptografischen Hashfahrens (nämlich SHA-512) und eines einheitlichen Pseudonymisierungs-Salts- bzw. Peppers erstellt. Da deterministische Verfahren bei gleichem Salt/Pepper gleiche Klartexte auf gleiche Ergebniswerte (Pseudonyme) abbilden, ist die Verkettbarkeit der Pseudonyme sichergestellt. Aufgrund der Ausgabelänge des SHA-512 ist die Wahrscheinlichkeit einer Kollision vernachlässigbar ($< 10^{-70}$). Da der Bereich DWH Usage keinen Zugriff auf den Pseudonymisierungs-Salt bzw. Pepper besitzt, ist eine Rückführung der Pseudonyme und damit die Aufdeckung der Klartexte für diesen Bereich ausgeschlossen. Dies gilt auch bei Umgehung der Einwegfunktionalität mittels der Erstellung einer Look-Up Tabelle, welche ohne die Verwendung oder bei Kenntnis des Salts bzw. Peppers aufgrund der limitierten Menge von möglichen Eingangswerten gebildet werden kann.

Im Ergebnis führt das umgesetzte Pseudonymisierungsverfahren, dazu, dass kein Mitarbeiter der Deutschen Telekom Gruppe das Nutzungsverhalten eines bestimmten Kunden ansehen, auswerten oder anderen darüber eine Mitteilung machen kann.

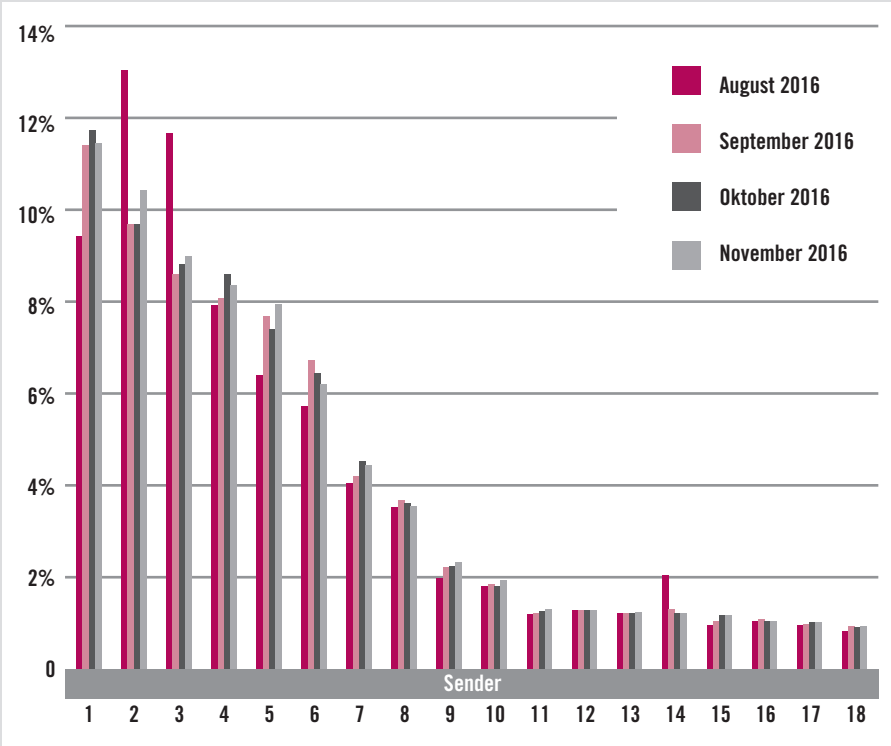
Die nachfolgende Abbildung skizziert das zugrunde liegende Organisations- und Berechtigungskonzept und die verschiedenen beteiligten Organisationsbereiche.



6.1.4. Statistikerstellung

Alle auf den Endkunden rückführbaren Attribute sind bereits mit der Account-ID und der Device-ID pseudonymisiert. Da verkettbare Pseudonyme genutzt werden, sind Zählungen möglich. Das heißt, es kann z.B. die Frage beantwortet werden, wie viele Haushalte oder wie viele Set-Top-Boxen zu einer bestimmten Zeit einen bestimmten Sender gesehen haben. In der erstellten anonymen Statistik sind Account-ID, Device-ID sowie die hierfür erstellten Pseudonyme letztlich nicht mehr vorhanden, und somit ist eine Rückführung von den reinen Zahlen auf die verhashten IDs unmöglich.

Auf der Basis der Verpflichtungen gegenüber Sendeanstalten übermittelt die Telekom ausschließlich anonyme Statistiken über das Nutzungsverhalten der Fernsehzuschauer, z.B. die Marktanteile anhand von relativen Zahlen, wie in folgender Abbildung veranschaulicht.



6.1.5. Opt-out

Jeder Entertain-Kunde wird durch die Datenschutzhinweise über die Erfassung der Daten zu statistischen Zwecken informiert. Die Telekom hat die Kunden vor Einführung dieser Auswertelösung darauf sowohl per E-Mail als auch durch „Pop-Ups“ in Entertain selbst hingewiesen.

Jeder Kunde hat die Möglichkeit, jederzeit der Erhebung und Auswertung der pseudonymen Nutzungsdaten zu widersprechen (Opt-out). Dieses Opt-out kann er über seine Set-Top-Box selbst vornehmen. Bei dem bisherigen Produkt (alte Set-Top-Box – Microsoft) war die PIN-Eingabe für das Opt-out erforderlich. Bei dem neuen Produkt (neue Set-Top-Box – Huawei) ist dies nicht mehr erforderlich. Mit dem Setzen des Opt-outs werden seine Nutzungsdaten weder zu pseudonymen Nutzungsprofilen noch für die anonymen Statistiken verwendet. Das Opt-out kann der Deutschen Telekom zudem über die sonst üblichen Kommunikationskanäle mitgeteilt werden.

6.2. Direktmarketing

6.2.1. Werbekampagne

Werbekampagnen zielen darauf ab, innerhalb eines festgelegten Zeitraums Werbung in einem bestimmten Zielmarkt oder für eine Zielgruppe einzusetzen, um Verkaufszahlen zu steigern oder die Bekanntheit eines Produkts oder einer Dienstleistung zu erhöhen. Um eine solche Kampagne anhand eines Zielmarktes oder einer Zielgruppe auszurichten, sind Analysen erforderlich, die auf vorhandenen Daten basieren. Reicht der vorhandene Datenpool bei einem Werbetreibenden nicht aus, werden die Daten mit Zusatzmerkmalen angereichert. Die gewünschten Merkmale, die nicht vorhanden sind, z.B. Altersklasse und geschätzte Einkommensklasse, werden dabei oftmals von externen Quellen beschafft.

Zur Selektion einer Zielgruppe beispielsweise kann auf Basis verschiedener Merkmale wie Alter, bisherige Kaufhistorie etc. die Wahrscheinlichkeit berechnet werden, ob Personen ein bestimmtes Produkt kaufen. Ebenso ist eine Selektion von Zielgruppen nach der Zugehörigkeit zu einem bestimmten Kundensegment sowohl offline (Direktmail, Telefon- und E-Mail-Werbung) als auch online (Displaywerbung) möglich. Pseudonyme Daten spielen bei einer solchen Kampagne eine wichtige Rolle, die im Folgenden dargestellt werden soll.

6.2.1.1. Nutzung pseudonymisierter Daten für den Datenabgleich und das Matching mit externen Quellen – Erstellung einer Analysedatenbank

Um die vorhanden Daten mit externen Daten anzureichern, werden zunächst die Datenbanken von Werbetreibendem und Datenlieferanten miteinander abgeglichen und danach „gematched“. Die hierbei erforderlichen Prozessschritte gestalten sich wie folgt:

1. **Direkte Erkennungsmerkmale** von Betroffenen (z.B. Name und Adresse) werden zwischen den Datenbanken durch einen „Datenhygieneprozess“ (Standardisierung, Uniformisierung, Parsing, Duplettenabgleich etc.) korrigiert und jeweils in einer identischen Datenqualität gespeichert.
2. Die dabei verarbeiteten Namens- und Adressfelder werden mit demselben Algorithmus in beiden Datenbanken zu einer verkettbaren ID-Nummer pseudonymisiert. Der verwendete Psedonymisierungsprozess variiert von einem einfachen „Memnum“ (z.B. eine 17-stellige Nummer, bestehend jeweils aus den ersten, jedoch über das „Memnum“ verteilten, drei Buchstaben des Vornamens, des Familiennamens, der Straße und Hausnummer sowie der Postleitzahl, bis hin zu einem selbst entwickelten Pseudonymisierungsalgorithmus. Danach werden die **direkten Erkennungsmerkmale** gelöscht.
3. Die Datenbank des Werbetreibenden und die des Datenlieferanten werden anhand der ID-Nummern gegenseitig abgeglichen. Sollten dieselben Pseudonyme (ID-Nummer) gefunden worden sein („Verkettbarkeit“), können die zugehörigen Datensätze „gematched“ werden. So kann ein Unternehmen beispielsweise Zusatzmerkmale in einer Marketingdatei von einer externen Quelle speichern.

Im Ergebnis entsteht eine Analysedatenbank mit sowohl eigenen als auch angekauften Merkmalen und ID-Nummern, ohne Namen und Adressen. Mit dieser Datenbank kann dann z.B. ein Profil des Kunden erstellt werden (z.B. Altersgruppe 65+ und wohnhaft in einer Großstadt). Im Anschluss können auf Basis dieser Information gezielt Personen mit ähnlichem Profil angesprochen werden (d.h. nur Personen aus einer Verbraucherdatenbank selektieren die älter sind als 65 Jahre und in einer Großstadt wohnen).

6.2.1.2. Nutzung pseudonymisierter Daten für die Datenselektion bei interessensbasierter Werbung

1. Die Analysten des Werbetreibenden selektieren die ID-Nummern für eine Werbekampagne aus der Inhouse-Analysedatenbank. Die Selektion erfolgt durch eine einfache Selektion der Merkmale (z.B. Altersgruppe >65 Jahre und Stadt = „Frankfurt am Main“ oder „Berlin“ oder „München“ oder „Stuttgart“), aber auch durch komplexere statistische Modelle.
2. Die selektierten ID-Nummern werden mit der Referenzdatei, in der Klardaten (Name, Adresse etc.) und ID-Nummern gespeichert sind, „gematched“.
3. Die Werbung wird letztlich an den ausgewählten Kunden (z.B. über E-Mail, Direktmail, Display) ausgesendet.

6.2.1.3. Den Erfolg einer Kampagne mit pseudonymisierten Daten messen

Nach einer Werbeansprache wird in der Regel der Erfolg der Kampagne gemessen. Die Reaktionen auf die Kampagne, wie z.B. Informationsanfrage oder gar eine Produktbeschaffung, werden gespeichert, die Werbebotschaft und der Erfolg des Werbekanals bewertet. Die Reaktionen werden als Rohdaten im CRM-System des Werbetreibenden für künftige Analysezwecke gespeichert. Wie bei der Erstellung einer Analysedatenbank werden hierbei die direkten Erkennungsmerkmale entfernt, bevor die Daten der Analyseabteilung zur Verfügung gestellt werden.

6.2.2. Datenvermarktung im Auftrag über eine Agentur/einen Databroker

Wenn werbliche Daten vermarktet werden, bauen die Dateneigentümer (die für die Erhebung, Verarbeitung und Nutzung der Daten verantwortlichen Stellen) einen Kontrollmechanismus ein, damit jeder Einsatz oder Verkauf der Daten ausschließlich durch den beauftragten Auftragdatenvermarkter (Agentur/Broker) abgewickelt wird und mit Kenntnis des Dateneigentümers erfolgt. So wird auch sichergestellt, dass die fällige Lizenzgebühr korrekt bezahlt wird. Eine Datenvermarktung erfolgt in der Regel anhand der folgenden Schritte:

1. Der Dateneigentümer schließt eine Vereinbarung zur Auftragdatenverarbeitung mit der Agentur/dem Datenbroker ab.
2. Eine Datei mit Selektionsmerkmalen und Pseudonymen (ID-Nummer), aber ohne direkte Erkennungsmerkmale, wird der Agentur/dem Datenbroker zur Verfügung gestellt.
3. Die Agentur/der Datenbroker selektiert die Pseudonyme (ID-Nummer) anhand der Selektionsmerkmale für ihre/seine Kunden und schickt diese Selektion an den Dateneigentümer.
4. Der Dateneigentümer übermittelt die Daten direkt an den Kunden der Agentur/des Datenbrokers, der seine Endkunden dann schriftlich anspricht (oder er schaltet einen Lettershop für eine schriftliche Werbung beim Endkunden ein).

6.2.3. Einschalten von Display-Werbung

Werbung generiert für Onlinedienste wie Verlage, soziale Medien und eCommerce-Shops einen wichtigen Umsatz und dient für manche Bereiche als einzige Einkommensquelle, um die Dienstleistung den Internetnutzern kostenlos zur Verfügung stellen zu können. Die geschaltete Werbung sollte für den Nutzer relevant sein, um unnötige Irritation zu vermeiden. Das nachfolgende Beispiel ist eine der Varianten der interessensbasierten Displaywerbung, d.h. auf dem Bildschirm des jeweiligen Nutzers wird Werbung unmittelbar eingeblendet.

1. Ein Nutzer registriert sich auf der Webseite des Onlinedienstes oder loggt sich ein.
2. Der Onlinedienst pseudonymisiert (durch Erstellung einer User-ID) die Kontaktdaten des Nutzers (P1).
3. Der Onlinedienst platziert ein Cookie mit dem Pseudonym/der User-ID (P1) auf dem Rechner des Nutzers.
4. In einem parallelen Verfahren sendet der Onlinedienst in regelmäßigen Abständen im Rahmen einer Auftragsdatenverarbeitung das Pseudonym/die User-ID-Nummer (P1) sowie Name und Adresse des Nutzers an eine Agentur.
5. Im Auftrag des Onlinedienstes verarbeitet die Agentur die separat als Datei bereitgestellten Daten (Pseudonym/User-ID-Nummer (P1), Name, Adresse) und berechnet ein internes Pseudonym (P2) anhand des Namens und der Adresse. Danach werden alle Merkmale bis auf die Pseudonyme (P1) und (P2) gelöscht.
6. Die Agentur hat eine Datei mit Pseudonymen (P2) und (geschätzte/berechnete) Verbrauchermerkmalen. Anhand des Pseudonyms (P2) werden die bei der Agentur über den Nutzer bereits gespeicherten Verbrauchermerkmale (Schätzmerkmale und Affinitäten) angereichert. Danach wird das Pseudonym (P2) gelöscht.
7. Das Pseudonym (P1) und die Merkmale der Agentur werden zusammen mit einer Quellen-ID über eine Werbeplätzeplattform – Demand Side Plattform (DSP) bzw. Sell Side Plattform (SSP)[13] oder eine andere Stelle – im Real Time Bidding-Markt selektierbar gemacht. So kann ein Werbetreibender, z.B. ein Gartenzentrum, gezielt Personen mit einem Haus mit Garten bewerben. Die Quellen-ID identifiziert den Onlinedienstleister und sorgt dafür, dass die Datenbestände voneinander unterschieden werden können und dass die Rechnungsstellung über die verschiedenen Onlinedienste möglich ist.
8. Erkennt der DSPs/SSPs oder ein anderer Player im Real Time Bidding-Markt ein Cookie mit einem Pseudonym (P1), kann er über dieses Cookie an das Display des jeweiligen Nutzers anonym aber zielgerichtet Werbung aussteuern.

Im Übrigen wird dem Betroffenen ermöglicht, wie bei der schriftlichen Werbung, keine interessebasierte Werbung zuzulassen bzw. **Widerspruch** hiergegen einzulegen. Die Industrieselbstregulierung des EDAAs (European Digital Advertisement Alliance), in Deutschland betrieben durch den Deutschen Datenschutzrat Online-Werbung (DDOW), bietet diese Möglichkeit mithilfe eines **Icons** an.[14]

13 DSPs helfen als technischer Dienstleister Werbetreibenden den richtigen Werbeplatz für ihre Zielgruppe zu vorher festgelegten Konditionen zu finden, indem sie als zentrale Plattform den effizienten Einkauf von Werbeinventar über verschiedene Angebots-Kanäle (Ad Networks, Ad Exchanges etc.) ermöglichen. SSPs hingegen wickeln den Verkauf von Werbeplätzen ab.

14 Siehe http://www.youronlinechoices.com/de/nutzungsbasierte-online-werbung/.

6.3. Pseudonymisierung im Leitfaden zum Datenschutz in medizinischen Forschungsprojekten - TMF Generische Lösungen 2.0 -

6.3.1. Überblick: TMF-Datenschutzleitfaden

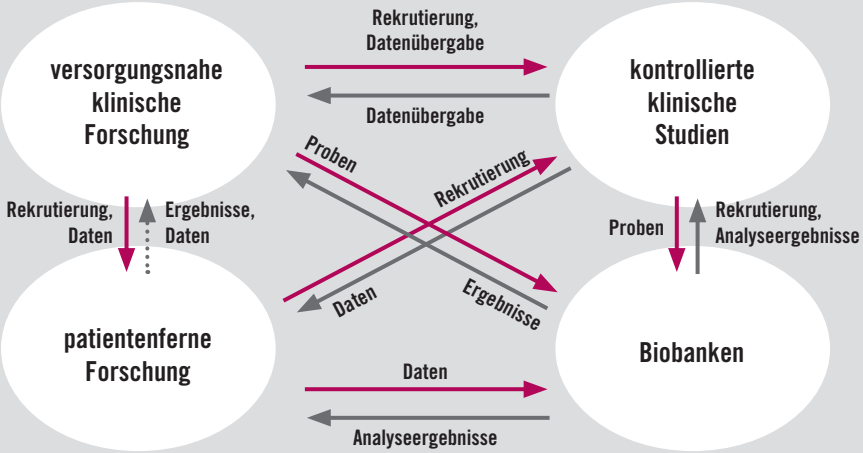
Der Leitfaden der TMF[15] zum Datenschutz in medizinischen Forschungsprojekten wurde erstmals im Jahr 2003 veröffentlicht und liegt nunmehr in der 2. Auflage (2014) vor. Die Idee zur Erstellung des Leitfadens entstand vor dem Hintergrund der Gründung deutschlandweiter Forschungsnetzwerke in den 1990er Jahren. Diese Netzwerke hatten sich mit einem fragmentierten Datenschutzrecht auf Landes- und Bundesebene, das zusätzlich vom EU-Recht überlagert wurde, auseinanderzusetzen. Das Ziel des Leitadens war somit, generische Lösungen für die Handhabung der komplexen Situation zu finden.

Im Rahmen der biomedizinischen Forschung werden grundsätzlich sensible Daten genutzt, da diese fast immer Gesundheitsdaten darstellen oder zumindest enthalten. Zum datenschutzrechtlichen Rahmen ist daher auch immer noch die ärztliche Schweigepflicht zu beachten, wenn die Daten aus dem Behandlungskontext stammen. Da die ärztliche Schweigepflicht zum Berufsrecht der Ärzte gehört und neben dem Datenschutzrecht anwendbar ist, kann der Datenschutzleitfaden der TMF nur am Rande auf diesen Rechtsrahmen hinweisen.

Der Leitfaden ist modular aufgebaut. Die einzelnen Module spiegeln typische Fallkonstellationen wider, um zu erreichen, dass diese als generische Grundlage für die Entwicklung entsprechender konkreter Datenschutzkonzepte genutzt werden können. Es besteht jedoch eine gewisse Interdependenz der Module, da die Daten häufig in verschiedenen Zusammenhängen genutzt werden bzw. von einem Kontext in einen anderen weitergereicht werden.

15 Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V.

Medizinische Forschungsbereiche: Module des TMF Datenschutzleitfadens:



Das besondere Merkmal des TMF Datenschutzleitfadens ist die Tatsache, dass die generischen Konzepte mit den Datenschutzaufsichtsbehörden in Deutschland in ausführlichen Verhandlungen abgestimmt wurden und von der Konferenz der Datenschutzbeauftragten des Bundes und der Länder seit 2003 empfohlen werden. Dies schafft für die Forschung eine gewisse Rechtssicherheit, weil die Befolgung des TMF-Leitfadens Gewähr dafür bietet, dass ein Datenschutzkonzept grundsätzlich den datenschutzrechtlichen Anforderungen genügt. Die seit Gründung der TMF innerhalb der TMF tagende „Arbeitsgruppe Datenschutz“ trägt außerdem dazu bei, die forschenden Institutionen bei der Umsetzung des Leitfadens zu unterstützen. Im Rahmen eines „Peer-review“ werden durch Vorstellung eines Datenschutzkonzeptes vor den dort anwesenden Kollegen Schwächen aufgedeckt und weitere Maßnahmen vorgeschlagen sowie schließlich per Votum der Mitglieder die Konformität des vorgestellten – und gegebenenfalls geänderten – Konzeptes festgestellt. Dieses Votum wird dann in aller Regel von der zuständigen Datenschutzaufsichtsbehörde als ausreichend angesehen.

6.3.2. Pseudonymisierung als technisch-organisatorische Maßnahme im Sinne der informationellen Gewaltenteilung

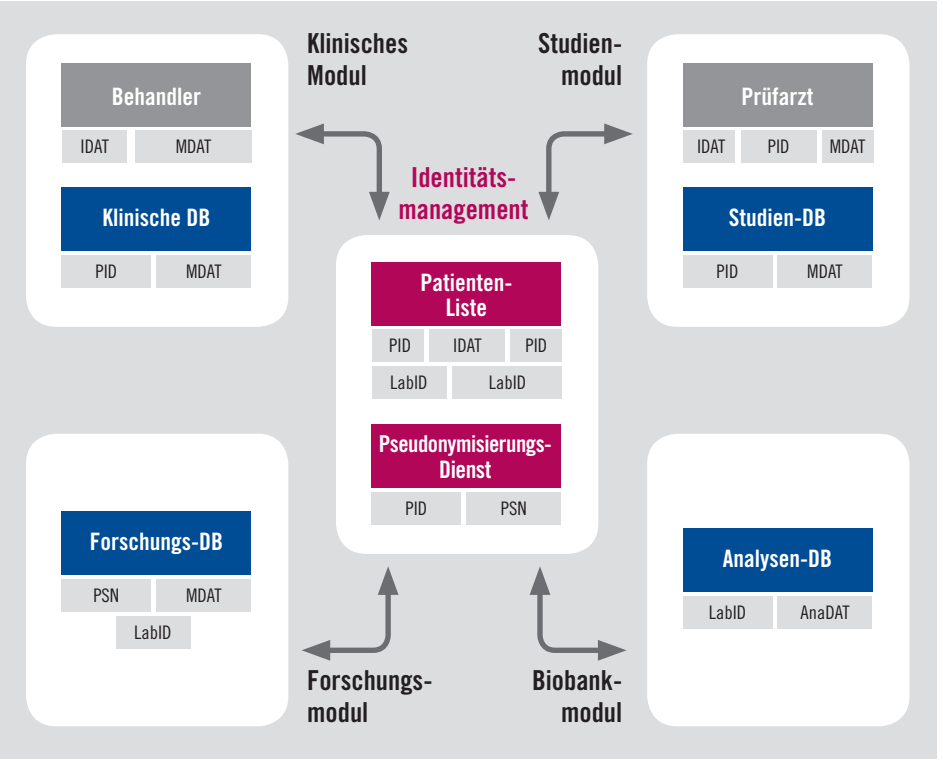
Die **informationelle Gewaltenteilung** ist ein Grundelement der generischen Datenschutzkonzepte der TMF. Die Pseudonymisierung ist das wesentliche Mittel, um die Trennung von identifizierenden Daten einerseits und Forschungsdaten andererseits zu erreichen. Sie dient als **technisch-organisatorische Maßnahme** zum Schutz der Probanden.

Die Datenverarbeitung für die Forschung erfolgt in aller Regel einwilligungsbasiert, d.h. die Probanden wurden vor der Nutzung, meist bereits bei der Erhebung der Daten, über die Nutzung für die Forschung aufgeklärt und über die Risiken der Nutzung belehrt (sog. informed consent). Eine Nutzung ist daher auch vor dem Hintergrund möglich, dass die Daten personenbezogen bleiben, eine Anonymisierung ist nur im Hinblick auf den Grundsatz der Datensparsamkeit nötig, nicht aber, um die Daten überhaupt erst nutzbar zu machen.

Allerdings hat sich der Pseudonymisierungsbegriff unter der DS-GVO[16] gegenüber dem BDSG[17] geändert. Was zunächst unscheinbar aussieht, hat doch erhebliche Konsequenzen. Während unter dem BDSG die Erschwerung der Re-Identifizierung genügte, verlangt die DS-GVO, dass eine Re-Identifizierung nicht mehr möglich ist. Letzteres entspricht dem Begriff der Anonymität. Das Erreichen von

de facto Anonymität für pseudonyme Datensätze erhöht die Anforderungen erheblich. Sie wird im Rahmen der biomedizinischen Forschung in aller Regel nicht zu erreichen sein, da es sich sehr oft um hochdimensionale Datensätze handelt, immer öfter auch um genomische Informationen, so dass eine De-Identifizierung bis zu dem Grad der Anonymität die Daten für die Forschung praktisch wertlos machen würde. Zudem ist die Anonymität von Daten ohnehin keine langfristig sichere Eigenschaft, sondern hängt u.a. ab von unbekanntem externen – eventuell erst künftig verfügbarem – Zusatzwissen. Daher sind Forschungsdaten meist nicht de facto anonym, sondern gröber de-identifiziert, um sie für den entsprechenden Forschungszweck nutzbar zu halten. Daten, aus denen die direkten Identifizierer entfernt wurden, enthalten meist noch viele Detailangaben, die zumindest in Summe eine Re-Identifizierung nicht ausschließen, so dass diese lediglich erschwert, aber de facto nicht unmöglich ist. Die Pseudonymisierung ist allerdings auch nur eine mögliche Maßnahme zum Schutz der Probanden, andere Maßnahmen sind denkbar und vor dem geschilderten Hintergrund auch sinnvoll. Es stellt sich hier die Frage, ob ein neuer Begriff gefunden werden muss, für Daten, die zwar „kodiert", aber nicht pseudonymisiert im Sinne der DS-GVO sind.

Das folgende Modell beruht noch auf dem Pseudonymisierungsbegriff des BDSG. Es ist grundsätzlich auch unter der DS-GVO anwendbar, müsste aber eventuell begrifflich angepasst werden, weil die verbleibenden Datensätze typischer Weise nicht den Anforderungen des Pseudonymisierungsbegriffs unter der DS-GVO entsprechen.



PID Patientenidentifikator (Pseudonym in der Behandlungsdatenbank) · **MDAT** Medizinische Daten · **IDAT** Identifizierende Daten eines Patienten · **LabID** Labordaten/Probennummer · **PSN** Pseudonym im Weiteren Forschungskontext · **AnaDAT** Analysedaten

16 Art. 4 Abs. 5 EU-DS-GVO „Pseudonymisierung" (ist) die Verarbeitung personenbezogener Daten in einer Weise, dass die personenbezogenen Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können, sofern diese zusätzlichen Informationen gesondert aufbewahrt werden und technischen und organisatorischen Maßnahmen unterliegen, die gewährleisten, dass die personenbezogenen Daten nicht einer identifizierten oder identifizierbaren natürlichen Person zugewiesen werden.

17 § 3 Abs. 6a BDSG (6a) Pseudonymisieren ist das Ersetzen des Namens und anderer Identifikationsmerkmale durch ein Kennzeichen zu dem Zweck, die Bestimmung des Betroffenen auszuschließen oder wesentlich zu erschweren.

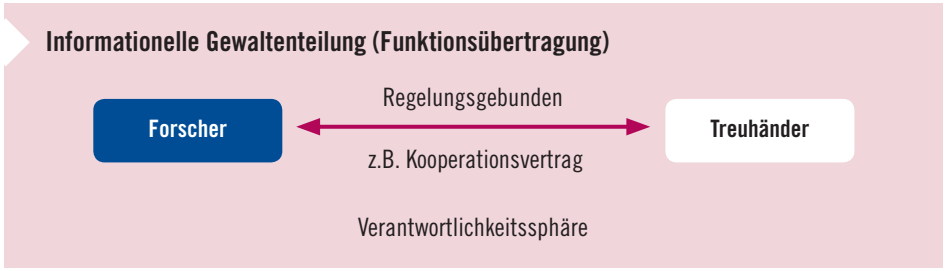
Ein Pseudonym muss grundsätzlich „nicht sprechend“ sein, d.h. es darf keine Bestandteile enthalten, die ihrerseits auf identifizierende Daten hinweisen, z.B. Geburtsdaten, Initialen etc. Die Zuordnung darf ausschließlich über eine Zuordnungsliste oder eine Zuordnungsregel erfolgen, die gut zu schützen sind.

6.3.3. Einbindung eines Treuhänders

Grundsätzlich ist die Einschaltung eines Treuhänders empfehlenswert, um dem Merkmal „gesondert aufbewahrt werden“ gem. Art. 4 Nr. 5 DS-GVO gerecht zu werden. Im TMF-Leitfaden wurde dies bereits unter bisherigem Datenschutzrecht für sinnvoll erachtet. Eine Treuhandstelle (TTP), die Patientenidentitätsdaten (IDAT) und Pseudonyme (PID) hält, muss bestimmte Anforderungen erfüllen, d.h. sie muss:

>> rechtlich unabhängig sein (eigene dauerhafte Rechtsform);
>> räumlich und personell klar getrennt sein;
>> vertraglich verpflichtet werden, das Datenschutzkonzept umzusetzen;
>> im Übrigen weisungsunabhängig sein.

Eine Auftragsdatenverwaltung des Treuhänders für die Forschungseinrichtung genügt diesen Anforderungen nicht.

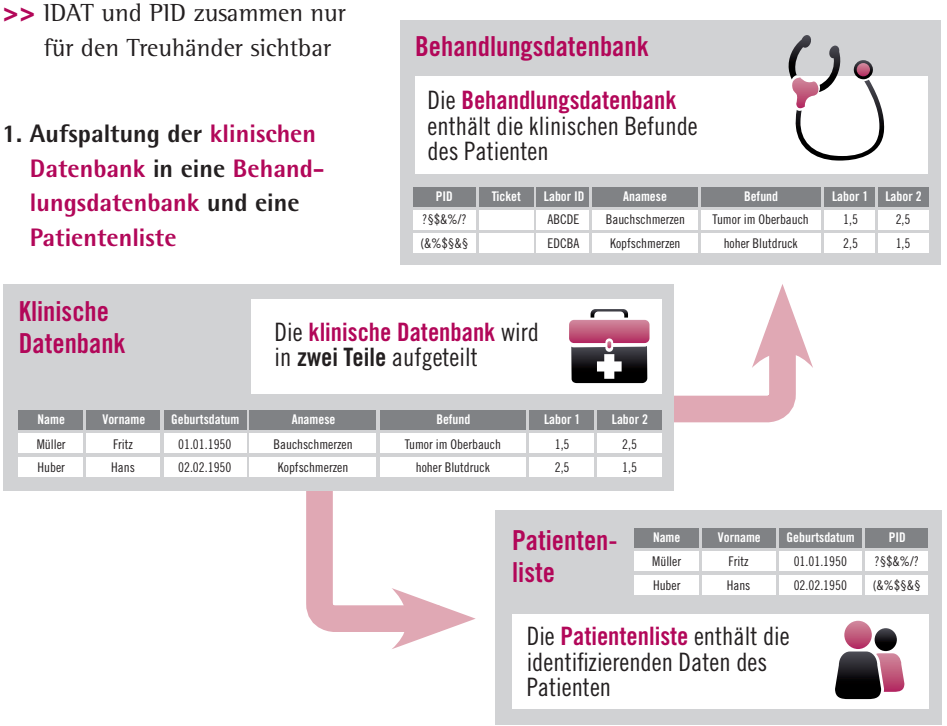


Beispiel: Wesentliche Elemente des klinischen Moduls

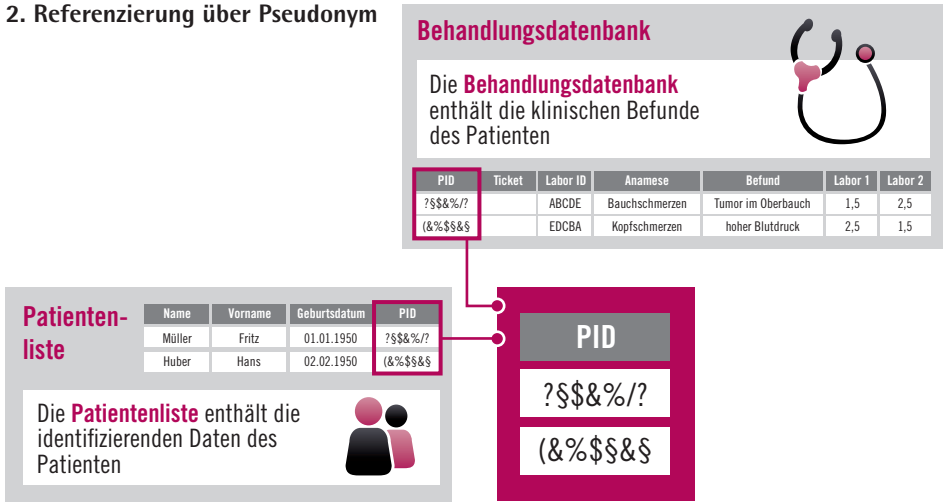
>> Zentraler Datenpool (z.B. EHR repository oder Datawarehouse)
>> Online-Zugang zu allen Daten für behandelndes Personal (IDAT & MDAT)
>> Speicherung der MDAT unter Pseudonym (PID)
>> Pseudonym (PID) nur dem Datawarehouse Staff und dem Treuhänder bekannt
>> MDAT and IDAT zusammen sichtbar nur an client workstations (im Behandlungskontext) nicht auf dem application server!
>> IDAT und PID zusammen nur für den Treuhänder sichtbar
>> Zugang im Behandlungskontext kontrolliert über 1x token
>> Kein “Public Use” des Datenpools (kein Zugang und keine Suchfunktion von außen),
>> Externe Forschung nur mit exportierten Daten

Der Prozess der Überführung einer klinischen Datenbank in eine Behandlungsdatenbank, die sowohl für die weitere Behandlung des Patienten dem medizinischen Personal zur Verfügung stehen als auch Daten für Forschungszwecke vorhalten soll, vollzieht sich in folgenden Schritten:

1. Aufspaltung der klinischen Datenbank in eine Behandlungsdatenbank und eine Patientenliste



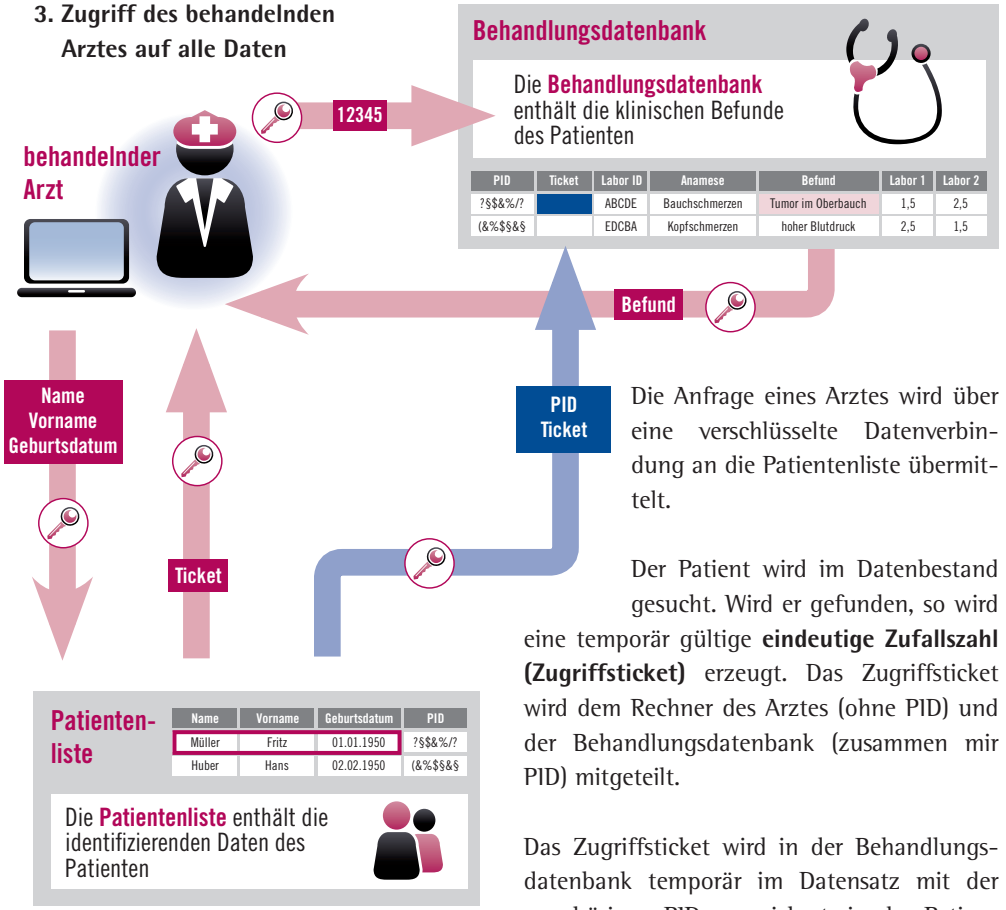
2. Referenzierung über Pseudonym



Die **Patientenliste** und die **Behandlungsdatenbank** referenzieren über einen geheimen, gemeinsamen, zufälligen Indentifikator **(PID)** aufeinander.

Die **Patientenliste** und die **Behandlungsdatenbank** sind logisch und **räumlich getrennt**. Sie unterliegen einer **ubabhängigen Administration**.

3. Zugriff des behandelnden Arztes auf alle Daten



Die Anfrage eines Arztes wird über eine verschlüsselte Datenverbindung an die Patientenliste übermittelt.

Der Patient wird im Datenbestand gesucht. Wird er gefunden, so wird eine temporär gültige **eindeutige Zufallszahl (Zugriffsticket)** erzeugt. Das Zugriffsticket wird dem Rechner des Arztes (ohne PID) und der Behandlungsdatenbank (zusammen mir PID) mitgeteilt.

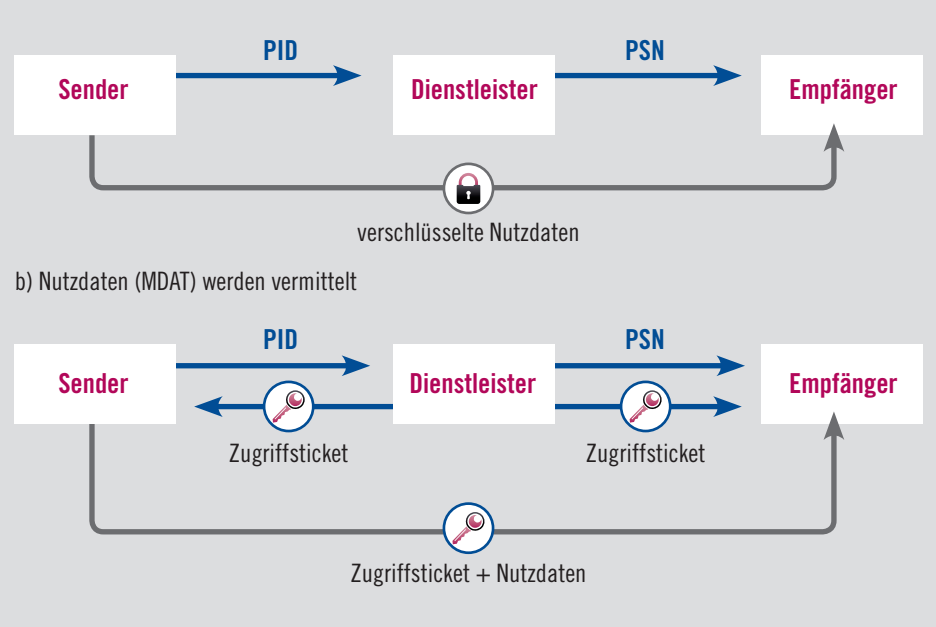
Das Zugriffsticket wird in der Behandlungsdatenbank temporär im Datensatz mit der zugehörigen PID gespeichert, in der Patientendatenbank nicht.

Mit dem Zugriffsticket kann der behandelnde Arzt die von ihm gewünschte Information in der Behandlungsdatenbank abfragen.

Das Zugriffsticket wird in der Behandlungsdatenbank nach erfolgter Abfrage (bzw. nach Timeout) gelöscht.

6.3.4. Umpseudonymisierung durch Pseudonymisierungsdienst (Treuhänder) beim Export der Daten in das Forschungsmodul

Zweck des Pseudonymisierungsdienstes ist der besondere Schutz der Daten in einer auf Langzeitspeicherung angelegten Forschungsdatenbank. Mittel dazu ist die Transformation des PID aus der Patientenliste in ein Pseudonym PSN, das in der Forschungsdatenbank als Kennung genutzt wird; hierfür wird ein **kryptographisches Verfahren** angewendet. Da der Pseudonymisierungsdienst die medizinischen Daten (MDAT) weder benötigt noch überhaupt sehen soll, werden diese in **asymmetrischer Verschlüsselung** durchgereicht:



Die Pseudonymisierung ist eine reine Maschinenfunktion, die keines Eingriffs durch das Personal bedarf. Die Daten werden nur von zugelassenen Absendern übernommen.

Ihre Notizen

Ihre Notizen