Praxishilfe zur Anonymisierung/Pseudonymisierung

Erarbeitet von Mitgliedern aus den nachfolgend genannten Verbänden

Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e. V. (GMDS)
Arbeitsgruppe „Datenschutz und IT-Sicherheit im Gesundheitswesen“

Berufsverband der Datenschutzbeauftragten Deutschlands (BvD) e. V.

Bundesverband Gesundheits-IT e. V.
Arbeitsgruppe „Datenschutz & IT-Sicherheit“

Version 2

Stand: 27. Januar 2024

Autoren (alphabetisch)

Tanja Albert	Daskeo Albert & Beyer GbR
Andrea Backer-Heuveldop	ds² Unternehmensberatung GmbH & Co. KG
Jürgen Bühse	PERGON Unternehmensberatung e. K.
Jamie Crookes	Compliant Digital GmbH & Co. KG
Joaquín M. González	Kassenärztliche Vereinigung Baden-Württemberg
Carla Haase	PRO Klinik Holding
Christoph Isele	Cerner Health Services Deutschland GmbH
David Koeppe	Vivantes - Netzwerk für Gesundheit GmbH
Michael Letter	5medical management GmbH
Georg Möller	SK-Consulting Group GmbH
Regina Mühlich	AdOrga Solutions GmbH
Bernd Schütze	Deutsche Telekom Healthcare and Security GmbH
David Seiler	Rechtsanwalt
Ulrike-Alexandra Seitzinger	Seitzinger Legal•HR•Privacy

Version 1

Stand: 29. Juni 2018

Autoren (alphabetisch)

Sonja Holst	Charité - Universitätsmedizin Berlin
Bernd Schütze	Deutsche Telekom Healthcare and Security GmbH
Gerald Spyra	Ratajczak und Partner mbB Rechtsanwälte

Geschlechtergerechte Sprache

Hinweis bzgl. geschlechtsneutraler Formulierung im gesamten Text:

- Eine gleichstellungsgerechte Gesellschaft erfordert eine geschlechterneutrale Sprache. Geschlechterneutrale Sprache muss im deutschen Umfeld drei Geschlechtern gerecht werden: Divers, Frauen und Männern.
- Im folgenden Text werden, soweit möglich und sinnvoll, entsprechende Formulierungen genutzt (z. B. Paarformeln, Ableitungen). Personenbezeichnungen, bei denen es sich um juristische Fachbegriffe handelt, die sowohl natürliche als auch juristische Personen bezeichnen können, werden im folgenden Text nicht durch Paarformeln ersetzt. Dies gilt auch für technische Fachbegriffe, Definitionen und Zitate aus Normen (z. B. DIN EN ISO) und gesetzlichen Vorschriften. Entsprechende Begriffe sind im Sinne der Gleichbehandlung geschlechtsneutral zu interpretieren.
- Wo aus Gründen der leichteren Lesbarkeit bei personenbezogenen Substantiven und Pronomen nur ein Geschlecht dargestellt wurde, impliziert dies jedoch keine Benachteiligung der anderen beiden Geschlechter, sondern soll im Sinne der sprachlichen Vereinfachung als geschlechtsneutral verstanden werden.

Haftungsausschluss

Das vorliegende Werk ist nach bestem Wissen erstellt, der Inhalt wurde von den Autoren mit größter Sorgfalt zusammengestellt. Dennoch ist diese Ausarbeitung nur als Standpunkt der Autoren aufzufassen, eine Haftung für die Angaben übernehmen die Autoren nicht. Die in diesem Werk gegebenen Hinweise dürfen daher nicht direkt übernommen werden, sondern müssen vom Leser für die jeweilige Situation anhand der geltenden Vorschriften geprüft und angepasst werden.

Die Autoren sind bestrebt, in allen Publikationen die Urheberrechte der verwendeten Texte zu beachten, von ihnen selbst erstellte Texte zu nutzen oder auf lizenzfreie Texte zurückzugreifen.

Alle innerhalb dieses Dokumentes genannten und ggf. durch Dritte geschützten Marken- und Warenzeichen unterliegen uneingeschränkt den Bestimmungen des jeweils gültigen Kennzeichenrechts und den Besitzrechten der jeweiligen eingetragenen Eigentümer. Allein aufgrund der bloßen Nennung ist nicht der Schluss zu ziehen, dass Markenzeichen nicht durch Rechte Dritter geschützt sind!

Copyright

Für in diesem Dokument veröffentlichte, von den Autoren selbst erstellte Objekte gilt hinsichtlich des Copyrights die folgende Regelung:

Inhaltsverzeichnis

Tabellenverzeichnis

Zusammenfassung

Personenbezogene Daten sind „alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person beziehen“. Dieser Begriff ist somit sehr weit gefasst. Er umfasst alle Daten, die einer individuellen natürlichen Person direkt oder indirekt zugeordnet werden können.

Wie die früheren nationalen Datenschutzregelungen enthält auch die EU-Datenschutz-Grundverordnung (DS-GVO) ein grundsätzliches Verbot der Verarbeitung personenbezogener Daten. Daher gilt nach wie vor, dass jede Verarbeitung personenbezogener Daten verboten ist, es sei denn, es gibt einen gesetzlich geregelten Erlaubnistatbestand (vgl. Art. 6 DS-GVO „Rechtmäßigkeit der Verarbeitung“). Dabei ist insbesondere zu beachten, dass je sensibler die zu verarbeitenden Daten sind, desto notwendiger ist die Gewährleistung eines angemessen hohen Schutzniveaus für diese Daten. Gewährleisten muss den Schutz personenbezogener Daten für die gesamte Dauer der Verarbeitung, also über den gesamten Lebenszyklus der Daten hinweg, der „Verantwortliche“. Art. 4 Ziff. 7 DS-GVO definiert einen Verantwortlichen als „die natürliche oder juristische Person, Behörde, Einrichtung oder andere Stelle, die allein oder gemeinsam mit anderen über die Zwecke und Mittel der Verarbeitung von personenbezogenen Daten entscheidet“. Folglich ist derjenige, welcher über die Mittel und Zwecke der Verarbeitung entscheidet, vollumfänglich für alles verantwortlich, was mit diesen Daten geschieht.

Die DS-GVO sieht die Pseudonymisierung als eine mögliche Maßnahme an, deren Einsatz zur Gewährleistung eines angemessenen Schutzniveaus beitragen kann. Die DS-GVO verweist an verschiedenen Stellen auf die Pseudonymisierung bzw. Anonymisierung als Schutzmaßnahme, wie z. B. bei den Anforderungen zu Privacy by Design/Default (Art. 25 DS-GVO) oder bei der Verarbeitung von personenbezogenen Daten zu Forschungszwecken, auf diese Maßnahme.

Der Begriff der Anonymisierung wird in der DS-GVO nicht explizit definiert. Dies bedeutet jedoch nicht, dass eine Anonymisierung mit Geltung der DS-GVO per se nicht mehr möglich ist. Vielmehr adressieren sowohl die DS-GVO als auch etwaige nationale Regelungen an mehreren Stellen die Möglichkeit einer Anonymisierung. Aus den Regelungen der DS-GVO bzgl. der Begrifflichkeiten „personenbezogene Daten“ und „Pseudonymisierung“ sowie der Definition der Anonymisierung in Richtlinie (EU) 2019/1024 ergeben sich Anforderungen, welche an eine Anonymisierung bzw. Pseudonymisierung zu stellen sind.

Sowohl die Pseudonymisierung als auch die Anonymisierung können bzw. sollten daher angewendet werden, wenn dies im Rahmen der Verarbeitung sinnvoll erscheint. Einerseits um die Risiken der Verarbeitung personenbezogener Daten für von dieser Verarbeitung betroffene Personen zu verringern, andererseits um den aus der DS-GVO resultierenden rechtlichen Anforderungen hinsichtlich eines angemessen hohen Schutzniveaus zu genügen.

Dabei ist zu beachten, dass sowohl die Anonymisierung als auch die Pseudonymisierung eine Verarbeitung im Sinne der DS-GVO darstellen. Dies bedeutet, dass für die Durchführung einer Anonymisierung oder einer Pseudonymisierung eine Rechtsgrundlage („Erlaubnistatbestand“) erforderlich ist. Dies ist nur eine der datenschutzrechtlichen Rahmenbedingungen, welche in der DS-GVO zu finden sind.

In dieser Ausarbeitung werden verschiedene Rahmenbedingungen besprochen, die bei einer Pseudonymisierung oder Anonymisierung zu beachten sind.

Vorwort zur zweiten Fassung

Die erste Version der „Praxishilfe zur Pseudonymisierung/Anonymisierung“ erschien im Juni 2018. Unmittelbar danach, ebenfalls im Juni 2019, wurde die Richtlinie (EU) 2019/1024 mit der darin enthaltenen europäischen Definition der Anonymisierung im Amtsblatt der Europäischen Union veröffentlicht. Da einerseits die Zielrichtung dieser Richtlinie (EU) 2019/1024 der öffentliche Sektor war, andererseits die in der Richtlinie enthaltene Definition weitestgehend den in der Praxishilfe enthaltenen Darstellungen entspricht, verzichteten wir 2019 auf eine Anpassung der gerade frisch veröffentlichten Praxishilfe.

Seit Ende 2022 wird insbesondere in Deutschland von verschiedenen Akteuren eine Definition des Begriffs „Anonymisierung“ gefordert. Aufgrund der Existenz einer europäischen Definition ist eine rechtskonforme Definition, d. h. eine „deutsche“ Definition, die weder den Regelungen der DS-GVO zu „personenbezogenen Daten“ widerspricht noch im Widerspruch zur Begriffsbestimmung in der Richtlinie (EU) 2019/1024 steht, nur schwer vorstellbar. Vermutlich ist den meisten, welche eine deutsche Definition fordern, die in der Richtlinie (EU) 2019/1024 enthaltene Definition unbekannt. Ein Grund, in dieser überarbeiteten Praxishilfe ausdrücklich auf diese bereits vorhandene Definition des europäischen Gesetzgebers hinzuweisen, um so den Wunsch nach einer „deutschen“ Begriffsbestimmung mit der Existenz der europäischen Definition zu beantworten.

Immer wieder tauchen auch Fragen nach dem Nachweis der Anonymität von Daten auf. Dieser Nachweis kann in der Tat nicht immer leicht erbracht werden. Zwar sind personenbezogene und pseudonyme Daten in Art. 4 Ziff. 1 und 5 DS-GVO legaldefiniert, jedoch erschweren die Offenheit der Begriffe und fehlende Konkretisierungen sowohl im Verordnungstext wie auch in den Erwägungsgründen der DS-GVO eine verlässliche Auslegung, insbesondere auch in Bezug auf anonyme Daten. Die Kriterien für die Feststellung oder das Fehlen des Personenbezugs sind somit nicht eindeutig festgelegt, was wiederum die Nachweismöglichkeit des Vorhandenseins oder Fehlens eines Personenbezugs erschwert.

Mit dem Thema „Künstliche Intelligenz“ wird die Nachfrage nach Daten immer größer. Inwieweit hier anonymisierte Daten tatsächlich sinnvoll für das Training von KI-Modellen im medizinischen Kontext der Unterstützung bei individuellen Behandlungsfragen genutzt werden können, sei dahingestellt. Aber der Bedarf an bereitgestellte Daten und damit auch die laute Forderung nach Daten übt Druck auf politische Akteure aus, die sich diesem Druck oftmals beugen.

Insgesamt gesehen ergibt sich aus den Entwicklungen der letzten Jahre eine Nachfrage an eine Überarbeitung der Praxishilfe, welcher wir hiermit nachkommen.

1 Einleitung

Sowohl die Pseudonymisierung als auch die Anonymisierung wird verwendet, um Risiken der Verarbeitung für die von der Verarbeitung der personenbezogenen Daten betroffenen Personen zu minimieren. D. h. die Methoden stellen Maßnahmen dar, welche dem Schutz von personenbezogenen Daten dienen. In Bezug auf die Pseudonymisierung schrieb der europäische Gesetzgeber in ErwGr. 28 DS-GVO: „Die Anwendung der Pseudonymisierung auf personenbezogene Daten kann […] die Verantwortlichen und die Auftragsverarbeiter bei der Einhaltung ihrer Datenschutzpflichten unterstützen".

Art. 4 Ziff. 7 DS-GVO definiert „Verantwortlicher" als „die natürliche oder juristische Person, Behörde, Einrichtung oder andere Stelle, die allein oder gemeinsam mit anderen über die Zwecke und Mittel der Verarbeitung von personenbezogenen Daten entscheidet". Wie aus dieser Formulierung ersichtlich wird, kann es einen oder auch mehrere Verantwortliche geben. Entsprechend wird in Art. 26 DS-GVO erklärt, unter welchen Umständen eine Verarbeitung durch gemeinsam Verantwortliche erfolgen kann. Dies gilt selbstverständlich auch für die Pseudonymisierung oder auch die Anonymisierung, die jeweils eine Verarbeitung darstellen. Auch diese Verarbeitungen können von einem oder auch von mehreren gemeinsamen Verantwortlichen durchgeführt und genutzt werden.

Eine Pseudonymisierung ersetzt nicht zwangsläufig andere Datenschutzmaßnahmen, sondern ist eher als begleitende Maßnahme zu verstehen (ErwGr. 28 S. 2 DS-GVO). Die DS-GVO nennt die Pseudonymisierung als begleitende Maßnahme an verschiedenen Stellen wie z. B.:

- Art. 6 Abs. 4 DS-GVO, um bei Zweckänderung geeignete Garantien für die Sicherheit abzubilden
- Art. 25 Abs. 1 DS-GVO, als eines der Mittel, um „Datenschutz durch Technikgestaltung und durch datenschutzfreundliche Voreinstellungen" („Privacy by Design/Privacy by Default") umzusetzen
- Art. 32 Abs. 1 lit. a DS-GVO, als eine der zu berücksichtigenden Anforderungen bei der Gewährleistung eines angemessenen Schutzniveaus
- Art. 89 Abs. 1 DS-GVO, als eine mögliche Maßnahme, um Rechte und Freiheiten betroffener Person bei der Verarbeitung personenbezogener Daten von im öffentlichen Interesse liegenden Archivzwecken, zu wissenschaftlichen oder historischen Forschungszwecken oder zu statistischen Zwecken zu gewährleisten.

Anonyme oder pseudonyme Daten können z. B. genutzt werden für:

- Schulungszwecke
- Krankheitsregister
- Klinische Studien
- Statistische Auswertungen/Analysen.

Die Vorgaben der DS-GVO enthalten die Rahmenbedingungen, unter welchen eine Pseudonymisierung anwendbar ist. Bzgl. der Ausgestaltung der Pseudonymisierung gibt es aber Bedarf an Auslegung. Art. 40 Abs. 2 lit. d DS-GVO verweist darauf, dass Verhaltensregeln die Pseudonymisierung personenbezogener Daten konkreter ausgestalten können. Diese Praxishilfe stellt keine Verhaltensregeln i. S. d. DS-GVO dar, will aber einen Beitrag dazu leisten, wie mit Pseudonymisierung und Anonymisierung unter Geltung der DS-GVO umzugehen ist.

2 Abgrenzung/Klarstellung

Die vorliegende Praxishilfe bezieht sich auf Daten in der Gesundheitsversorgung, d. h. besondere Kategorien von Daten im Sinne von Art. 9 Abs. 1 DS-GVO. Grundsätzlich ist eine Pseudonymisierung oder Anonymisierung natürlich auch bei anderen Daten sinnvoll. Die vorliegend dargestellten Ausführungen und Methoden sind daher i. d. R. auch auf diese (vielfach weniger sensiblen) Daten übertragbar.

Diese Praxishilfe stellt keine Verhaltensregel i.S.v. Art. 40 Abs. 2 lit. d DS-GVO dar[1]. Sie beschreibt allerdings, welche Rahmenbedingungen beim Vorgehen bzgl. Anonymisierung oder Pseudonymisierung aus Sicht der DS-GVO mindestens beachtet werden sollten. Weiterhin werden einige Methoden zu diesem Thema vorgestellt, ohne dass von den Verfassern ein Anspruch auf Vollständigkeit bzgl. der Darstellung erhoben wird.

[1] Vgl. hierzu Schwartmann R, Weiß, S. (Hrsg.) Entwurf für einen Code of Conduct zum Einsatz DS-GVO konformer Pseudonymisierung. (Stand Oktober 2019. Online, zitiert am 2024-01-25; verfügbar unter https://www.de.digital/DIGITAL/Redaktion/DE/Digital-Gipfel/Download/2019/p9-code-of-conduct.pdf?blob=publicationFile&v=1

3 Allgemeines

Im Rahmen der folgenden Darstellung werden des Öfteren Beispiele zur Veranschaulichung genutzt. Alle Beispiele basieren auf dem folgenden onkologischen Beispieldatensatz:

Vorname	Nachname	Geschlecht	Geb.-Datum	PLZ	ICD-2019[2]	Diagnose
Heike	Richter	W	11.05.1983	10115	C43.9	Bösartiges Melanom der Haut, nicht näher bezeichnet
Jan	Schröder	M	03.12.1965	10115	D22.9	Melanozytennävus, nicht näher bezeichnet
Hugo-Egon	Meyer	M	27.08.1977	10178	C85.9	Non-Hodgkin-Lymphom, Typ nicht näher bezeichnet
Eckbert	Schneider	M	23.12.1981	10247	D44.8	Neubildung unsicheren oder unbekannten Verhaltens: Beteiligung mehrerer endokriner Drüsen
Jürgen	Stillstand	M	29.11.1985	10319	C18.4	Bösartige Neubildung: Colon transversum
Hiltrud	Niemand	W	15.07.1987	10407	D46.1	Refraktäre Anämie mit Ringsideroblasten
Uwe	Müller	M	31.03.1988	10435	C16.3	Bösartige Neubildung: Antrum pyloricum
Michael	Matuschek	M	13.04.1968	10439	D46.2	Refraktäre Anämie mit Blastenüberschuss
Anke	Schmidt	W	01.04.1978	10585	C50.3	Bösartige Neubildung unterer innerer Quadrant Brustdrüse
Kunigunde	Gewaltig	W	21.01.1969	10707	C91.10	Chronische lymphatische Leukämie: Ohne Angabe einer kompletten Remission
Franz	Herrlich	M	17.11.1967	10717	D12.8	Gutartige Neubildung: Rektum
Berthold	Koch	M	28.08.1991	10717	D12.6	Gutartige Neubildung: Kolon, nicht näher bezeichnet
Frieda	Fischer	W	15.11.1987	10787	C50.8	Bösartige Neubildung: Brustdrüse, mehrere Teilbereiche überlappend
Gerfriede	Jensen	W	23.07.1983	10827	C50.1	Bösartige Neubildung: Zentraler Drüsenkörper der Brustdrüse
Käthe	Albers	W	27.05.1975	10963	C83.0	Non-Hodgkin-Lymphom: Kleinzellig (diffus)

Tabelle 1: Beispieldaten mit onkologischen Erkrankungen

[2] In den folgenden Tabellen wurde auf die Angabe der Jahreszahl aus Platzgründen verzichtet.

4 Begriffsbestimmungen

Die nachfolgend dargestellten Begriffe sind in europäischen Gesetzesakten legaldefiniert, d. h. die Auslegung dieser Begriffe erfolgt autonom nach dem europäischen Recht unter Berücksichtigung der Systematik und der Zielsetzung der Rechtsakte[3]. Die Interpretation europäischer Rechtsvorgaben aus rein deutscher Sichtweise führte in der Vergangenheit oft zu Überraschungen bei diesen Rechtsanwendern, wenn der EuGH über Rechtsfragen dann aus europarechtlicher Sicht entschied. Es ist daher dringend zu empfehlen, die Vorgabe europäische Rechtsakte aus europäischer Sicht zu interpretieren und diese nicht ausschließlich aus deutscher Sicht zu bewerten.

Auch die Auslegung europäischer Gesetzestexte erfolgt entsprechend der juristischen Methodenlehre, wobei natürlich die Auslegung der europäischen und nicht der deutschen Sichtweise die Auslegung bestimmen muss. Der EuGH[4] führt hierzu aus: „Bei der Prüfung dieser Bestimmungen sind nicht nur ihr Wortlaut, sondern auch ihr Zusammenhang und die Ziele zu berücksichtigen, die mit der Regelung, zu der sie gehören, verfolgt werden“. Wie der EuGH darstellt, sind somit insbesondere die Ziele des jeweiligen Rechtsaktes zu berücksichtigen. Im Falle der DS-GVO werden die Ziele in Art. 1 DS-GVO festgelegt, welche entsprechend Art. 1 Abs. 2 DS-GVO insbesondere auch den Schutz der Grundrechte und Grundfreiheiten der betroffenen Personen betreffen; aus europäischer Sicht sind diese Grundrechte und Grundfreiheiten in der Charta der Europäischen Union[5] zu finden. In diesem Zusammenhang sei darauf hingewiesen, dass Datenschutz in Art. 8 in der Charta der Grundrechte im Gegensatz zum deutschen Grundgesetz eigenständig geregelt ist und somit kein abgeleitetes Recht darstellt.

Allerdings wird die Möglichkeit der Auslegung durch den Wortlaut des Gesetzestextes beschränkt; keine Auslegung kann die Bedeutung des Wortlauts einer Regelung ändern, wenn der Wortlaut eindeutig ist. Die Erwägungsgründe eines europäischen Rechtsaktes sind hinsichtlich der Interpretation bei einem nicht eindeutig interpretierbaren Wortlaut vorrangig zu berücksichtigen, sind sie doch Bestandteil der jeweiligen europäischer Rechtsakte[6]. Der EuGH weist darauf hin, dass „Begründungserwägungen eines Gemeinschaftsrechtsakts rechtlich nicht verbindlich sind und weder herangezogen werden können, um von den Bestimmungen des betreffenden Rechtsakts

[3] EuGH, Urt. v. 25. Oktober 2011, Az. C-509/09, C-161/10, Rn. 38: „[…] dass die Bestimmungen der Verordnung nach ständiger Rechtsprechung autonom und unter Berücksichtigung ihrer Systematik und ihrer Zielsetzungen auszulegen sind“. Online, zitiert am 2023-11-01; verfügbar unter https://dejure.org/2011,48 bzw. Volltext des Urteils unter https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A62009CJ0509

[4] EuGH, Urt. v. 25. Oktober 2011, Az. C-509/09, C-161/10, Rn. 54. Online, zitiert am 2023-11-01; verfügbar unter https://dejure.org/2011,48 bzw. Volltext des Urteils unter https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A62009CJ0509

[5] Charta der Grundrechte der Europäischen Union. Online, zitiert am 2023-11-01; verfügbar unter https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX:12012P/TXT

[6] Siehe auch „Gemeinsamer Leitfaden des Europäischen Parlaments, des Rates und der Kommission für Personen, die an der Abfassung von Rechtstexten der Europäischen Union mitwirken“, Abschnitt 10.5: „Die Erwägungsgründe müssen in möglichst knapper Form die Gründe für die wesentlichen Vorschriften des verfügenden Teils des Rechtsakts angeben.“ Online, zitiert am 2023-11-01; verfügbar unter https://op.europa.eu/de/publication-detail/-/publication/3879747d-7a3c-411b-a3a0-55c14e2ba732/language-de

abzuweichen, noch, um diese Bestimmungen in einem Sinne auszulegen, der ihrem Wortlaut offensichtlich widerspricht“[7].

4.1 Personenbezogene Daten

Entsprechend Art. 4 Ziff. 1 DS-GVO sind personenbezogene Daten „alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person beziehen“. Die Begriffe „identifiziert“ und „identifizierbar“ werden dabei als gleichberechtigte Alternativen verwendet. Damit fallen nicht nur Informationen darunter, welche direkt eine Person identifizieren, sondern auch alle Informationen, welche über Zwischenschritte eine Person identifizieren („identifizierbar machen“). Eine Identifikation einer natürlichen Person liegt somit bereits dann vor, wenn diese Person hinreichend individualisiert werden kann, da sie sich aufgrund der vorhandenen Informationen ausreichend von anderen Personen in der vorhandenen Datenmenge unterscheidet[8]. D. h. soweit und solange die Informationen aus sich heraus Rückschluss auf eine einzelne Person zulassen, handelt es sich um Daten einer bestimmten Person[9]. (Beispiel: Bundeskanzlerin der BRD = Dr. Angela Merkel – schließlich gab es bislang nur eine Bundeskanzlerin.) Dies gilt ebenso, wenn andere Informationen existieren, welche in Verbindung mit den vorhandenen Daten indirekt einen Rückschluss auf eine einzelne Person zulassen. Da ErwGr. 26 DS-GVO auf die Mittel Bezug nimmt, die „nach allgemeinem Ermessen wahrscheinlich zur Identifizierung der natürlichen Person genutzt werden“, müssen Mittel sowohl von dem für die Verarbeitung Verantwortlichen als auch von einem „Dritten“ betrachtet werden. Für die Einstufung eines Datums als „personenbezogenes Datum“ ist es somit nicht erforderlich, dass sich alle zur Identifizierung der betreffenden Person erforderlichen Informationen in den Händen einer einzigen natürlichen oder juristischen Person befinden, sondern auch das Wissen von Dritten muss berücksichtigt werden.[10]

Der Begriff „identifizierbar“ muss somit im Sinne von „als Einzelperson wahrnehmbar“ bzw. einer „Einzelperson zuordenbar“ verstanden werden, wobei bei der Bewertung auch evtl. vorhandenes Zusatzwissen, über welches der Verantwortliche selbst oder beliebige Dritte verfügen, einfließen kann. Dem Verantwortlichen muss allerdings bekannt sein, dass es „Dritte“ gibt, welche über die erforderlichen Kenntnisse und Mittel verfügt, um die betroffene Person zu identifizieren.[11]

---

[7] EuGH, Urt. v. 19. Juni 2014, Az. C-345/13, Rn. 31. Online, zitiert am 2023-11-01; verfügbar unter https://dejure.org/2014,13697 bzw. Volltext des Urteils unter https://eur-lex.europa.eu/legal-content/de/TXT/?uri=CELEX:62013CJ0345

[8] Bzgl. indirekter Identifizierbarkeit siehe auch:

- EuGH, Urt. v. 19. Oktober 2016, Az. C-582/14, Rn. 41: „[…] dass es für die Einstufung einer Information als personenbezogenes Datum nicht erforderlich ist, dass die Information für sich genommen die Identifizierung der betreffenden Person ermöglicht.“ Online, zitiert am 2023-11-01; verfügbar unter https://dejure.org/2016,33959 bzw. Urteil unter https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A62014CJ0582
- BGH, Urt. v. 16. Mai 2017, Az. VI ZR 135/13, Rn. 28: „[…] dass es für die Einstufung einer Information als personenbezogenes Datum nicht erforderlich sei, dass die Information für sich genommen die Identifizierung der betreffenden Person ermögliche.“ Online, zitiert am 2023-11-01; verfügbar unter https://dejure.org/2017,15139 bzw. Volltext Urteil unter https://openjur.de/u/2117724.html

[9] Karg M. (2015) Anonymität, Pseudonyme und Personenbezug revisited. DuD: 520-526

[10] EuGH, Urt. v. 19.Oktober 2016, Az. C-582/14, Rn. 43. Online, zitiert am 2023-11-01; verfügbar unter https://dejure.org/2016,33959 bzw. Urteil unter https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A62014CJ0582

[11] Schlussanträge des Generalanwalts M. Campos Sánchez-Bordona vom 12. Mai 2016, Rn. 64-68. Online, zitiert am 2023-11-01; verfügbar unter https://eur-lex.europa.eu/legal-content/de/TXT/?uri=CELEX:62014CC0582

Die Identifizierbarkeit ist damit Dreh- und Angelpunkt hinsichtlich der Beurteilung, ob Daten als anonym oder pseudonym angesehen werden können.

In Bezug auf die in Art. 9 Abs. 1 DS-GVO benannten besonders sensiblen Daten ist dabei zu beachten, dass entsprechend der Rechtsprechung des EuGH die Zuordnung eines Datums als „sensibles Datum" weit zu verstehen ist[12]: Auch wenn ein Datum aufgrund der eigenen Bedeutung nach kein sensibles Datum darstellen würde, so ist zu prüfen, ob Daten, aus denen „mittels gedanklicher Kombination oder Ableitung" auf in Art. 9 Abs. 1 DS-GVO genannte Datenkategorien geschlossen werden kann, als sensible Daten i. S. d. Art. 9 Abs. 1 DS-GVO anzusehen sind. Ist somit ein Datum als „personenbezogen oder personenbeziehbar" klassifiziert, so muss bei der Prüfung, ob es sich um ein sensibles Datum i. S. d. Art. 9 Abs. 1 DS-GVO handelt, entsprechend vorgegangen werden. Auch indirekt mögliche Aussagen sind dabei prüfen.

> Beispiel: Eine Person besucht eine Arztpraxis. Die Standortdaten, also Straße, Postleitzahl und Ort, der Arztpraxis stellen eigentlich keine sensiblen Daten i. S. v. Art. 9 Abs. 1 DS-GVO dar. Da aber bekannt ist, dass eine Arztpraxis aufgesucht wird und dies i. d. R. für eine medizinische Betreuung erfolgt, ist diese Information damit als sensibles Datum aufzufassen.

Weiterhin urteilte der EuGH, dass ein Datensatz, der sowohl sensible als auch nicht sensible Daten enthält, insgesamt als sensibles Datum i. S. v. Art. 9 Abs. 1 DS-GVO anzusehen ist.[13]

4.2 Pseudonymisierung

Der Begriff der Pseudonymisierung wird in Art. 4 Ziff. 5 DS-GVO definiert. Dort heißt es:

> „"Pseudonymisierung" 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".

Dieser Definition folgend charakterisiert eine Pseudonymisierung daher Nachfolgendes:

- Die Pseudonymisierung ist eine Verarbeitung personenbezogener Daten.
- Pseudonyme Daten sind Daten, die ohne weitere Informationen einer spezifischen Person nicht zuordenbar sind.
- Die zur Zuordenbarkeit benötigten Informationen stehen dem Verantwortlichen nicht zur Verfügung, sondern
  - werden gesondert aufbewahrt und
  - sind durch technische und organisatorische Maßnahmen vor dem Zugriff durch den Verantwortlichen geschützt.

→ Für den Verantwortlichen besteht bei der Verarbeitung von pseudonymisierten Daten keine Möglichkeit der Identifizierung der betroffenen Person.

---

[12] EuGH, Urt. v. 2022-08-01, Az. C-92/09, C-93/09, Rn. 119, 120, 125. Online, zitiert am 2023-11-01; verfügbar unter https://dejure.org/2010,236 bzw. Volltext abrufbar unter https://eur-lex.europa.eu/legal-content/DE/TXT/?qid=1698904362512&uri=CELEX%3A62020CJ0184

[13] EuGh, Urt. v. 20263-07-04, Az. C-252/21, Rn. 89 sowie Leitsatz 2. Online, zitiert am 2023-12-29; verfügbar unter https://eur-lex.europa.eu/legal-content/de/TXT/?uri=CELEX:62021CJ0252

Hinweis: ErwGr. 26 führt aus, dass zur Feststellung, ob eine natürliche Person identifizierbar ist, alle Mittel berücksichtigt werden sollten, die von dem Verantwortlichen oder einer anderen Person nach allgemeinem Ermessen wahrscheinlich genutzt werden, um die natürliche Person direkt oder indirekt zu identifizieren.
Bei der Feststellung, ob Mittel nach allgemeinem Ermessen wahrscheinlich zur Identifizierung der natürlichen Person genutzt werden können, sollten alle objektiven Faktoren, wie die Kosten der Identifizierung und der dafür erforderliche Zeitaufwand, herangezogen werden, wobei die zum Zeitpunkt der Verarbeitung verfügbare Technologie und technologische Entwicklungen zu berücksichtigen sind.

4.3 Pseudonyme Daten

Entsprechend der in der DS-GVO enthaltenen Definition von Pseudonymisierung sind – in Anlehnung an ErwGr. 26 S. 2 DS-GVO – pseudonyme Daten demnach solche Daten, welche der oder die Verantwortlichen keiner spezifischen Person zuordnen können, jedoch für andere durch die Einbeziehung weitergehender Informationen („Zuordnungsregeln“) die grundsätzliche Möglichkeit der Zuordnung besteht. Dafür ist es nicht erforderlich, dass die betroffene Person durch die „Re-Identifizierung“ mit bürgerlichem Namen zu identifizieren ist[9]. Ausreichend ist vielmehr, wenn durch das Datum bzw. die Daten die betroffene Person individualisiert wird und Aussagen über deren sachliche und persönliche Verhältnisse möglich sind; ein Name muss nicht vorhanden sein[14].

Maßstab für die Beurteilung, ob pseudonyme Daten vorliegen, bilden die Anforderungen in Art. 4 Ziff. 5 DS-GVO i. V. m. ErwGr. 26 DS-GVO. Zu beachten ist, dass nach ErwGr. 28 S. 2 DS-GVO durch „die ausdrückliche Einführung der ‘Pseudonymisierung‘ in dieser Verordnung nicht beabsichtigt ist, andere Datenschutzmaßnahmen auszuschließen“. Pseudonymisierung stellt also nur eine weitere Maßnahme dar; alle Vorgaben der DS-GVO müssen daher vollumfänglich auch für pseudonyme Daten gewährleistet werden.

4.4 Anonymisierung

Anonymisierung wird in Art. 2 Ziff. 7 der Richtlinie (EU) 2019/1024[15] wie folgt definiert:

> „Anonymisierung“ den Prozess, in dessen Verlauf Dokumente in anonyme Dokumente umgewandelt werden, die sich nicht auf eine identifizierte oder identifizierbare natürliche Person beziehen, oder personenbezogene Daten so anonym gemacht werden, dass die betroffene Person nicht oder nicht mehr identifiziert werden kann.

Diese Begriffsbestimmung entspricht der Definition der internationalen Norm ISO/IEC 29100[16]:

---

[14] Artikel-29-Datenschutzgruppe. WP 136 „Stellungnahme 4/2007 zum Begriff 'personenbezogene Daten'“, S. 16: […] ein Name zur Identifizierung einer Person jedoch keineswegs immer notwendig ist“. Online, zitiert am 2023-10-14; verfügbar unter http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf

[15] Richtlinie (EU) 2019/1024 des Europäischen Parlaments und des Rates vom 20. Juni 2019 über offene Daten und die Weiterverwendung von Informationen des öffentlichen Sektors. Online, zitiert am 2023-10-14; verfügbar unter https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A32019L1024&qid=1697259421370

[16] ISO/IEC 29100:2011: „Information technology - Security techniques - Privacy framework. Online, zitiert am 2023-11-22; verfügbar unter https://www.iso.org/standard/45123.html bzw. download pdf-Datei kostenlos unter https://standards.iso.org/ittf/PubliclyAvailableStandards/index.html
Deutsche Übersetzung der Norm von 2011 wurde 2020 vom Beuth-Verlag veröffentlicht: https://www.beuth.de/de/norm/din-en-iso-iec-29100/325198919

„anonymity: characteristic of information that does not permit a personally identifiable information principal to be identified directly or indirectly“,

Ins Deutsche übersetzt „Anonymität: Merkmale von Informationen, die eine direkte oder indirekte Identifizierung des Betroffenen nicht zulassen“.

Wo die europäische Richtlinie den Prozess der Anonymisierung adressiert, definiert die internationale Norm ISO/IEC 29100 den Begriff der Anonymität; das eine ist das Resultat des anderen.

Europäische Richtlinien müssen durch nationale Rechtsakte umgesetzt werden.[17] Im Falle der Begriffsdefinition zur Anonymisierung erfolgte dies durch § 3 Ziff. 12 Datennutzungsgesetz:

> „Anonymisierung“ ist der Prozess, in dessen Verlauf personenbezogene Daten in Daten umgewandelt werden, die sich nicht auf eine identifizierte oder identifizierbare natürliche Person beziehen, oder derart in Daten umgewandelt werden, dass die betroffene Person nicht oder nicht mehr identifiziert werden kann.

4.4.1 Deutsche vs. EU-Regelung

Die deutsche und die europäische Regelung sind nicht wirklich deckungsgleich: Art. 2 Ziff. 7 der Richtlinie (EU) 2019/1024 adressiert „Dokumente“, § 3 Ziff. 12 Datennutzungsgesetz „Daten“.

Sieht man sich die englische Definition an, der Sprache, in welcher der Text zwischen den Mitgliedsstaaten verhandelt wurde, so findet sich auch im englischen Text der Richtlinie (EU) 2019/1024 der Begriff „document“. Es handelt sich bei der Verwendung des Begriffs „Dokument“ in der deutschen Version der Art. 2 Ziff. 7 der Richtlinie (EU) 2019/1024 daher um keinen Übersetzungsfehler, wie er in Texten der EU immer wieder einmal vorkommt.

Bei der Interpretation muss man jedoch beachten, dass Art. 3 Ziff. 6 der Richtlinie (EU) 2019/1024 den Begriff „Dokument“ definiert:

> „Dokument“
> a) jeden Inhalt unabhängig von der Form des Datenträgers (auf Papier oder in elektronischer Form oder als Ton-, Bild- oder audiovisuelles Aufnahme); oder
> b) einen beliebigen Teil eines solchen Inhalts.

Unter dieser Definition des Wortes „Dokument“ ist die deutsche Umsetzung unter Verwendung „Daten“ dem Sinngehalt der Regelung sehr ähnlich.

Dennoch empfiehlt es sich, im Zweifelsfall die Begriffsbestimmung in Art. 2 Ziff. 7 der Richtlinie (EU) 2019/1024 zumindest parallel anzuwenden, denn wie der BGH urteilte[18]: Der Gesetzgeber muss die

---

[17] Deutscher Bundestag, Wissenschaftliche Dienste: Kurzinformation – Umsetzung von EU-Richtlinien und Verfassungsrecht. Online, zitiert am 2023-10-31; verfügbar unter https://www.bundestag.de/resource/blob/899872/33b2422d86eab34c741b67207ab1bda3/WD-3-045-22-pdf-data.pdf

[18] Siehe

- EuGH, Urt. v. 2019-10-01, Az. C-673/17. Online, zitiert am 2023-10-14; verfügbar unter https://dejure.org/2019,31907 bzw. Volltext Urteil beim EuGH unter https://curia.europa.eu/juris/document/document.jsf?text=&docid=218462&pageIndex=0&doclang=DE&mode=req&dir=&occ=first&part=1
- Darauf aufbauend: BGH, Urt. v. 2020-05-28, Az. I ZR 7/16. Siehe insbesondere Abschnitt „b“ des Urteilsspruches. Online, zitiert am 2023-10-14; verfügbar unter https://dejure.org/2020,12443 bzw. Volltext unter https://juris.bundesgerichtshof.de/cgi-bin/rechtsprechung/document.py?Gericht=bgh&Art=en&nr=107623&pos=0&anz=1

europäischen Vorgaben beachten, sodass gegebenenfalls auch eine europarechtskonforme Auslegung nationaler Normen erfolgen muss.

4.4.2 Von der EU-Vorgabe abweichende Regelungen

Einige Datenschutzgesetze von deutschen Bundesländern enthalten eigene Begriffsdefinitionen, z. B.

- § 3 BbgDSG[19]
- § 11 Abs. 2 HmbDSG[20]
- § 2 Ziff. 4 HDSIG[21]
- § 4 DSG NRW[22]
- § 13 Abs. 2 LDSG SH[23]
- § 28 Abs. 3 ThürDSG[24]

Entsprechend Art. 4 Abs. 1 AEUV handelt es sich beim Datenschutz um eine mit den Mitgliedstaaten geteilte Zuständigkeit. Insbesondere auch durch die DS-GVO übte die Union ihre Zuständigkeit für den Datenschutz i.S.v. Art. 2 Abs. 2 S. 2 AEUV[25] aus; mitgliedstaatliche Aktivitäten kommen gemäß Art. 2 Abs. 1 Hs. 2 AEUV nur in Betracht, wenn die europäischen Regelungen die Mitgliedstaaten zu eigenen Maßnahmen ermächtigt. Vor Wirkeintritt der Richtlinie (EU) 2019/1024 und der darin enthaltenen Definition der Anonymisierung konnte man vielleicht aus Art. 6 Abs. 2 und 3 DS-GVO ein Recht der Mitgliedstaaten ableiten, den Begriff der Anonymisierung zu konkretisieren; in der Literatur wird jedoch die Meinung vertreten, dass die existierenden landesrechtlichen Regelungen nicht den Vorgaben der DS-GVO genügen[26].

Spätestens seit dem Wirkeintritt der Richtlinie (EU) 2019/1024 werden nationale Definitionen durch das europäische Recht überlagert: Entsprechend Art. 288 Abs. 2 AEUV[27] gilt, dass Unionsrecht vorrangig gegenüber nationalem Recht anzuwenden ist.[28]

---

[19] Gesetz zum Schutz personenbezogener Daten im Land Brandenburg (Brandenburgisches Datenschutzgesetz - BbgDSG). Online, zitiert am 2023-10-14; verfügbar unter https://bravors.brandenburg.de/gesetze/bbgdsg

[20] Hamburgisches Datenschutzgesetz (HmbDSG). Online, zitiert am 2023-10-14; verfügbar unter https://dsgvo-gesetz.de/hmbdsg/

[21] Hessisches Datenschutz- und Informationsfreiheitsgesetz (HDSIG). Online, zitiert am 2023-10-14; verfügbar unter https://dsgvo-gesetz.de/hdsig/

[22] Datenschutzgesetz Nordrhein-Westfalen (DSG NRW). Online, zitiert am 2023-10-14; verfügbar unter https://recht.nrw.de/lmi/owa/br_text_anzeigen?v_id=3520071121100436275

[23] Schleswig-Holsteinisches Gesetz zum Schutz personenbezogener Daten (LDSG SH). Online, zitiert am 2023-10-14; verfügbar unter https://dsgvo-gesetz.de/ldsg-sh/

[24] Thüringer Datenschutzgesetz (ThürDSG). Online, zitiert am 2023-10-14; verfügbar unter https://dsgvo-gesetz.de/thuerdsg/

[25] Vertrag über die Arbeitsweise der Europäischen Union. Art. 2 Abs. 2: „Übertragen die Verträge der Union für einen bestimmten Bereich eine mit den Mitgliedstaaten geteilte Zuständigkeit, so können die Union und die Mitgliedstaaten in diesem Bereich gesetzgeberisch tätig werden und verbindliche Rechtsakte erlassen. Die Mitgliedstaaten nehmen ihre Zuständigkeit wahr, sofern und soweit die Union ihre Zuständigkeit nicht ausgeübt hat. Die Mitgliedstaaten nehmen ihre Zuständigkeit erneut wahr, sofern und soweit die Union entschieden hat, ihre Zuständigkeit nicht mehr auszuüben.“ Online, zitiert am 2023-10-14; verfügbar unter https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX:12012E/TXT

[26] Meyer S. (2021) Landesrechtliche Legaldefinitionen der „Anonymisierung“ im Anwendungsbereich der DS-GVO. Kompetenzielle und inhaltliche Vereinbarkeit mit dem Unionsrecht. ZD: 669-674

[27] Konsolidierte Fassung des Vertrags über die Arbeitsweise der Europäischen Union. Online, zitiert am 2023-10-14; verfügbar unter https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX:12012E/TXT

[28] Bzgl. Umfang der vorrangigen Anwendung siehe auch EuGH, Urt. V. 05.12.2017 Az. C-42/17. Online, zitiert am 2023-10-14; verfügbar unter https://dejure.org/2017,46354

Auch einige Verbände verwenden in ihren Veröffentlichungen eigene Begriffsbestimmungen zur Anonymisierung, so finden sich eigene Begriffsbestimmungen z. B. in

- Bundesverband der Deutschen Industrie e. V. (BDI): Broschüre „Anonymisierung personenbezogener Daten“, Kapitel 3.3[29]
- Stiftung Datenschutz: Grundsatzregeln für die Anonymisierung personenbezogener Daten, Kapitel 2[30]
- Stiftung Datenschutz: Praxisleitfaden für die Anonymisierung personenbezogener Daten, Kapitel 2.2[31]

Ob die jeweiligen Organisationen die Richtlinie (EU) 2019/1024 nicht kannten oder einfach nur ignorierten ist letztlich unerheblich: Auch hier gilt selbstverständlich, dass das EU-Recht anwendbar und die Definition der Richtlinie (EU) 2019/1024 somit vorrangig zu beachten und zu verwenden ist.

Auch deutsche Aufsichtsbehörden definieren die Anonymisierung mitunter ohne Beachtung der europäischen Begriffsbestimmung. So findet sich beispielsweise auch im 2020 erschienenen Positionspapier „zur Anonymisierung unter der DSGVO unter besonderer Berücksichtigung der TK-Branche“[32] des BfDI im Abschnitt 1.2 eine eigene Begriffsbestimmung. Natürlich können auch die Datenschutz-Aufsichtsbehörden keine Legaldefinition erlassen, dies kann nur der Gesetzgeber; Datenschutz-Aufsichtsbehörden sind Exekutivorgane und wenden das Recht an, erlassen es aber nicht.

4.5 Anonyme Daten

Entsprechend ErwGr. 52 der Richtlinie (EU) 2019/1024 sind anonyme Informationen solche Informationen, welche

> „sich nicht auf eine identifizierte oder identifizierbare natürliche Person beziehen, bzw. Informationen, die sich auf personenbezogene Daten beziehen, die so anonymisiert wurden, dass die betroffene Person nicht oder nicht mehr identifiziert werden kann.“

Daraus ergibt sich im Umkehrschluss, dass anonyme Daten weder direkt personenbezogene Daten noch pseudonymisierte Daten sein können. D. h., anonyme Daten sind Daten, bei denen während

[29] Bundesverband der Deutschen Industrie e. V. (BDI) „Anonymisierung personenbezogener Daten“. Stand 2020-02-11. Online, zitiert am 2023-10-14; verfügbar unter https://bdi.eu/publikation/news/anonymisierung-personenbezogener-daten/

[30] Stiftung Datenschutz: Grundsatzregeln für die Anonymisierung personenbezogener Daten. Stand 2022-12. Online, zitiert am 2023-10-14; verfügbar unter https://stiftungdatenschutz.org/fileadmin/Redaktion/Dokumente/Anonymisierung_personenbezogener_Daten/SDS_Studie_Grundsatzregeln_Web_01.pdf

[31] Stiftung Datenschutz: Praxisleitfaden für die Anonymisierung personenbezogener Daten. Stand: 2022-12. Online, zitiert am 2023-10-14; verfügbar unter https://stiftungdatenschutz.org/fileadmin/Redaktion/Dokumente/Anonymisierung_personenbezogener_Daten/SDS_Studie_Praxisleitfaden-Anonymisieren-Web_01.pdf

[32] Bundesbeauftragte für den Datenschutz und die Informationsfreiheit (BfDI): Positionspapier zur Anonymisierung unter der DSGVO unter besonderer Berücksichtigung der TK-Branche. Stand: 2020-06-29. Online, zitiert am 2023-10-14; verfügbar unter https://www.bfdi.bund.de/DE/Fachthemen/Inhalte/Telefon-Internet/Positionen/Positionspapier-Anonymisierung-DSGVO-TKG.html bzw. pdf-Datei unter https://www.bfdi.bund.de/SharedDocs/Downloads/DE/Konsultationsverfahren/1_Anonymisierung/Positionspapier-Anonymisierung.pdf?__blob=publicationFile&v=6

des gesamten Lebenszyklus der Daten keine Zuordnungsmöglichkeit zu einer spezifischen betroffenen Person existiert[33].

Am 26. April 2023 urteilte das EuG[34], dass es sich bei der Beurteilung, ob es sich bei an Dritten übermittelte Daten um personenbezogene Daten handelt, auf das Verständnis abzustellen ist, das dieser Dritte bei der Bestimmung der Frage hat, ob die ihm übermittelten Informationen sich auf „identifizierbare Personen" beziehen. Wenn nur der Sender, nicht jedoch der Empfänger eine Re-Identifizierung durchführen kann, seien diese Daten anonym. Gegen diese Entscheidung legte der Europäischen Datenschutzbeauftragte Rechtsmittel ein[35], das Verfahren ist zum Zeitpunkt der Erstellung dieser Praxishilfe beim EuGH anhängig.

4.6 Pseudonyme vs. anonyme Daten: Kurzdarstellung der Unterschiede

Sowohl pseudonyme als auch anonyme Daten sind für den Verantwortlichen keiner spezifischen betroffenen Person zuordenbar. Der Unterschied zwischen anonymen und pseudonymen Daten besteht darin, dass bei pseudonymen Daten außerhalb der Zugriffsmöglichkeiten des für die Verarbeitung Verantwortlichen grundsätzlich eine Zuordnungsmöglichkeit besteht oder bestehen könnte, bei anonymen Daten hingegen für niemanden eine Zuordnungsmöglichkeit vorhanden ist.

Pseudonyme Daten	Anonyme Daten
Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zuordenbar	Betroffene Person kann mit Hilfe der Daten nicht mehr identifiziert werden
Zusätzliche Informationen müssen gesondert aufbewahrt werden, sodass die Verantwortlichen keine Zugriffsmöglichkeit auf diese Informationen haben	Es existieren keine zusätzlichen Informationen, mit denen Daten einer spezifischen Person zuordenbar sind.
Pseudonymisierung ist reversibel	Anonymisierung ist irreversibel
Re-Identifizierung möglich	Re-Identifizierung nicht möglich

[33] Voigt P, von dem Bussche A. The EU General Data Protection Regulation (GDPR) - A Practical Guide. Springer Verlag, 2017. ISBN 978-3-319-57958-0. PP 13-16, chapter „2.1.2.2 Anonymisation and Pseudonymisation": „Anonymised data is either information that does not relate to an identified or identifiable individual or personal data that was rendered anonymous in such a manner that the person is not or no longer identifiable."

[34] EuG: Urt. v. 2023-04-26, AZ. T-557/20. Online, zitiert am 2023-10-14; verfügbar unter https://eur-lex.europa.eu/legal-content/de/TXT/?uri=CELEX:62020TJ0557

[35] EuGH: Rechtssache C-413/23 P. Online, zitiert am 2023-10-14; verfügbar unter https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A62023CN0413&qid=1697259421370

5 Rechtliche Rahmenbedingungen

5.1 Erlaubnistatbestand Pseudonymisierung/Anonymisierung

Bei der Pseudonymisierung als auch bei der Anonymisierung handelt es sich um eine Verarbeitung gemäß Art. 4 Ziff. 2 DS-GVO.[36] Es ist daher sowohl für eine Anonymisierung als auch für eine Pseudonymisierung von Gesundheitsdaten ein Erlaubnistatbestand gem. Art. 9 Abs. 2 DS-GVO[37] und ergänzend[38] Art. 6 Abs. 1 DS-GVO[39] erforderlich. Für Daten, welche nicht zu den in Art. 9 Abs. 1 DS-GVO genannten besonderen Kategorien zählen, müssen nur die in Art. 6 Abs. 1 DS-GVO genannten Rechtmäßigkeitsvoraussetzungen erfüllt sein .

Dabei ist zu beachten, dass Art. 5 Abs. 1 lit. b DS-GVO keine strenge Zweckbindung fordert[40], sondern lediglich verlangt, dass der Sekundärzweck mit dem Primärzweck „vereinbar" ist („dürfen nicht in einer mit diesen Zwecken nicht zu vereinbarenden Weise weiterverarbeitet werden"). Bei der Beurteilung, ob ein Sekundärzweck mit dem Primärzweck vereinbar ist, muss Art. 6 Abs. 4 DS-GVO berücksichtigt werden.

5.1.1 Einwilligung

Existiert kein gesetzlicher Erlaubnistatbestand, so ist zur Pseudonymisierung oder Anonymisierung die Einholung einer rechtsgültigen Einwilligung[41] der betroffenen Person bzw. Personen erforderlich. Hierbei sind alle von der DS-GVO genannten Vorgaben/Rahmenbedingungen einzuhalten.[42]

5.1.2 Sonderfall „Zweckvereinbarkeit"

Entsprechend Art. 5 Abs. 1 lit. b Hs. 1 DS-GVO müssen personenbezogene Daten „für festgelegte, eindeutige und legitime Zwecke erhoben werden und dürfen nicht in einer mit diesen Zwecken nicht zu vereinbarenden Weise weiterverarbeitet werden". Diese Anforderung ist unter dem Begriff

---

[36] EDSA (2021-02-02) EDPB Document on response to the request from the European Commission for clarifications on the consistent application of the GDPR, focusing on health research. Rn. 43. Online, zitiert am 2023-11-22; verfügbar unter https://edpb.europa.eu/our-work-tools/our-documents/other-guidance/edpb-document-response-request-european-commission_en

[37] Zu beachten: „Zudem ist Art. 9 Abs. 2 DSGVO, da er eine Ausnahme vom Grundsatz des Verbots der Verarbeitung besonderer Kategorien personenbezogener Daten vorsieht, eng auszulegen". EuGH, Urt. v-2023-07-04, Az. C-252/21, Rn. 76. Online, zitiert am 2023-12-29; verfügbar unter https://eur-lex.europa.eu/legal-content/de/TXT/?uri=CELEX:62021CJ0252

[38] EuGH, Urt. v. 2023-12-21, Az. C-667/21, Rn. 79. Online, zitiert am 2023-12-29; verfügbar unter https://eur-lex.europa.eu/legal-content/de/TXT/?uri=CELEX:62021CJ0667

[39] Zu beachten: „[…] sind die in Art. 6 Abs. 1 Unterabs. 1 Buchst. b bis f DSGVO vorgesehenen Rechtfertigungsgründe eng auszulegen, da sie dazu führen können, dass eine Verarbeitung personenbezogener Daten trotz fehlender Einwilligung der betroffenen Person rechtmäßig ist". EuGH, Urt. v-2023-07-04, Az. C-252/21, Rn. 93. Online, zitiert am 2023-12-29; verfügbar unter https://eur-lex.europa.eu/legal-content/de/TXT/?uri=CELEX:62021CJ0252

[40] Roßnagel A, Geminn C. (2021) Vertrauen in Anonymisierung. ZD: 487-490

[41] Siehe auch EuGH Urt. v. 2021-01-18, Az. C-61/19, Tenor: Verantwortlicher muss nachweisen, dass betroffene Person Einwilligung durch aktives Verhalten bekundet hat und vorher Information über alle Umstände im Zusammenhang mit der Verarbeitung erhielt, welcher der Person erlaubten, die Konsequenzen dieser Einwilligung leicht zu ermitteln, sodass gewährleistet ist, dass die Einwilligung in voller Kenntnis der Sachlage erteilt wird. Online, zitiert am 2023-12-29; verfügbar unter https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A62019CA0061&qid=1703360472835

[42] Näheres siehe: Ausarbeitung der GMDS AG DIG „Anforderungen der DS-GVO an die Einwilligung". Online, zitiert am 2023-10-14; verfügbar unter http://ds-gvo.gesundheitsdatenschutz.org/html/einwilligung.php

„Zweckbindung“ allgemein bekannt. Eine spätere Zweckänderung ist daher grundsätzlich nicht zulässig.

In ErwGr. 50 S. 1 DS-GVO findet sich, dass eine **Verarbeitung für andere Zwecke** als die, für die die personenbezogenen Daten ursprünglich erhoben wurden, **zulässig** sein sollte, **wenn** die Verarbeitung **mit den Zwecken, für die die personenbezogenen Daten ursprünglich erhoben wurden, vereinbar** ist. Dies entspricht auch der Forderung der Zweckbindung in Art. 5 Abs. 1 lit. b DS-GVO, welche nur Verarbeitungen zu Zwecken verbietet, die nicht mit den Zwecken, zu denen die Daten erhoben wurden, vereinbar ist.

ErwGr. 50 S. 2 DS-GVO führt weiterhin aus, dass in Fällen der Vereinbarkeit des neuen Zweckes mit dem ursprünglichen Zweck „keine andere gesonderte Rechtsgrundlage erforderlich [ist] als diejenige für die Erhebung der personenbezogenen Daten“. Hieraus wird in der juristischen Literatur teilweise abgeleitet, dass in diesen Fällen kein weiterer Erlaubnistatbestand benötigt wird, andere lehnen diese Sichtweise ab, da Art. 6 Abs. 1 DS-GVO grundsätzlich die Erfüllung der im ersten Absatz genannten Bedingungen fordert.

Zu beachten ist, dass bei einer Zweckänderung entsprechend Art. 6 Abs. 4 DS-GVO eine Prüfung auf Zweckvereinbarkeit mit dem ursprünglichen Zweck erfolgen muss. Art. 6 Abs. 4 DS-GVO enthält einen nicht abschließenden Katalog, welcher als Grundlage zur Abklärung der Vereinbarkeit von neuem und ursprünglichem Zweck bearbeitet werden muss.

Bei der „Zweckvereinbarkeit“ als Rechtsgrundlage muss daher beachtet werden:

1) Ob bei Vorhandensein einer Zweckvereinbarkeit kein weiterer Erlaubnistatbestand benötigt wird, ist rechtlich nicht geklärt, auch wenn viel für die Annahme spricht. Wer Sicherheit benötigt (z. B. weil bei einer gerichtlich erzwungenen Löschung von Daten ggf. ein Produktrückruf erforderlich wird), sollte bzgl. der neuen Verarbeitung einen eigenen Erlaubnistatbestand nachweisen können.
2) Grundsätzlich muss bei der Verarbeitung von in Art. 9 Abs. 1 DS-GVO genannten besonderen Kategorien von Daten immer mindestens einer der in Art. 9 Abs. 2 DS-GVO genannten Erlaubnistatbestände die Verarbeitung legitimierend, ergänzend muss auch mindestens eine der in Art. 6 Abs. 1 DS-GVO genannten Rechtmäßigkeitsvoraussetzungen erfüllt sein. [43] Somit reicht eine Interessenabwägung nach Art. 6 Abs. 1 lit. f DS-GVO alleine nicht aus, um eine Verarbeitung von genetischen oder Gesundheitsdaten zu legitimieren: Es muss immer auch mindestens ein Erlaubnistatbestand entsprechend Art. 9 Abs. 2 lit. a-j DS-GVO erfüllt sein.
3) Die Vereinbarkeit des Zweckes, zu dem die Daten ursprünglich erhoben wurden (siehe ErwGr. 50 S. 1 DS-GVO, eine weitere Weitergabe zu später neu aufgekommenen Zwecken wird damit vermutlich eingeschränkt), mit dem neuen Zweck muss nachgewiesen werden. Dabei müssen auf jeden Fall mindestens die Vorgaben in Art. 6 Abs. 4 lit. a bis e DS-GVO betrachtet und bewertet werden.
4) Auch bei Annahme, dass die Auffassung, der ursprüngliche Erlaubnistatbestand reiche für eine zweckändernde Verarbeitung bei einer Zweckvereinbarkeit von altem und neuem Zweck aus, richtig ist, gilt der ursprüngliche Erlaubnistatbestand ausschließlich für den Verantwortlichen, der ursprünglich die Daten für den ursprünglichen Zweck erhob und

[43] EuGH, Urt. v. 2023-12-21, Az. C-667/21, Rn. 79. Online, zitiert am 2023-12-29; verfügbar unter https://eur-lex.europa.eu/legal-content/de/TXT/?uri=CELEX:62021CJ0667

verarbeitet. Jeder andere Verantwortliche, der Daten für den neuen Zweck verarbeiten will, braucht auf jeden Fall einen eigenen Erlaubnistatbestand. Selbst dann, wenn eine gemeinsame Verantwortlichkeit für Datenverarbeitungsvorgänge entsprechend Art. 26 DS-GVO („Gemeinsam Verantwortliche“) vorliegt.

5) Will ein Verantwortlicher Daten anonymisieren, um die anonymisierten Daten Dritten zu deren Zwecken zu übergeben, wird i. d. R. die Anonymisierung nicht mit dem ursprünglichen Zweck vereinbar sein, d. h. allein für eine Anonymisierung wird dieser Erlaubnistatbestand kaum anwendbar sein.

5.1.3 Sonderfall Forschung

Die DS-GVO privilegiert Verarbeitungen personenbezogener Daten zum Zwecke der wissenschaftlichen Forschung an unterschiedlichen Stellen. Insbesondere enthält die DS-GVO privilegierende Bestimmungen für wissenschaftliche und historische Forschungszwecke[44]. Hinsichtlich der Nutzung besonderer Kategorien personenbezogener Daten findet sich in Art. 9 Abs. 2 lit. j DS-GVO ein datenschutzrechtlicher Erlaubnistatbestand zur Nutzung von Daten zu Zwecken der wissenschaftlichen Forschung. Hiernach ist eine Verarbeitung gestattet, wenn die Verarbeitung gemäß Art. 89 Abs. 1 DS-GVO erforderlich ist und ein nationales oder europäisches Recht, welches den besonderen Anforderungen von Art. 89 Abs. 1 DS-GVO genügt, die Verarbeitung erlaubt oder sogar fordert.

Die Verarbeitung zu wissenschaftlichen Forschungszwecken stellt jeden Verantwortlichen vor die besondere Herausforderung, die geplante Verarbeitung an den allgemein üblichen Definitionen des Forschungsbegriffes zu messen, da man nicht bei jeder Form der Forschung automatisch davon ausgehen kann, dass die Verarbeitung den Regelungen zur wissenschaftlichen Forschung genügt. Da seitens der Gesetzgeber keine genaue Definition des Forschungsbegriffes vorgenommen wurde, aus den Erwägungsgründen der DS-GVO jedoch ersichtlich wird, was der Gesetzgeber mit „Forschung“ adressieren will, müssen die Ziele des europäischen Gesetzgebers vom geplanten Vorhaben nachweisbar adressiert werden. Sollte die geplante Forschung einen transparenten Gewinn für die Volksgesundheit darstellen, ist dies sicherlich der Fall. Auch Grundlagenforschung oder technologische (Weiter-)Entwicklungen werden vom europäischen Gesetzgeber entsprechend verschiedenernErwägungsgründe der DS-GVO dem Begriff „Forschung“ zugeordnet. Dabei ist die privat finanzierte Forschung der öffentlichen Forschung gleichgestellt.

5.1.3.1 Forschung ohne Einwilligung nach § 27 BDSG

In Deutschland erfolgte eine nationale Konkretisierung dieser Öffnungsklausel in § 27 BDSG. Die Regelung in § 27 BDSG gestattet die Verarbeitung besonderer Kategorien von Daten i. S. d. Art. 9 Abs. 1 DS-GVO zu Forschungszwecken auch ohne Einwilligung, wenn

a) die Verarbeitung zu diesen Zwecken erforderlich ist und
b) die Interessen der Verantwortlichen an der Verarbeitung die Interessen der betroffenen Person an der Nicht-Verarbeitung ihrer Daten erheblich überwiegen.

[44] Zu den Begriffsbestimmungen bzgl. „Forschung“, „wissenschaftliche Forschung“ und „historische Forschung“ siehe Ausarbeitung „Datenschutzrechtliche Anforderungen an die medizinische Forschung unter Berücksichtigung der EU Datenschutz-Grundverordnung (DS-GVO)“, herausgegeben von der Arbeitsgruppe „Datenschutz und IT-Sicherheit im Gesundheitswesen“ der GMDS und der Arbeitsgruppe „Datenschutz und Datensicherheit im Gesundheits- und Sozialwesen“ der GDD. Online, zitiert am 2023-10-14; verfügbar unter http://ds-gvo.gesundheitsdatenschutz.org/html/forschung.php

§ 27 Abs. 1 BDSG stellt somit einen Erlaubnistatbestand zur Verarbeitung von Gesundheitsdaten zu Forschungszwecken dar, wobei allerdings zu beachten ist, dass spezifischeres Bundes- oder Landesrecht vorrangig gelten kann (§ 1 Abs. 1 Nr. 2 BDSG)[45]. Weiterhin gilt § 27 Abs. 1 BDSG nur für die Verarbeitung von Daten i.S.v. Art. 9 Abs. 1 DS-GVO. Für Daten, welche nicht unter Art. 9 Abs. 1 DS-GVO fallen, müssen andere Erlaubnistatbestände gefunden werden, wie die in Art. 6 DS-GVO angegebenen Möglichkeiten.

Die von § 27 Abs. 1 BDSG geforderte Interessenabwägung stellt hohe Anforderungen: die Interessen des oder der Verantwortlichen müssen nicht nur überwiegen, sondern sie müssen **erheblich** überwiegen. Dementsprechend kann ein erhebliches Überwiegen des Interesses an der Forschung angenommen werden, wenn ein Forschungsvorhaben „erhebliche Verbesserungen für die Gesundheit oder soziale Sicherheit der Bevölkerung mit sich bringt“[46].

Entsprechend ErwGr. 47 DS-GVO stellt ein Faktor, der bei jeder Interessensabwägung berücksichtigt werden muss, die „vernünftigen Erwartungen“ der betroffenen Person dar, die auf der Beziehung der Person zu dem Verantwortlichen beruhen. Ein Patient, der zu seiner Behandlung in ein Krankenhaus oder eine Arztpraxis geht, wird als Erwartung regelhaft die Verarbeitung der Gesundheitsdaten zur Behandlung aufweisen: Die betroffene Person geht in die Gesundheitseinrichtung zum Zwecke der Behandlung. Zweck der Gesundheitseinrichtung ist i. d. R. ebenfalls die Behandlung und Betreuung von Patienten. Forschung mit diesen für die Behandlung preisgegebenen Daten wird somit i. d. R. nicht zu den „vernünftigen Erwartungen“ gehören. Eine Ausnahme kann – je nach öffentlicher Wahrnehmung – vorliegen, wenn die Behandlung in einem forschenden Universitätsklinikum erfolgt. Ein weiterer zu beachtender Punkt liegt nach ErwGr. 47 DS-GVO in dem Umstand, ob eine betroffene Person zum Zeitpunkt der Erhebung der personenbezogenen Daten und angesichts der Umstände, unter denen sie erfolgt, vernünftigerweise absehen kann, dass möglicherweise eine Verarbeitung für diesen Zweck erfolgen wird. Laut ErwGr 47 DS-GVO können die Interessen der betroffenen Person insbesondere dann überwiegen, wenn personenbezogene Daten in Situationen verarbeitet werden, in denen eine betroffene Person vernünftigerweise nicht mit einer weiteren Verarbeitung rechnen muss. Weiterhin zu beachten: Je sensibler die Daten, desto schwerer ist entsprechend der Rechtsprechung des EuGH[47] der Eingriff in die in den Art. 7 und 8 der Charta verankerten Grundrechte der betroffenen Person.

Die DSK sieht keine Möglichkeit, die „vernünftigen Erwartungen“ einer betroffenen Person durch Informationen nach Art. 13, 14 DS-GVO zu erweitern.[48] Insbesondere, da die betroffene Person keine

---

[45] Beispiele für vorrangige bereichsspezifische Regelungen sind, soweit sie den Vorgaben der DS-GVO entsprechen:
- Bundesrecht: Sozialgesetzbücher, Arzneimittelgesetz, Gendiagnostikgesetz usw.
- Landesrecht: Krankenhausgesetze, Krebsregistergesetze

[46] Buchner B, Tinnefeld M-T. § 27 BDSG RN. 12 in Kühling/Buchner, Kommentar zur Datenschutz-Grundverordnung/BDSG. C.H.Beck Verlag 2. Auflage 2018. ISBN 978-3-406-71932-5

[47] EuGH Urt. v. 2023-12-07, Az. C-26/22, C-64/22, Rn. 94. Online, zitiert am 2023-12-12; verfügbar unter https://eur-lex.europa.eu/legal-content/de/TXT/?uri=CELEX:62022CJ0026

[48] Datenschutzkonferenz (DSK): Orientierungshilfe der Aufsichtsbehörden zur Verarbeitung von personenbezogenen Daten für Zwecke der Direktwerbung unter Geltung der Datenschutz-Grundverordnung (DS-GVO): „Die Erwartungen der betroffenen Person können dabei nicht durch die nach der DS-GVO vorgesehenen Pflichtinformationen (Art. 13, 14 DS-GVO) erweitert werden." (Seite 4, Stand 2022-02-18) Online, zitiert am 2024-01-11; verfügbar unter https://www.datenschutzkonferenz-online.de/media/oh/OH-Werbung_Februar%202022_final.pdf

Pflicht zur Kenntnisnahme der Informationen trifft, erscheint diese Auffassung zutreffend: Inwieweit sollen Informationen, die nicht zur Kenntnis genommen werden müssen, Erwartungen der betroffenen Person beeinflussen? Hier wird mehr erforderlich sein, damit die vernünftigen Erwartungen einer betroffenen Person berührt werden.

Gemäß § 27 Abs. 3 BDSG sind besondere Kategorien personenbezogener Daten im Sinne des Art. 9 Abs. 1 DS-GVO zu anonymisieren, sobald dies nach dem Forschungs- oder Statistikzweck möglich ist, es sei denn, berechtigte Interessen der betroffenen Person stehen einer Anonymisierung entgegen. Ist dies der Fall, sind die Merkmale, mit denen Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren Person zugeordnet werden können, gesondert zu speichern (§ 27 Abs. 3 S. 2 BDSG). Die Neuregelung des § 27 Abs. 3 BDSG verlangt somit, Forschung grundsätzlich mit anonymen Daten durchzuführen und in Ausnahmefällen mit pseudonymen Daten.

Von einer Unvereinbarkeit einer Anonymisierung mit der Zweckerreichung eines Forschungsvorhabens kann zumindest immer dann ausgegangen werden, wenn eine fortlaufende Zuordnung von neuen Daten zu bereits vorhandenen Daten erforderlich ist[49]. Gleiches gilt, wenn im Rahmen des Forschungsvorhabens die betroffene Person ggf. kontaktiert werden muss, z. B., weil Forschungsergebnisse ihre Behandlung beeinflussen könnten. Weiterhin schließt die Nutzung von Biomaterialien, deren enthaltene genetische Informationen prinzipiell einen Personenbezug erlauben, eine Anonymisierung aus[49].

Auch wenn eine dingliche Übereignung der Biomaterialien durch die betroffene Person erfolgt ist, was in vielen Einverständniserklärungen von Patienten bei der Aufnahme in einer Klinik abgefragt wird, kann daraus keinerlei Rückschluss auf den Willen der Patienten bezüglich der Verarbeitung der enthaltenen genetischen Informationen gezogen werden; von einer im Rahmen der eigenen Behandlung erforderlichen Verarbeitung abgesehen.

Eine Re-Identifizierung, die bei einer Verarbeitung pseudonymer Daten grundsätzlich möglich ist, darf entsprechend § 27 Abs. 3 S. 3. BDSG nur durchgeführt werden, wenn dies der Forschungs- oder Statistikzweck erfordert[50].

5.2 Nachweispflichten

Entsprechend ErwGr. 26 DS-GVO sollen die Vorgaben der DS-GVO nicht für anonyme Daten gelten. Oder anderes ausgedrückt: Für anonyme Daten findet die DS-GVO keine Anwendung, denn die Anwendbarkeit der DS-GVO hängt nach Art. 2 Abs. 1 DS-GVO – vorbehaltlich der in Art. 2 Abs. 2 DS-

---

Anders Datenschutzbehörde Österreich, Bescheid v. 2023-08-01, Geschäftszahl 2023-0.544.853: „Wie den Feststellungen zu entnehmen ist, informiert die Beschwerdegegnerin ihre Kunden in transparenter Weise in ihrer Datenschutzerklärung s […] Folglich hat die Beschwerdeführerin damit zu rechnen, dass bei Benützung von Personenverkehrszügen der Beschwerdegegnerin ihre Jahreskarte elektronisch validiert wird.“ Online, zitiert am 2024-01-11; verfügbar unter https://www.ris.bka.gv.at/Dokumente/Dsk/DSBT_20230801_2023_0_544_853_00/DSBT_20230801_2023_0_544_853_00.html

[49] Buchner B, Tinnefeld M-T. § 27 BDSG RN. 24 in Kühling/Buchner, Kommentar zur Datenschutz-Grundverordnung/BDSG. C.H.Beck Verlag 2. Auflage 2018. ISBN 978-3-406-71932-5

[50] Bzgl. „erforderlich“ siehe Kap. 4.7 „Erforderlichkeit, Notwendigkeit“ in der Ausarbeitung „Datenschutzrechtliche Anforderungen an die medizinische Forschung unter Berücksichtigung der EU Datenschutz-Grundverordnung (DS-GVO)“, herausgegeben von der Arbeitsgruppe „Datenschutz und IT-Sicherheit im Gesundheitswesen“ der GMDS und der Arbeitsgruppe „Datenschutz und Datensicherheit im Gesundheits- und Sozialwesen“ der GDD. Online, zitiert am 2023-10-14; verfügbar unter http://ds-gvo.gesundheitsdatenschutz.org/html/forschung.php

GVO beschriebenen Ausnahmen – in erster Linie von der Verarbeitung personenbezogener Daten ab – was die Verarbeitung anonymer Daten nicht beinhaltet.

Allerdings muss der Verantwortliche, u. a. aufgrund der Rechenschaftspflicht nach Art. 5 Abs. 2 DS-GVO, zu jedem Zeitpunkt der Verarbeitung (und damit insbesondere auch während der gesamten Speicherdauer) nachweisen können, dass es sich bei den anonymisierten Daten zu jedem Zeitpunkt der Verarbeitung um anonyme Daten handelt. Der Nachweis der Anonymisierung stellt also keinen einmaligen Vorgang dar, sondern muss als Prozess betrachtet werden, welcher den gesamten Lebenszyklus der Daten begleitet.[51, 52]

Der EuGH legt diese in Art. 5 Abs. 2 DS-GVO verankerte Pflicht zum Nachweis der DS-GVO-konformen Verarbeitung sehr weit aus, auch im Sinne einer Beweislastumkehr[53, 54].

Aus diesem Grund, aber auch in Anbetracht sich ändernder technischer Möglichkeiten, muss dieser Nachweis der Anonymisierung mit einer entsprechenden Sorgfalt erfolgen und die dabei angewandte Methodik mindestens dem Stand der Technik entsprechen (siehe hierzu auch Kapitel 9.14). So bieten z. B. Bayes'sche Netzwerke heute Möglichkeiten der Zusammenführung von Daten, die noch vor wenigen Jahren so für nicht möglich gehalten wurden. D. h., Daten, welche 2016 noch als „anonym" angesehen wurden, müssen heute in Anbetracht moderner IT-Verfahren evtl. als „einer Person zuordenbar" und somit als „pseudonym" angesehen werden. Daher sind Anonymisierungsverfahren regelmäßig, spätestens aber bei neuen Erkenntnissen/Algorithmen/Techniken aus Mathematik und IT-Wissenschaft anlassbezogen zu evaluieren.

Weiterhin muss beachtet werden, dass „neue" Daten evtl. auch neue Möglichkeiten bieten. So kann es z. B. vorkommen, dass durch eine Unternehmensübernahme Daten hinzukommen[55], welche bei einer Zusammenführung eine Re-Identifikation ermöglichen. Ein Beispiel:

---

[51] EDSA (2021-02-02) EDPB Document on response to the request from the European Commission for clarifications on the consistent application of the GDPR, focusing on health research. Rn. 47. Online, zitiert am 2023-11-22; verfügbar unter https://edpb.europa.eu/our-work-tools/our-documents/other-guidance/edpb-document-response-request-european-commission_en

[52] Siehe z. B. Gerichtsurteile:

- EuGH Urt. v. 2022-02-24, Az. C-175/20, Rn. 81 i. V. m. Rn. 77. Online, zitiert am 2023-12-12; verfügbar unter https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A62020CJ0175
- BverwG Urt. v. 2022-03-02, Az. 6 C 7.20, Rn. 49-51. Online, zitiert am 2023-12-12; verfügbar unter https://www.bverwg.de/020322U6C7.20.0
- OLG Stuttgart Urt. v. 2023-11-22, AZ. 4 U 20/23, Rn. 395-400. Online, zitiert am 2023-12-12; verfügbar unter https://www.landesrecht-bw.de/bsbw/document/JURE235012226

[53] Siehe z. B. EuGH, Urt. v. 2022-02-24, Az. C-175/20, Rn. 81 i. V. m. Rn. 77. Online, zitiert am 2023-10-14; verfügbar unter https://dejure.org/2022,3279 bzw. Volltext unter https://curia.europa.eu/juris/document/document.jsf?text=&docid=254583&pageIndex=0&doclang=DE&mode=req&dir=&occ=first&part=1

[54] EuGH, Urt. v. 2023-12-21, Az. C-667/21. Rn. 103. Online, zitiert am 2023-12-29; verfügbar unter https://eur-lex.europa.eu/legal-content/de/TXT/?uri=CELEX:62021CJ0667

[55] Das Bundesverwaltungsgericht in Österreich stellte fest, dass im Falle einer Gesamtrechtsnachfolge keine Weitergabe personenbezogener Daten an Dritte stattfindet und der Rechtsnachfolger das Recht zur Datenverwendung in jenem Umfang übernimmt, wie es bereits dem Rechtsvorgänger zustand. Bescheid v. 2023-08-22, AZ W137 2251172-1. Online, zitiert am 2023-10-14; verfügbar unter https://www.ris.bka.gv.at/Dokument.wxe?Abfrage=Bvwg&Entscheidungsart=Undefined&SucheNachRechtssatz=True&SucheNachText=True&GZ=W137+2251172-1&VonDatum=01.01.2014&BisDatum=11.10.2023&Norm=&ImRisSeitVonDatum=&ImRisSeitBisDatum=&ImRisSeit=Undefined&ImRisSeitForRemotion=Undefined&ResultPageSize=100&Suchworte=&Position=1&SkipToDocu

Ein amerikanischer Dienstleister, dem unter dem EU-US Privacy Shield Daten im Rahmen einer Auftragsverarbeitung überlassen wurden, kauft einen anderen, europäischen Dienstleister auf, sodass jetzt auch diese vormals „europäischen“ Daten zusätzlich von dem amerikanischen Dienstleister verarbeitet werden.

Das EU-U.S. Data Privacy Framework schließt Verarbeitungen zu Zwecken des amerikanischen Auftragsverarbeiters nicht aus, schränkt diese jedoch ein. So findet sich im Kapitel „Data Privacy Framework Principles“ unter Abschnitt 5 „Data Integrity and Purpose Limitation“[56] zwar die Vorgabe, dass ein Unternehmen personenbezogene Daten nicht in einer Weise verarbeiten darf, die mit den Zwecken unvereinbar ist, jedoch ist unter Fußnote 6 zu lesen: „[…] können Beispiele für vereinbare Verarbeitungszwecke solche sein, die vernünftigerweise den Kundenbeziehungen, der Einhaltung von Vorschriften und rechtlichen Erwägungen, der Rechnungsprüfung, der Sicherheit und der Betrugsprävention, der Wahrung oder Verteidigung der gesetzlichen Rechte der Organisation oder anderen Zwecken dienen, die mit den Erwartungen einer vernünftigen Person angesichts des Kontextes der Datensammlung übereinstimmen.“

So können amerikanische Dienstleister die Daten eines europäischen Verantwortlichen z. B. für eigene Zwecke der Sicherheit oder Betrugsverarbeitung verarbeiten. Auch ist die Nutzung der Daten zu eigenen Forschungszwecken[57] möglich.

Die aus Art. 5 Abs. 2 DS-GVO resultierende Nachweispflicht fordert, dass der Nachweis erbracht wird, dass eine erfolgte Anonymisierung auch mit sich ändernden technischen Möglichkeiten oder mit neu erworbenem Zusatzwissen unumkehrbar ist. Eine regelmäßige Überprüfung des Vorhandenseins des immer noch richtigen Status „anonym“ bzw. „pseudonym“ sowie des davon abgeleiteten Schutzfaktors für die personenbezogenen Daten ist somit aus verschiedenen Gründen zwingend erforderlich und der dazugehörige Prozess muss – ggf. inkl. evtl. erforderlicher Kriterien, wann eine Prüfung spätestens erfolgen muss – dokumentiert sein, ebenso müssen die Überprüfungen selbst sowie die Ergebnisse nachvollziehbar dokumentiert werden.

Diese Nachweispflicht obliegt entsprechend den Vorgaben von Art. 5 DS-GVO dem Verantwortlichen, welcher die Daten für den ursprünglichen Zweck erhob, verarbeitete und anschließend für einen anderen Zweck anonymisierte. Die Verpflichtung ist auch vertraglich nicht delegierbar, da die DS-GVO diese Pflicht unmittelbar an den Verantwortlichen richtet. Gleichwohl kann der Verantwortliche einzelne Aufgaben an andere natürliche oder juristische Personen delegieren, wenn die Delegation dieser Aufgaben weiterhin unbestritten Aufgabe des Verantwortlichen bleibt.

---

mentPage=true&ResultFunctionToken=397ad383-d665-4713-9559-ad2fb1d604bd&Dokumentnummer=BVWGT_20230822_W137_2251172_1_00

[56] EU-U.S. Data Privacy Framework: 5. Data Integrity and Purpose Limitation. Online, zitiert am 2023-10-14; verfügbar unter https://www.dataprivacyframework.gov/s/article/5-DATA-INTEGRITY-AND-PURPOSE-LIMITATION-dpf?tabset-35584=2

[57] Siehe auch EU-U.S. Data Privacy Framework: 14. Pharmaceutical and Medical Products. Online, zitiert am 2023-10-14; verfügbar unter https://www.dataprivacyframework.gov/s/article/14-Pharmaceutical-and-Medical-Products-dpf?tabset-35584=2

5.3 Betroffenenrechte

5.3.1 Anonyme Daten und Betroffenenrechte

Entsprechend ErwGr. 26 gelten für anonyme Daten die Anforderungen der DS-GVO nicht. Die sich aus den Art. 12 bis 22 DS-GVO ergebenden Anforderungen an die Wahrung der Betroffenenrechten sind ebenfalls nicht einschlägig. Für die Zeit vor der Anonymisierung sind die jeweiligen Betroffenenrechte jedoch vollumfänglich zu beachten. So ist die betroffene Person bei einer Anonymisierung ggf. u. a. über den Anonymisierungsvorgang und – sofern vorhanden – die Zweckänderung bei der Verarbeitung der Daten zu informieren.

5.3.2 Pseudonyme Daten und Betroffenenrechte

Pseudonyme Daten gelten als personenbezogene Daten, daher gelten auch die Betroffenenrechte in vollem Umfang. Pseudonyme Daten weisen die Besonderheit auf, dass der Verantwortliche, der die pseudonymen Daten verarbeiten will, die betroffene Person nicht identifizieren kann. Nur der Verantwortliche, der die Daten ursprünglich personenbezogen erhoben und – idealerweise – dem anderen Verantwortlichen die pseudonymisierten Daten zur Verfügung stellte, sollte eine Re-Identifikation durchführen können. Jeder Auftragsverarbeiter gehört in die Sphäre eines Verantwortlichen, sodass die Zuordnung pseudonyme Daten oder nicht, vom jeweiligen Verantwortlichen „geerbt“ wird.

Die Tatsache, dass der die pseudonymen Daten verarbeitende Verantwortliche keine Kontaktmöglichkeit zur betroffenen Person besitzt, macht es notwendig, dass jegliche Kommunikation mit der betroffenen Person nur über den Verantwortlichen laufen kann, welcher die Daten (ursprünglich) erhoben hat, d. h. welcher die betroffenen Personen – ggf. mittels einer Zuordnungstabelle – identifizieren und somit auch kontaktieren kann.

5.3.3 Information bei Zweckänderung

Häufig werden Daten anonymisiert/pseudonymisiert, um diese dann für einen anderen Zweck als dem ursprünglichen zu verwenden. Gemäß Art. 13 Abs. 4 bzw. Art. 14 Abs. 4 DS-GVO muss der Verantwortliche vorher (insbesondere noch vor der Pseudonymisierung/Anonymisierung) der bzw. den betroffenen Personen Informationen über diesen weiteren Zweck und alle anderen maßgeblichen Informationen gemäß Art. 13 und/oder Art. 14 Abs. 2 DS-GVO zur Verfügung stellen.

5.4 Privacy by Design/Default

Entsprechend Art. 25 DS-GVO muss die Pseudonymisierung bzw. Anonymisierung über den vollständigen Lebenszyklus der Daten, d. h. von der Erhebung bis zur endgültigen Vernichtung (z. B. eine vollständige Löschung), aufrechterhalten werden. Änderungen in der technischen Entwicklung (Stand der Technik, ggf. auch Stand der Wissenschaft, siehe hierzu auch Kapitel 9.14) müssen während dieser Zeit betrachtet, hinsichtlich der Auswirkungen auf die pseudonymisierten/anonymisierten Daten bewertet und ggf. erforderliche Maßnahmen abgeleitet und umgesetzt werden[58].

[58] Hinweise bzgl. des Vorgehens hierzu finden sich in der Praxishilfe „Datenschutz durch Technikgestaltung und durch datenschutzfreundliche Voreinstellungen (Art. 25 DS-GVO)“. Online, zitiert am 2023-10-14; verfügbar unter https://www.gesundheitsdatenschutz.org/html/privacy_design_default.php

5.5 Datenschutz-Folgenabschätzung

Entsprechend Art. 35 Abs. 4 DS-GVO müssen Aufsichtsbehörden eine Liste mit Verarbeitungsvorgängen, für die eine Datenschutz-Folgenabschätzung erforderlich ist, veröffentlichen. Dementsprechend veröffentlichten auch die Aufsichtsbehörden des Bundes und der jeweiligen Bundesländer Listen mit Kriterien, wann eine Datenschutz-Folgenabschätzung erforderlich ist. Dabei unterscheiden die Aufsichtsbehörden zwischen Verarbeitungen von öffentlichen und nicht-öffentlichen Stellen.

2018 veröffentlichte die DSK eine abgestimmte Liste von Verarbeitungsvorgängen für den nicht-öffentlichen Bereich [59], in welcher festgelegt wurde, welche Verarbeitungen auf jeden Fall eine Datenschutz-Folgenabschätzung erfordern; diese Liste wurde inhaltlich von allen Aufsichtsbehörden übernommen. In dieser Liste findet sich:

Ziffer	Kriterium
15	Anonymisierung von besonderen personenbezogenen Daten nach Artikel 9 DS-GVO nicht nur in Einzelfällen (in Bezug auf die Zahl der betroffenen Personen und die Angaben je betroffener Person) zum Zweck der Übermittlung an Dritte

Als typisches Einsatzfeld für dieses Kriterium nennt die DSK „Anonymisierung von besonderen Arten personenbezogener Daten nach Artikel 9“, d. h. bei jeder Anonymisierung von genetischen oder Gesundheitsdaten wird eine Datenschutz-Folgenabschätzung erforderlich sein.

Für den öffentlichen Bereich finden sich Listen im Internet:

- Bundesbeauftragte für den Datenschutz und die Informationsfreiheit (BfDI) https://www.bfdi.bund.de/DE/Fachthemen/Inhalte/Technik/Datenschutz-Folgenabschaetzungen.html
- Bayerische Landesbeauftragte für den Datenschutz (BayLfD) https://www.datenschutz-bayern.de/datenschutzreform2018/DSFA_Blacklist.pdf
- Berliner Beauftragte für Datenschutz und Informationsfreiheit https://www.datenschutz-berlin.de/datenschutz/datenschutz-folgenabschaetzung/
- Landesbeauftragte für den Datenschutz und für das Recht auf Akteneinsicht Brandenburg https://www.lda.brandenburg.de/sixcms/media.php/9/DSFA_Muss_Liste_oeffentlich_21062019.pdf
- Landesbeauftragte für Datenschutz und Informationsfreiheit Bremen https://www.datenschutz.bremen.de/sixcms/media.php/13/Liste%20von%20Verarbeitungsvorg%C3%A4ngen%20nach%20Artikel%2035.pdf
- Hamburgische Beauftragte für Datenschutz und Informationsfreiheit der Freien und Hansestadt Hamburg https://datenschutz-hamburg.de/fileadmin/user_upload/HmbBfDI/Datenschutz/Informationen/Liste_Art_35-4_DSGVO_HmbBfDI-oeffentlicher_Bereich_v2.0a.pdf

[59] Datenschutzkonferenz (DSK): Liste von Verarbeitungsvorgängen nach Art. 35 Abs. 4 DS-GVO für den nicht-öffentlichen Bereich, Version 1.1. Online, zitiert am 2023-10-14; verfügbar unter https://www.datenschutzkonferenz-online.de/anwendungshinweise.html bzw. pdf-Datei unter https://www.datenschutzkonferenz-online.de/media/ah/20181017_ah_DSK_DSFA_Muss-Liste_Version_1.1_Deutsch.pdf

- Landesbeauftragter für Datenschutz und Informationsfreiheit Mecklenburg-Vorpommern https://www.datenschutz-mv.de/static/DS/Dateien/DS-GVO/HilfsmittelzurUmsetzung/MV-DSFA-Muss-Liste-Oeffentlicher-Bereich.pdf
- Landesbeauftragte für den Datenschutz Niedersachsen https://www.lfd.niedersachsen.de/startseite/datenschutzrecht/ds_gvo/liste_von_verarbeitungsvorgangen_nach_art_35_abs_4_ds_gvo/muss-listen-zur-datenschutz-folgenabschatzung-179663.html
- Landesbeauftragte für Datenschutz und Informationsfreiheit Nordrhein-Westfalen https://www.ldi.nrw.de/liste-von-verarbeitungsvorgaengen-nach-art-35-abs-4-ds-gvo-fuer-den-oeffentlichen-bereich
- Landesbeauftragte für den Datenschutz und die Informationsfreiheit Rheinland-Pfalz https://www.datenschutz.rlp.de/de/themenfelder-themen/datenschutz-grundverordnung/datenschutz-folgenabschaetzung/
- Landesbeauftragter für den Datenschutz Sachsen-Anhalt https://datenschutz.sachsen-anhalt.de/informationen/internationales/datenschutz-grundverordnung/liste-datenschutz-folgenabschaetzung
- Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein https://www.datenschutzzentrum.de/dsgvo/#dsfa
- Thüringer Landesbeauftragter für den Datenschutz und die Informationsfreiheit (TLfDI) https://www.tlfdi.de/fileadmin/tlfdi/datenschutz/dsfa_muss-liste_04_07_18.pdf

Die Listen beinhalten - direkt oder indirekt – Verarbeitungen, welche eine Anonymisierung und/oder Pseudonymisierung adressieren und somit eine Datenschutz-Folgenabschätzung vor Beginn der Verarbeitung verlangen.

Kriterium	Bundesland
Verarbeitung von Patientendaten (Ziff. 4, 7 und evtl. 9)	– Bund
Umfangreiche Verarbeitung inklusive Anonymisierung vertraulicher oder höchstpersönlicher Daten im Rahmen der amtlichen Statistik oder zum Zweck der Übermittlung an Dritte	– BayLfD
Die Verarbeitung von Daten gemäß Art. 9 Abs. 1 und 10 DS-GVO und von anderen Daten, die dem Sozial-, einem Berufs- oder besonderen Amtsgeheimnis unterliegen, soweit sie einem anderen Zweck als demjenigen dient, zu dem die Daten erhoben wurden.	– Berlin – Mecklenburg-Vorpommern – Rheinland-Pfalz
Umfangreiche Verarbeitung personenbezogener Daten im Rahmen der amtlichen Statistik, deren Erhebung, Speicherung und Verarbeitung, insbesondere der Anonymisierungs- oder Pseudonymisierungsprozesse und statistische Aufbereitung vor/für die Übermittlung der Informationen an Dritte.	– Brandenburg
Anonymisierung von besonderen Kategorien personenbezogener Daten nach Artikel 9 DSGVO nicht nur in Einzelfällen (in Bezug auf die Zahl der betroffenen Personen und die Angaben je betroffener Person) zum Zweck der Übermittlung an Dritte.	– Bremen – Hamburg
Anonymisierung von besonderen Kategorien personenbezogener Daten nach Art. 9 DS-GVO nicht nur in Einzelfällen (in Bezug auf die Zahl der betroffenen Personen und die Angaben je betroffener Person) zum Zweck der Übermittlung an Dritte.	– Sachsen-Anhalt

Kriterium	Bundesland
Umfangreiche Verarbeitung personenbezogener Daten im Rahmen der amtlichen Statistik, deren Erhebung, Speicherung und Verarbeitung, insbesondere der Anonymisierungsprozesse sowie deren Anonymisierung und statistische Aufbereitung vor/für die Übermittlung der Informationen an Dritte (Verarbeitung der personenbezogenen Daten im Rahmen der amtlichen Statistik).	– Niedersachsen
Umfangreiche Verarbeitung personenbezogener Daten im Rahmen der amtlichen Statistik, deren Erhebung, Speicherung und Verarbeitung, insbesondere der Anonymisierungsprozesse sowie deren Anonymisierung und statistische Aufbereitung vor/für die Übermittlung der Informationen an Dritte	– Nordrhein-Westfalen – Hamburg

Demnach erfordert die Anonymisierung von in Art. 9 Abs. 1 DS-GVO genannten Daten ggf. eine Datenschutz-Folgenabschätzung entsprechend Art. 35 DS-GVO, unabhängig davon, ob die Anonymisierung oder Pseudonymisierung von einer öffentlichen oder nicht-öffentlichen Stelle durchgeführt wird. Hinweise zum Umgang mit einer Datenschutz-Folgenabschätzung finden sich im Internet. [60]

[60] Hinweise bzgl. des Vorgehens hierzu finden sich in der Praxishilfe „Datenschutz-Folgenabschätzung gemäß Art. 35 DS-GVO“. Online, zitiert am 2023-10-14; verfügbar unter https://www.gesundheitsdatenschutz.org/html/dsfa.php

6 Sonderfall: Genetische Daten/Biomaterial

Bereits 2005 wurde auf der 27. Internationalen Konferenz der Beauftragten für Datenschutz und den Schutz der Privatsphäre in Montreux in Bezug auf genetische Daten Folgendes festgehalten[61]:

„Die Datenschutzbeauftragten

1. […]
7. sind sich bewusst, dass aufgrund des rasch wachsenden Kenntnisstandes im Bereich der Genetik Daten über die menschliche DNA zu den sensibelsten überhaupt werden können, und dass die Gewährleistung eines angemessenen rechtlichen Schutzes dieser Daten angesichts der beschleunigten Wissensentwicklung wachsende Bedeutung erlangt,
8. […]“

Aktuellere Untersuchungen haben ergeben, dass genetische Daten nicht als anonym angesehen werden können[62]. Genetische Daten können daher maximal pseudonymisiert werden bzw. bei entsprechender Anpassung bestenfalls als pseudonyme Daten angesehen werden. Entsprechend urteilte der EuGH in Bezug auf Fingerabdrücke, dass diese unter den Begriff eines personenbezogenen Datums fallen, da Fingerabdrücke „objektiv unverwechselbare Informationen über natürliche Personen enthalten und deren genaue Identifizierung ermöglichen“[63]; gleiches gilt für die Gensequenz eines Menschen.[64]

[61] 27. Konferenz vom 14. - 16. September 2005 in Montreux. Online, zitiert am 2023-10-14; verfügbar unter https://www.bfdi.bund.de/SharedDocs/Downloads/DE/GPA/2005_27GPA_MontreuxDSInGlobalisiertenWelt.html

[62] Siehe z. B.

- McGuire A, Gibbs RA (2006) No longer de-identified. Science 312: 370–371
- El Emam K. (2011) Methods for the de-identication of electronic health records for genomic research. Genome Medicine 3:25. Online, zitiert am 2023-10-14; verfügbar unter https://genomemedicine.biomedcentral.com/articles/10.1186/gm239
- El Emam K, Jonker E, Arbuckle L, Malin B (2011) A systematic review of re-identification attacks on health data. PLoS One 2011; 6: e28071. Online, zitiert am 2023-10-14; verfügbar unter https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3229505/
- Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y (2013) Identifying personal genomes by surname inference. Science 339: 321–324

Diskussion des Themas z. B. Hansson et al. (2016) The risk of re-identification versus the need to identify individuals in rare disease research. Eur J Hum Genet 24(11): 1553–1558. Online, zitiert am 2023-10-14; verfügbar unter https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5110051/

[63] EuGH, Urt. v. 17. Oktober 2013, Az. C-291/12, Rn. 27. Online, zitiert am 2023-11-01; verfügbar unter https://dejure.org/2013,27711 bzw. Volltext abrufbar unter https://curia.europa.eu/juris/document/document.jsf?text=&docid=143189&pageIndex=0&doclang=DE&mode=req&dir=&occ=first&part=1

[64] So auch EDSA (2021-02-02) EDPB Document on response to the request from the European Commission for clarifications on the consistent application of the GDPR, focusing on health research. Rn. 50, 51. Online, zitiert am 2023-11-22; verfügbar unter https://edpb.europa.eu/our-work-tools/our-documents/other-guidance/edpb-document-response-request-european-commission_en

6.1 Biomaterial und Einwilligung

Biomaterial enthält durch seinen genetischen Inhalt prinzipiell auch Daten, durch die ggf. Rückschlüsse auf etwaige dem Betroffenen verwandte Personen wie beispielsweise Kinder, Eltern und/oder Enkel möglich sind.[65] Unter den Vorgaben der DS-GVO können diese Daten mittels Einwilligung kaum verarbeitet werden, da eine Einwilligung immer nur für eigene Daten und deren Verarbeitung, nicht aber für die Daten Dritter gegeben werden kann.

Daraus folgt, dass wenn eine Person heute beispielsweise mittels einer Spende Biomaterial abgibt, ein heute noch nicht geborenes Enkelkind dieser Person aufgrund der Kenntnis dieses gespendeten Biomaterials durch eine heute noch gar nicht bekannte Erkrankung, die aber dem Erbgut innewohnt, diskriminiert werden kann. Auch diesen verwandten Personen muss ein „Recht auf Nichtwissen“ zugestanden werden, eine Einwilligung der spendenden Person kann dies nicht legitimieren.

Für die Verarbeitung derartiger personenbezogener bzw. personenbeziehbarer Daten ist daher ein gesetzlicher Erlaubnistatbestand erforderlich. Wenn seitens der Gesellschaft ein Konsens besteht, dass die Vorratsdatenspeicherung von Biomaterial mit recht allgemeiner Zweckbindung („medizinische Forschung“) gewünscht ist, muss somit ein Gesetz diese Forschung explizit erlauben, sodass eine Einwilligung zur Legitimierung der Verarbeitung nicht erforderlich ist und die Grundlage für die Rechtmäßigkeit der Verarbeitung die bereichsspezifische Norm ist. Denn eine Einwilligung alleine des Spenders kann die damit verbundene Verarbeitung von Informationen über Dritte (= genetisch Verwandte) enthaltenen Informationen nicht legitimieren.

[65] In einer Studie wurde schon 2018 gezeigt, dass etwa 60 % der weißen US-amerikanischen Bevölkerung durch eine DNA-Probe allein durch Nutzung von in öffentlicher Ahnenforschungsdatenbanken enthaltenen genetischen Daten Verwandter re-identifiziert werden könnten.
Ehrlich et. Al. (2018) Identity inference of genomic data using long-range familial searches. Science 365(6415): 690-694. https://doi.org/10.1126/science.aau4832

7 Exkurs: HIPAA und De-Identification – der amerikanische Weg

Ähnlich wie in Deutschland gilt in den Vereinigten Staaten von Amerika (USA) nicht nur Bundesrecht, sondern die einzelnen Mitgliedsstaaten haben eigene, souveräne Rechte. Entsprechend der amerikanischen Verfassung gelten Bundesgesetze und ebenso von den USA anerkannte und gegengezeichnete völkerrechtliche Verträge als höchste Rechtsquelle. Die amerikanische Verfassung enthält auch Beschränkungen für das Bundesrecht.

Alle fünfzig Bundesstaaten besitzen eine eigene Verfassung und das Recht, unabhängig vom Bundesrecht eigenes Recht in allen Bereichen zu schaffen, die nicht ausdrücklich durch die Verfassung an den Bund übergeben wurden. Das Datenschutzrecht ist überwiegend vom Recht der Bundesstaaten dominiert, sodass auch Fragen, was als „anonym“ oder „pseudonym“ anzusehen ist, vielfach vom jeweiligen Landesrecht beantwortet wird – wenn nicht aufgrund der amerikanischen Verfassung für einen bestimmten Bereich das Bundesrecht vorrangig gilt. In diesem Fall gelten in einem Bundesstaat je nach Anwendungsfall zwei verschiedene Interpretationen, jeweils voneinander abgegrenzt allein durch den Umstand, wann was anzuwenden ist.

Die Gesundheitsversorgung wird überwiegend vom amerikanischen Bundesrecht dominiert. Die wichtigsten Regelungen sind:

- Health Insurance Portability and Accountability Act of (HIPAA)
    - 1996 wurde HIPAA eingeführt, 2003 durch Privacy Rule umgesetzt,
    - 2009 durch Health Information Technology for Economic and Clinical Health Act (HITECH),
    - 2022 Datenschutzvorgaben durch Coronavirus Aid, Relief, and Economic Security Act (CARES Act) eingeschränkt
- Genetic Information Nondiscrimination Act (GINA); GINA ergänzt HIPAA Privacy Rule mit Regelungen zum Umgang mit genetischen Daten
    - 2008 in Kraft getreten
    - 2016 durch den Equal Employment Opportunity Commission (EEOC) überarbeitet, um die Zulässigkeit von Wellness-Programmen am Arbeitsplatz zu klären.

Hinsichtlich des Themas Anonymisierung/Pseudonymisierung in Bezug auf Patientendaten ist überwiegend HIPAA die maßgebliche Regelung, jedoch müssen ggf. datenschutzrechtliche Regelungen des jeweiligen Bundesstaates zusätzlich betrachtet werden, wenn Daten amerikanischer Patienten anonymisiert und pseudonymisiert werden sollen.

7.1 Health Insurance Portability and Accountability Act

Der amerikanische „Health Insurance Portability and Accountability Act[66]“ (HIPAA) kennt den Begriff der De-identification, also den Prozess, um Personen vor der Identifikation zu schützen. Das Vorgehen hierzu wird in Abschnitt 164.514 von HIPAA beschrieben[67]. Das U. S. Department of Health & Human Services veröffentlichte eine diesbezügliche Leitlinie, in welcher vorgegeben wird, wie De-Identifikation in Bezug auf medizinische Daten umgesetzt werden kann[68].

§ 16.4514(b) HIPAA enthält eine Spezifizierung bzgl. der Umsetzung der De-Identifikation und kennt zwei Methoden:

1) § 16.4514(b)(1)
   Eine Person mit angemessenem Wissen und Erfahrung bzgl. allgemein anerkannten statistischen und wissenschaftlichen Prinzipien und Methoden hinsichtlich einer De-Identifikation
   i. entscheidet, dass das angewandte Verfahren zur De-Identifikation nur sehr geringe Risiken auch in Kombination mit Informationen, die anderen Stellen (Dritten i. S. d. DS-GVO) zur Verfügung stehen, hinsichtlich der Re-Identifizierung der betroffenen Person(en) beinhaltet und
   ii. dokumentiert die angewandten Methoden und die Analyse bzgl. der Risiken.
2) § 16.4514(b)(2) (nach Angaben des U.S. Department of Health & Human Services[68])
   i. Es müssen nachfolgende Informationen von betroffenen Personen oder von Verwandten, Arbeitgebern oder Haushaltsmitgliedern entfernt werden:
      a) Namen
      b) Alle ortsbezogenen Angaben kleiner als ein Staat, einschließlich Straßenadresse, Stadt, Bezirks, Postleitzahl, außer den ersten 3 Ziffern der Postleitzahl
      c) Alle Datumsangaben, die direkt auf eine Person schließen lassen, außer dem Jahr
      d) Telefonnummern
      l) Fahrzeugnummern und Seriennummer, einschließlich Nummernschilder
      e) Faxnummern
      m) Geräte-Identifizierer sowie Seriennummern
      f) E-Mailadressen
      n) Web Universal Resource Locators (URLs)
      g) Sozialversicherungsnummern
      o) Internet Protocol (IP) Adressen
      h) Patienten-IDs aus elektronischen Patientenakten
      p) Biometrische Identifikatoren, Fingerprint und Stimme eingeschlossen
      l) Nummern der Leistungsempfänger der Krankenkasse

[66] U.S. Government Publishing Service: H. Rept. 104-736 - HEALTH INSURANCE PORTABILITY AND ACCOUNTABILITY ACT OF 1996. Online, zitiert am 2023-10-14; verfügbar unter https://www.gpo.gov/fdsys/search/pagedetails.action?granuleId=CRPT-104hrpt736&packageId=CRPT-104hrpt736 bzw. Zusammenfassung unter https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html

[67] §164.514 Other requirements relating to uses and disclosures of protected health information. Online, zitiert am 2023-10-14; verfügbar unter http://www.hipaasurvivalguide.com/hipaa-regulations/164-514.php

[68] U.S. Department of Health & Human Services: Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. Online, zitiert am 2023-10-14; verfügbar unter https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html

q) Fotografien, die das gesamte Gesicht zeigen, und alle vergleichbaren Bildnisse
j) Kontonummern, Abrechnungsnummern
k) Jede andere ein Individuum identifizierende Nummer, Merkmal oder Code, ausgenommen Angaben bzgl. § 16.4514(c) HIPAA (Angaben zur Re-Identifikation)
k) Nummern von Zeugnissen, Zertifikaten, Attesten, Führerscheinnummer

ii. Der oder die Verantwortliche/-n hat keine tatsächliche Kenntnis davon, dass die Informationen allein oder in Kombination mit anderen Informationen zur Identifizierung einer Person, die Gegenstand der Informationen ist, verwendet werden könnten.

§ 16.4514(c) HIPAA sieht die Möglichkeit einer Re-Identifikation vor. Dazu kann einer betroffenen Person im Rahmen der De-Identifikation ein Code oder ein anderes Identifizierungskennzeichen zugewiesen werden, um dadurch eine Möglichkeit zur Re-Identifikation zu erhalten.

Dieses wiederum hat zur Konsequenz, dass eine De-Identifikation nach HIPAA somit keine Anonymisierung im Sinne des europäischen Rechts darstellt. Eine De-Identifikation nach HIPAA kann jedoch den aus der DS-GVO resultierenden Anforderungen bezüglich einer Pseudonymisierung genügen, wenn der Verantwortliche keinen Zugriff auf die ggf. vorhandene Möglichkeit einer Re-Identifikation hat.

8 Hands on: Wie geht man vor?

Zunächst müssen die Daten überprüft und kategorisiert werden: Welche Art von Daten liegt vor?

Abbildung 1: Prüfschema, welche Art von Daten vorliegen

Liegen personenbezogene Daten vor, welche pseudonymisiert oder anonymisiert werden sollen, müssen als Erstes die direkten und indirekten identifizierenden Daten bestimmt werden; eine andere Bezeichnung für diese identifizierenden Daten ist „direkte Identifikatoren“ bzw. „Quasi-Identifikatoren“.

Dabei ist zu beachten, dass jede Pseudonymisierung und erst recht jede Anonymisierung von Merkmalsträgern auch immer mit einem Informationsverlust der zugehörigen Daten verbunden ist, da ja Daten des ursprünglichen Datensatzes verändert werden. Um diesen Verlust so gering wie möglich zu halten, können u. U. diejenigen Merkmalsträger schwächer pseudonymisiert/anonymisiert werden, welche einem geringeren Re-Identifikationsrisiko ausgesetzt sind. Inwieweit dies möglich ist, ohne den Schutz der Pseudonymisierung bzw. Anonymisierung zu verlieren, hängt zum einen vom Schutzbedarf der Daten selbst ab, zum anderen aber auch davon, ob die direkt und indirekt identifizierenden Daten unterschiedliche Risiken zur Re-Identifikation beinhalten und ob durch eine schwächere Pseudonymisierung/Anonymisierung dieser Daten ggf. die stärker geschützten Daten kompromittiert werden und so eine Re-Identifikation erleichtert wird.

8.1 Vorüberlegungen

8.1.1 Festlegung: Pseudonymisierung oder Anonymisierung?

Eine Pseudonymisierung soll eine Re-Identifizierung erschweren bzw. im Falle einer Anonymisierung vollständig verhindern. Hierzu werden direkte oder indirekte Merkmale in Datensätzen entfernt oder manipuliert, sodass nach Auffassung des Verantwortlichen eine Re-Identifikation nicht mehr möglich ist.

Bei einer Pseudonymisierung ist zu beachten, dass alleine durch die Vernichtung des Zuordnungsschlüssels/Pseudonyms Daten nicht automatisch zu anonymen Daten werden, da dies alleine keine ausreichende Gewähr für die Nicht-Identifizierbarkeit der Daten ist. Die Artikel-29

Datenschutzgruppe schrieb bereits 2014, dass zusätzliche Schritte unternommen werden sollten, bevor ein pseudonymer Datenbestand als anonymisiert betrachtet werden kann.[69]

Es ist daher vorab zu prüfen, ob eine Pseudonymisierung ausreichend ist oder eine Anonymisierung erforderlich (und auch möglich) ist. Eine Anonymisierung setzt die Anforderungen der DS-GVO weitestgehend außer Kraft (siehe Kapitel 5), schränkt die Aussagekraft von Auswertungen aber deutlich ein oder verringert die Aussagekraft ggf. sogar bis zur Bedeutungslosigkeit. Gerade im medizinischen Kontext seltener Erkrankungen wird dies eher die Regel als die Ausnahme sein. Andere medizinische Daten wie genetische Daten oder Erbkrankheiten können nicht anonymisiert werden, hier kann nur eine Pseudonymisierung erfolgen.

8.1.2 Pseudonymisierung durch denselben Verantwortlichen

Entsprechend ErwGr. 29 DS-GVO sind Pseudonymisierungsmaßnahmen bei demselben Verantwortlichen möglich. Dieser muss die erforderlichen technischen und organisatorischen Maßnahmen treffen, um für die jeweilige Verarbeitung zu gewährleisten, dass zusätzliche Informationen, mit denen die personenbezogenen Daten einer speziellen betroffenen Person zugeordnet werden können, seitens des Verantwortlichen im Rahmen der Verarbeitung nicht zugreifbar sind. Auch in diesem Fall müssen die in Art. 4 Ziff. 5 DS-GVO enthaltenen Vorgaben an eine Pseudonymisierung erfüllt werden. D. h., in Fällen der Pseudonymisierung durch denselben Verantwortlichen muss gewährleistet sein, dass:

a) Organisatorische Vorgaben existieren, in denen festlegt wird
   - welche Organisationseinheit innerhalb des Verantwortlichen die Pseudonymisierung durchführt,
   - welche Organisationseinheit den Zuordnungsschlüssel aufbewahrt (und vor Missbrauch schützt),
   - welche Organisationseinheit die pseudonymisierten Daten zu welchen Zwecken verarbeitet und
   - unter welchen Voraussetzungen eine Re-Identifizierung von betroffenen Personen möglich sein soll.

b) Die Zuordnungsschlüssel bzw. die Zuordnungsregel müssen entsprechend Art. 4 Ziff. 5 Hs. 2 DS-GVO gesondert aufbewahrt werden, um eine Nichtzuordnung der Informationen zu einer bestimmten Person zu gewährleisten. Dies beinhaltet, dass die Organisationseinheit, welche die pseudonymisierten Daten verarbeitet, keinen Zugriff auf die Zuordnungsregel bekommt.

c) In entsprechend eingesetzten informationstechnischen Systemen muss dies durch entsprechende Rechte- und Rollenkonzepte abgebildet werden.

d) Das Verfahren muss entsprechend dokumentiert werden, sowohl um der Nachweispflicht nach Art. 5 Abs. 2 DS-GVO zu genügen, aber auch im Rahmen der Dokumentationspflichten nach Art. 30 DS-GVO („Verzeichnis von Verarbeitungstätigkeiten“). ErwGr. 29 DS-GVO fordert ausdrücklich, dass hierbei die „befugten Personen“ bei diesem Verantwortlichen angegeben werden müssen. Dies dient u. a. auch dem Nachweis, dass die pseudonymen

[69] Artikel-29-Datenschutzgruppe (2014) Stellungnahme 5/2014 zu Anonymisierungstechniken. S. 26. Online, zitiert am 2023-10-14; verfügbar unter https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/index_en.htm#maincontentSec4

Daten verarbeitenden Personen andere sind als die Personen, die Zugriff auf den Zuordnungsschlüssel besitzen.

Um diesen Anforderungen zu genügen, wird oftmals ein sog. „Treuhänder“ (oder auch „Datentreuhänder“ genannt sowie als „Trust Center“ bezeichnet) eingesetzt.

8.1.3 Datentreuhänder: Rahmenbedingungen

Ein Datentreuhänder soll als neutraler Intermediär agieren, welcher die Interessen von Datengebenden (= Patient oder Proband) und Datennutzenden gleichermaßen wahren soll; wobei die Interessen sich z. T. sehr deutlich unterscheiden und nicht immer vom Datentreuhänder in Einklang gebracht werden können. Die Datengeber haben i. d. R. kein kommerzielles Interesse (ggf. aber ein Interesse an der Weiterentwicklung medizinischer Behandlungsmethoden), wohl aber ein Interesse am Datenschutz. Insbesondere in der medizinischen Forschung (z. B. in einem Klinikum), wenn Daten aus der Behandlung von Patienten ohne Einwilligung der jeweiligen Patienten für Forschungszwecke verarbeitet werden, ist von einem starken Interesse der Patienten an der Wahrung ihrer Privatgeheimnisse auszugehen. Ein Datentreuhänder, z. B. in der Person des Datenschutzbeauftragten, muss diese Interessen wahrnehmen und sehr genau prüfen, ob die Daten für die vorgesehenen Zwecke verarbeitet werden dürfen, ob Betroffenenrechte gewahrt werden usw.

In der „Datenstrategie der Bundesregierung“[70] definierte die Bundesregierung einen Datentreuhänder wie folgt:

> „Eine Datentreuhandstelle kann mit der Aufgabe betraut sein, einen standardisierten Zugang zu Daten für zugelassene Stellen zu entwickeln und umzusetzen. Zudem besitzen Datentreuhänder eine Beratungsfunktion gegenüber ihren Nutzerinnen und Nutzern und bieten je nach Spezialisierung verschiedene Dienste, wie beispielsweise die Verwaltung von Daten im Sinne der Nutzerinnen und Nutzer. Datentreuhänder können aber auch datenschutzrechtliche Interessen und Gestaltungsrechte für eine Vielzahl von Verbraucherinnen und Verbrauchern geltend machen.“

Datentreuhänder werden mitunter dafür eingesetzt, im Auftrag betroffener Personen zu entscheiden, wer Daten zu welchen Zwecken unter welchen Bedingungen erhalten darf, und werden dabei z. T. in ähnlicher Weise aktiv, wie die in § 26 TTDSG angeführte „anerkannten Dienste zur Einwilligungsverwaltung“ oder der in Art. 2 Ziff. 11 Verordnung (EU) 2022/868 benannte „Datenvermittlungsdienst“.

Grundsätzlich kann jede Person – natürliche, juristische, private oder öffentlich-rechtliche – die Rolle eines Datentreuhänders einnehmen, was sowohl entgeltlich oder unentgeltlich erfolgen kann. D. h., ein Datentreuhänder führt entweder einen Auftrag (unentgeltlich) oder eine Geschäftsbesorgung (entgeltlich) entsprechend §§ 662 ff BGB aus. Grundsätzlich kann man zwischen internem und externem Datentreuhänder unterscheiden.

Im Kontext der Pseudonymisierung/Anonymisierung werden in dieser Praxishilfe die Managementfunktionen eines Datentreuhänders betrachtet, nach welcher ein Datentreuhänder bei

[70] Datenstrategie der Bundesregierung, S. 110. (2021). Online, zitiert am 2023-11-14; verfügbar unter https://www.bundesregierung.de/breg-de/suche/datenstrategie-der-bundesregierung-1845632 bzw. pdf unter https://dserver.bundestag.de/btd/20/082/2008260.pdf

Bedarf auch die De-Identifikation durchführen und insbesondere die Zuordnungsregeln verwalten kann. Dabei agiert der Datentreuhänder in dieser Beziehung weisungsfrei, d. h. Verantwortliche erhalten keinen ungewollten Zugriff auf geschützte Daten, nur dem Datentreuhänder ist eine Re-Identifikation möglich.

8.1.4 Anonymisierung durch denselben Verantwortlichen

Eine Anonymisierung bei demselben Verantwortlichen ist grundsätzlich möglich, wenn gleichzeitig die Originaldateien[71] gelöscht werden. Solange ein binärer Vergleich zwischen den Originaldaten und den veränderten Daten ein identifizierendes Ergebnis liefern kann und somit eine Identifizierung von Personen grundsätzlich möglich ist, können die Daten nicht als „anonym" angesehen werden.

Etwas anderes kann gelten, wenn die Originaldaten zur Erstellung von synthetischen Daten genutzt werden. Wenn auch mit Hilfe der Originaldaten eine Person nicht mehr als identifizierbar anzusehen ist, gelten die Daten gemäß Art. 2 Ziff. 7 der Richtlinie (EU) 2019/1024 als anonym.

8.1.5 Zu beachtende Einflussfaktoren

8.1.5.1 Rechtliche Vorgaben

Teilweise existieren Vorgaben, wie zu verfahren ist. Beispielsweise schreibt § 27 Abs. 3 BDSG vor (Hervorhebung durch Autoren):

> „Ergänzend zu den in § 22 Absatz 2 genannten Maßnahmen sind zu wissenschaftlichen oder historischen Forschungszwecken oder zu statistischen Zwecken verarbeitete besondere Kategorien **personenbezogener Daten** im Sinne des Artikels 9 Absatz 1 der Verordnung (EU) 2016/679 **zu anonymisieren**, sobald dies nach dem Forschungs- oder Statistikzweck möglich ist, **es sei denn**, **berechtigte Interessen der betroffenen Person stehen dem entgegen**. Bis dahin sind die Merkmale gesondert zu speichern, mit denen Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren Person zugeordnet werden können. Sie dürfen mit den Einzelangaben nur zusammengeführt werden, soweit der Forschungs- oder Statistikzweck dies erfordert."

§ 27 Abs. 3 BDSG schreibt für wissenschaftliche oder historische Forschungszwecke eine Anonymisierung zwingend vor, außer wenn

a) der Forschungs- oder Statistikzweck mit anonymen Daten nicht erreicht werden kann oder
b) berechtigte Interessen der betroffenen Person einer Anonymisierung entgegenstehen.

Verzichtet man auf eine Anonymisierung, so kann dies nur basierend auf einem der beiden Gründe erfolgen und aufgrund der Nachweispflichten müssen die Gründe dargelegt und so dokumentiert werden, dass die Begründung alleine durch die Dokumentation auch von Dritten nachvollzogen werden kann.

[71] Die Originaldateien müssen tatsächlich gelöscht sein, d. h. es dürfen somit natürlich auch keine Backup-Dateien noch vorhanden sein, die eine Re-Identifikation ermöglichen würde. Entsprechendes muss in einem Löschkonzept beachtet werden. Siehe auch „Leitfaden für die Erstellung von Löschkonzepten im Gesundheitswesen" (Stand 2020-06-20) Online, zitiert am 2023-11-22; verfügbar unter https://gesundheitsdatenschutz.org/html/loeschkonzept.php

In § 40b Abs. 6 Ziff. 1 lit. b-e AMG findet sich die Pflicht zur Pseudonymisierung, wenn Daten einer klinischen Prüfung an den Sponsor[72], an die Behörde oder an die EU-Datenbank[73] weitergegeben werden. Entsprechend § 42a AMG müssen Daten pseudonymisiert werden, bevor eine Ethik-Kommission Daten erhalten darf.

Gleiches gilt für Studien zu Medizinprodukten. Gemäß § 29 Ziff. 1 lit. b-e MPDG müssen die Daten pseudonymisiert werden, bevor die Daten an einen Sponsor gegeben, Daten an den Hersteller zur Konformitätsbewertung übermittelt oder Daten im Falle von Vorkommnissen an Behörden übermittelt werden. Das BfArM darf Patientendaten entsprechend § 86 Abs. 2-4 MPDG nur anonymisiert erhalten; pseudonymisiert ausschließlich nur dann, wenn eine Anonymisierung nicht möglich ist.

Die Rechtsgrundlagen der Verarbeitung sind daher immer daraufhin zu prüfen, ob sie Vorgaben zur Pseudonymisierung oder Anonymisierung enthalten.

8.1.5.2 Zeitpunkt der Pseudonymisierung/Anonymisierung

Idealerweise werden Daten bereits pseudonym bzw. anonym erhoben. Wenn von vornherein feststeht, dass Daten z. B. für ein Forschungsvorhaben nur in pseudonymisierter oder anonymisierter Form verarbeitet werden, ist dies allein schon aus dem in Art. 5 Abs. 1 lit. c DS-GVO enthaltenem Grundsatz der Datenminimierung erforderlich. Aber auch aus dem Gedanken der Risikominimierung heraus ist ein frühestmöglicher Schutz personenbezogener Daten geboten.

Im medizinischen Kontext ist dies bei Daten der Patientenversorgung nicht möglich, da u. a. die gesetzliche Pflicht eine Dokumentation der Patientenbehandlung verlangt. Werden Daten nur zu medizinischer Forschung erhoben, kann ggf. aber schon bei der Erhebung der Daten entsprechend vorgegangen werden.

[72] Zu beachten: Seit Januar 2022 gilt die Verordnung EU) 536/2014 über klinische Prüfungen mit Humanarzneimitteln (https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A32014R0536), Art. 75 VO 536/2014 sieht ein Co-Sponsoring vor. Hier gilt die Pflicht zur Pseudonymisierung gegenüber allen Sponsoren.

[73] Verordnung (EU) Nr. 536/2014 des Europäischen Parlaments und des Rates vom 16. April 2014 über klinische Prüfungen mit Humanarzneimitteln und zur Aufhebung der Richtlinie 2001/20/EG Text von Bedeutung für den EWR, Art. 81 „EU-Datenbank“. Online, zitiert am 2023-11-14; verfügbar unter https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:32014R0536#d1e4039-1-1

Grundsätzlich lassen sich verschiedene Ansätze hinsichtlich des Zeitpunkts einer Anonymisierung/ Pseudonymisierung beobachten:

	Zweckorientiert	Anlassbezogen
Beispiel	- Datenerhebung für Forschungszwecke[74]	- Weitergabe Patientendaten an Forscher (Sekundärnutzung) - Publikation von Forschungsergebnissen - Aufforderung von Behörde wie z. B. BfArM - Übermittlung an med. Register - Datenweitergabe an andere Forschungseinrichtungen zu deren Forschungszwecken
Zeitpunkt	Bei Erhebung	Zeitpunkt der Entscheidung zur Weitergabe
Umfang	Vollständige Anonymisierung/ Pseudonymisierung aller Daten	Selektive Anonymisierung/ Pseudonymisierung entsprechend beabsichtigter Weitergabe

Tabelle 2: Zeitpunkt, wann eine Anonymisierung/Pseudonymisierung erfolgt

Da gerade im medizinischen Kontext regelmäßig eine Anonymisierung/Pseudonymisierung erforderlich ist, sollten Verantwortliche für ihren Bereich die dafür erforderlichen technischen und organisatorischen Maßnahmen bereits frühzeitig festlegen und die Voraussetzungen schaffen, dass zumindest die gesetzlich erforderlichen Vorgaben (z. B. Weitergabe an Krebsregister oder Übermittlung an Sponsor bei Forschung) ermöglicht werden.

8.1.5.3 Rücknahmefestigkeit

2007 schrieb die Artikel-29-Datenschutzgruppe: „Rücknehmbar pseudonymisierte Daten sind als indirekt bestimmbare Informationen über Personen anzusehen.“[75] Je höher die Rücknahmefestigkeit, desto unwahrscheinlicher ist von einer (unerwünschten) Umkehr einer Pseudonymisierung bzw. Anonymisierung auszugehen.

Die Rücknahmefestigkeit hängt natürlich in weiten Teilen von der gewählten Methode ab.

8.1.5.4 Anzahl der Personen im Datensatz

Der Grad der Wirksamkeit einer De-Identifikation hängt maßgeblich von der „Mächtigkeit der Menge, in der sich der Betroffene verbirgt“[76], ab. Die Anzahl der Personen bestimmt die Menge der Pseudonyme und auch in der heutigen Zeit gilt, dass die Rechenleistung zum Auffinden möglicher Verkettungsmöglichkeiten in einem Datensatz abhängig von der Anzahl der zu betrachtenden Datenmenge ist.

Die Artikel-29-Datenschutzgruppe führte daher schon 2007 aus: „Die Menge der möglichen Pseudonyme sollte so groß sein, dass bei zufälliger Auswahl nicht zweimal das gleiche Pseudonym gewählt wird.“[75]

---

[74] § 27 Abs. 3 BDSG schreibt für wissenschaftliche oder historische Forschungszwecke eine Anonymisierung bzw., falls eine Anonymisierung nicht möglich ist, Pseudonymisierung zwingend vor, daher ist schon bei Erhebung der Daten entsprechend vorzugehen.

[75] Artikel-29-Datenschutzgruppe. WP 136 „Stellungnahme 4/2007 zum Begriff 'personenbezogene Daten'“, S. 21. Online, zitiert am 2023-11-02; verfügbar unter http://ec.europa.eu/justice/article-29/documentation/opinion-recommen dation/files/2007/wp136_en.pdf

[76] Der Bayerische Landesbeauftragte für den Datenschutz: Arbeitspapier „Datenschutzfreundliche Technologien“, Abschnitt „Anonymisierung“. Online, zitiert am 2023-11-02; verfügbar unter https://www.datenschutz-bayern.de/technik/grundsatz/apdsft.htm#nr4

In einem kleineren Datensatz erkennt teilweise schon der Mensch auf den ersten Blick Zusammenhänge und kann ggf. eine Person identifizieren, in größeren Datenmengen braucht man eine entsprechende Rechenleistung. Heute kann man die erforderliche Rechenleistung für entsprechende Auswertungen in der Cloud anmieten, muss also kein eigenes Rechenzentrum dafür unterhalten, aber natürlich ist die Schutzwirkung einer Anonymisierung/Pseudonymisierung auch unter diesen Umständen weiterhin von der Anzahl der Personen im Datensatz abhängig.

8.1.5.5 Verkettungsmöglichkeit

Unter „Verkettungsmöglichkeit“ wird in diesem Kontext die Möglichkeit verstanden, vorhandene eigentlich nicht einander zugeordnete Merkmale in einem Datensatz einem einzelnen Akteur (dies kann z. B. eine Datensatz-ID sein) zuordnen zu können, wodurch u. U. weitergehende Aussagen über eine betroffene Person ermöglicht werden. Die Verkettungsmöglichkeit von einzelnen Transaktionen oder Datensätzen hat daher eine direkte Auswirkung auf das Risiko einer Re-Identifikation.

8.1.5.6 Konkrete Einzelangaben: Indirekt identifizierende Daten

Je konkreter die Einzelangaben in einem Datensatz sind, desto geringer kann die Wirksamkeit einer Anonymisierung/Pseudonymisierung sein. Beispielsweise ist die Angabe „Beruf/Amt = Bundeskanzlerin“ eindeutig einer Person zuordenbar, da es in Deutschland bis ins Jahr 2024 hinein nur eine Bundeskanzlerin gab. Entsprechende Werte müssen daher mit anderen Werten zusammengefasst werden, im Beispiel wäre die Angabe „Beruf/Amt = öffentlicher Dienst“ möglich, wenn hinreichend andere Personen im Datensatz vorhanden sind, auf welche die Angabe zutrifft.

Ist eine solche Anpassung nicht möglich, wird aus diesen indirekten identifizierenden Daten die jeweilige Person immer identifizierbar bleiben.

8.1.5.7 Verfügbarkeitsoptionen

Mitunter werden bei einer Pseudonymisierung Daten bestimmte Eigenschaften zugewiesen, welche erlauben, dass bestimmte, vorher festgelegte Eigenschaften eines Pseudonyms aufgedeckt werden können. Dies wird in der Literatur z. T. als „Verfügbarkeitsoption“ bezeichnet.[77]

Eine entsprechende Verfügbarkeitsoption kann z. B. die Zusammenführung von Daten aus unterschiedlichen Quellen ermöglichen („Verkettbarkeit“), wobei die Daten der verschiedenen Quellen trotz Pseudonymisierung demselben Pseudonym zugeordnet werden. Im Rahmen von Krebsregistern ist diese Möglichkeit beispielsweise die Grundvoraussetzung dafür, dass die im Krebsregister eingehenden Meldungen richtig zusammengeführt und doppelte Meldungen eliminiert werden können usw. Für andere medizinische Register gilt dies natürlich analog.

Eine weitere Verfügbarkeitsoption kann die Möglichkeit darstellen, dass ein Pseudonym unter bestimmten, zuvor festgelegten Voraussetzungen entschlüsselt und so eine Re-Identifizierung der betroffenen Person ermöglicht werden kann. Dies kann z. B. im Rahmen einer klinischen Studie wichtig sein, wenn Probanden über aus der Studie gewonnene Erkenntnisse, die für ihre Behandlung wichtig sind, informiert werden sollen/müssen.

---

[77] So z. B. im „Whitepaper zur Pseudonymisierung der Fokusgruppe Datenschutz der Plattform Sicherheit, Schutz und Vertrauen für Gesellschaft und Wirtschaft im Rahmen des Digital-Gipfels 2017“, S. 19 Kapitel „4.2.1. Verfügbarkeitsanforderungen“. Online, zitiert am 2023-11-02; verfügbar unter https://www.bmi.bund.de/SharedDocs/downloads/DE/veroeffentlichungen/2017/digital-gipfel-kurzfassung-one-pager-fokusgruppe.pdf?__blob=publicationFile&v=3

8.2 Identifizierung von direkten und indirekten identifizierenden Daten

Will man eine Pseudonymisierung bzw. Anonymisierung durchführen, müssen in einem ersten Schritt sowohl direkte als auch indirekte Identifikationsmerkmale im Datensatz identifiziert und bzgl. der Notwendigkeit der Änderung/Löschung hinsichtlich des Prozesses einer Pseudonymisierung oder Anonymisierung im Sinne des Datenschutzes bewertet werden.

Die zur Verarbeitung vorgesehenen Daten sind in drei Kategorien einzuteilen:

a) Direkte Identifikationsmerkmale: Alle Daten, welche eine direkte Identifizierung zulassen. Beispiele für direkte Identifikationsmerkmale sind insbesondere Namen (der bürgerliche Name sowie alle sonstigen Namen wie beispielsweise der Rufname oder der Chat Name), unter denen die Person bekannt ist.

b) Indirekte Identifikationsmerkmale: Alle Daten, welche in Verbindung mit anderem indirektem oder externem Wissen eine Identifikation potenziell ermöglichen. Beispiele für indirekte Identifikationsmerkmale sind:

- Personenbezeichner (z. B. Patienten-ID, Sozialversicherungsnummer, Steuernummer, Autokennzeichen, Kontonummer, Versicherungsnummer, Geburtsdatum)
- Erscheinungsmerkmale (z. B. Körpergröße, Haarfarbe, Kleidung, Tätowierungen)
- Biometrische Kennzeichen (z. B. Gesicht, Stimmprofile, Fingerabdrücke)
- Genetische Daten
- Digitale Zertifikate, welche eine Identifikationsmöglichkeit beinhalten (z. B. Zertifikate zur elektronischen Unterschrift)
- Identifikationsmerkmale basierend auf elektronischer Kommunikation (z B. Telefonnummer, Faxnummer, E-Mail-Adresse, IP-Adresse)
- Demographische Daten (z. B. Religion, Geburtsland, Muttersprache, Vorstrafen)
- Zuordnungsmerkale (z. B. Beruf, Funktion, Anschriften, Vorstrafen, Name der Mutter/des Vaters)
- Ausreißervariablen (z. B. seltene Diagnosen, Behandlungsbesonderheiten, körperliche Fehlbildungen, für die untersuchte Population untypische Merkmale).

Je nachdem, welche weiteren Informationen dem oder den Verantwortlichen zur Verfügung stehen, sind indirekte Identifikationsmerkmale ggf. als direkte Identifikationsmerkmale anzusehen. Typisches Beispiel hierfür sind Personenbezeichner wie beispielsweise die Patienten-ID in medizinischen Informationssystemen.

Zu beachten: Auch die Kombination von Merkmalen kann ein indirektes Identifikationsmerkmal darstellen.

> Beispiel: Nur etwa 1 % aller Patienten mit Brustkrebs weisen ein männliches Geschlecht auf[78]. Dies führt dazu, dass es Krankenhäuser gibt, die in 2 - 3 Jahren nur einen Mann mit der Diagnose „Brustkrebs“ behandeln.
>
> Die Angaben Geschlecht, Behandlungsjahr und Diagnose können unter diesen Umständen die Identifikation einer Einzelperson darstellen und sind folglich als indirektes Identifikationsmerkmal zu behandeln.

Dieser Umstand stellt insbesondere bei der Erforschung seltener Erkrankungen eine Herausforderung dar.[79] Es kann vorkommen, dass eine Anonymisierung in diesen Fällen nicht

[78] Deutsche Krebsgesellschaft: Brustkrebs bei Männern, Stand 2022-06-01. Online, zitiert am 2023-11-22; verfügbar unter https://www.krebsgesellschaft.de/onko-internetportal/basis-informationen-krebs/krebsarten/brustkrebs/brustkrebs-bei-maennern.html

möglich ist, ohne dass zur Anonymisierung Daten derart verändert werden müssten, dass eine sinnvolle Bearbeitung der Forschungsfrage nicht mehr möglich ist. Aber eine Pseudonymisierung ist i. d. R. möglich.

c) Nicht identifizierende Daten: alle anderen Daten, die weder direkte oder indirekte Identifikationsmerkmale darstellen.

Gerade bei der Analyse der sogenannten indirekten Identifikationsmerkmale ist häufig eine individuelle, kontextbezogene Betrachtung erforderlich. So kann bspw. in einem Fall die Haarfarbe ein indirektes Identifikationsmerkmal darstellen, durch die eine Person eindeutig identifizierbar wird. In anderen Fällen beinhaltet das Datum „Haarfarbe“ vielleicht keine Re-Identifikationsmöglichkeit. Desgleichen können medizinische Bilddaten neben den zur Identifizierung geeigneten Metadaten (z. B. DICOM StudyUID) auch in sich selbst eine Identifikationsmöglichkeit enthalten, z. B., wenn eine 3D-Rekonstruktion des Kopfes eine Gesichtserkennung ermöglichen würde. Es wird aktuell auch an automatisierten Verfahren zur Erkennung von indirekten Identifikationsmerkmalen gearbeitet[80], sodass in Zukunft vielleicht auch hier automatisierte Verfahren bei der Erkennung unterstützen können.

8.3 Arten von Pseudonymen und ihre Unterscheidungsmöglichkeiten

Pseudonyme können einerseits durch den Inhaber der Zuordnungsregel unterschieden werden:

a) Pseudonyme werden ausschließlich vom Betroffenen selbst vergeben. Ein Beispiel hierfür ist der „Nickname“ des Nutzers im Chat.

b) Pseudonyme werden vom ursprünglichen Verarbeiter vergeben, wie dies beispielsweise bei der IP-Adress-Vergabe durch einen Internet-Provider erfolgt.

c) Pseudonyme können von einem vertrauenswürdigen Dritten vergeben werden, wie dies beispielsweise häufig bei der Einschaltung einer sog. Trusted-Third-Party („Datentreuhänder“, siehe Abschnitt 8.1.3) in medizinischen Forschungsnetzen geschieht.

Eine weitere Unterscheidungsmöglichkeit bzgl. Pseudonyme besteht in der Art ihrer Generierung:

a) Das Pseudonym wird durch eine schlüsselabhängige Einweg- oder Hashfunktion aus invarianten Daten (z. B. Identitätsdaten) erzeugt.

b) Das Pseudonym wird (willkürlich) nach einem festen Einweg-Algorithmus vom Benutzer aus einem Geheimnis (z. B. Passphrase) erzeugt.

c) Das Pseudonym wird (zufällig) frei gewählt oder nach einem Zufallsverfahren erzeugt.

Eine dritte Unterscheidungsmöglichkeit bei Pseudonymen besteht in ihrer gesellschaftlichen Verwendung. Pseudonyme können als

- Personenpseudonyme, z. B.
    - Öffentliches Personenpseudonym (z. B. Telefonnummer)
    - Nichtöffentliches Personenpseudonym (z. B. Kontonummer, Probanden-ID in der medizinischen Forschung oder auch die Beschäftigten-ID)
    - Anonymes Personenpseudonym (z. B. Genom)

---

[79] Hansson et al. (2016) The risk of re-identification versus the need to identify individuals in rare disease research. Eur J Hum Genet 24(11): 1553–1558. Online, zitiert am 2023-10-14; verfügbar unter https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5110051/

[80] Z. B.: Jadhav PS, Borkar GM (2023) Quasi-identifier recognition with echo chamber optimization-based anonymization for privacy preservation of cloud storage. Concurrency Computat Pract Exper. 2023;e7906. https://doi.org/10.1002/cpe.7906

oder auch als

- Rollenpseudonyme, wie beispielsweise
  - Geschäftsbeziehungspseudonym (z. B. Chat- Name)
  - Transaktionspseudonym (z. B. PIN, TAN)

wie auch als

- Mehrfachpseudonyme, wie beispielsweise
  - Gruppenpseudonyme, wo ein Pseudonym von mehr als einer natürlichen Person genutzt wird oder
  - übertragbare Pseudonyme (wie z. B. Dynamische IP-Adressen

genutzt werden.

8.4 Methoden zur Pseudonymisierung/Anonymisierung

Sowohl bei der Pseudonymisierung als auch bei der Anonymisierung gibt es unterschiedliche Methoden, die im Einzelfall danach evaluiert werden sollten, wie mit ihnen am besten der individuell verfolgte Zweck erreicht werden kann.

8.4.1 Nichtangabe

Das zu schützende Datum wird bei dieser Methode nicht verwendet, sondern weggelassen, z. B. durch Löschung oder Nicht-Exportieren von Spalten einer Datenbanktabelle oder auch von Teilbereichen der Werte.

Beispiel: Die Daten aus Tabelle 1 wurden entpersonalisiert, indem Vorname und Nachname sowie Tag und Monat beim Geburtsdatum sowie die letzten Ziffern der ICD-Kodierung (Reduzierung auf einen 3-stelligen Wert) weggelassen wurden:

Geschlecht	Geb.-Datum	PLZ	ICD
w	1983	10115	C43
m	1965	10115	D22
m	1977	10178	C85
m	1981	10247	D44
m	1985	10319	C18
w	1987	10407	D46
m	1988	10435	C16
m	1968	10439	D46
w	1978	10585	C50
w	1969	10707	C91
m	1967	10717	D12
m	1991	10717	D12
w	1987	10787	C50
w	1983	10827	C50
w	1975	10963	C83

Tabelle 3: Entpersonalisierung von Daten durch Nutzung der Methode der Nichtangabe

Dabei muss beachtet werden, dass sich durch das Weglassen von Teilbereichen auch die Information selbst ändert. So kann z. B. die Reduzierung des ICD-Wertes im obigen Beispiel zu einer ggf. nicht unerheblichen Änderung der Diagnosen führen:

Original		Nichtangabe	
ICD	**Diagnose**	**Diagnose**	**ICD**
C16.3	Bösartige Neubildung: Antrum pyloricum	Bösartige Neubildung des Magens	C16
C18.4	Bösartige Neubildung: Colon transversum	Bösartige Neubildung des Kolons	C18
C50.1	Bösartige Neubildung: Zentraler Drüsenkörper der Brustdrüse	Bösartige Neubildung der Brustdrüse	C50
C50.3	Bösartige Neubildung unterer innerer Quadrant Brustdrüse	Bösartige Neubildung der Brustdrüse	C50
C83.0	Non-Hodgkin-Lymphom: Kleinzellig (diffus)	Nicht follikuläres Lymphom	C83
D12.6	Gutartige Neubildung: Kolon, nicht näher bezeichnet	Gutartige Neubildung des Kolons, des Rektums, des Analkanals und des Anus	D12
D12.8	Gutartige Neubildung: Rektum	Gutartige Neubildung des Kolons, des Rektums, des Analkanals und des Anus	D12

Tabelle 4: Änderung des Informationswertes einer Diagnose bei Änderung des ICD durch Nichtangabe

In diesem Zusammenhang ist es notwendig, die Nutzer bzw. Anwender darüber zu informieren, dass bestimmte Daten verändert wurden.

8.4.2 Maskierung/Ersetzung

Bei dieser Methode werden die zu schützenden Daten durch einen konstanten oder sich ändernden Wert, ein Zeichen oder eine Zeichenkette ersetzt.

Beispiel: Die Daten aus Tabelle 1 wurden maskiert, indem der Tag und der Monat des Datums jeweils auf „01“ geändert wurden, die Namen auf feste Zeichenkette unter Beibehaltung der Geschlechterzuordnung:

Vorname	Nachname	Geschlecht	Geb.-Datum	PLZ	ICD
Anne	Musterfrau	w	01.01.1983	10115	C43.9
Max	Mustermann	m	01.01.1965	10115	D22.9
Max	Mustermann	m	01.01.1977	10178	C85.9
Max	Mustermann	m	01.01.1981	10247	D44.8
Max	Mustermann	m	01.01.1985	10319	C18.4
Anne	Musterfrau	w	01.01.1987	10407	D46.1
Max	Mustermann	m	01.01.1988	10435	C16.3
Max	Mustermann	m	01.01.1968	10439	D46.2
Anne	Musterfrau	w	01.01.1978	10585	C50.3
Anne	Musterfrau	w	01.01.1969	10707	C91.10
Max	Mustermann	m	01.01.1967	10717	D12.8
Max	Mustermann	m	01.01.1991	10717	D12.6
Anne	Musterfrau	w	01.01.1987	10787	C50.8
Anne	Musterfrau	w	01.01.1983	10827	C50.1
Anne	Musterfrau	w	01.01.1975	10963	C83.0

Tabelle 5: Maskiertes Geburtsdatum

8.4.3 Mischung/Shuffeling

Bei der Nutzung dieser Methode werden die in den Datensätzen enthaltenen Werte vertauscht („verwürfelt“). Dabei ist zu beachten, dass etwaige Informationen, welche eine Person eindeutig identifizieren wie z. B. eine Telefonnummer oder eine Kreditkartennummer, zur Auflösung des Personenbezugs zusätzlich mit einer weiteren Methode verfremdet werden müssen, um einen Personenbezug ausschließen zu können.

Die Grundlage für diese Durchmischung sollte eine Zufallsverteilung sein, die jedem Datenfeld die Daten bzw. die Teilmenge der Daten eines anderen Datenfeldes zuordnet, wodurch letztlich ein neuer Datensatz entsteht. Bei einer zufälligen Vertauschung kann grundsätzlich nicht ausgeschlossen werden, dass ein Datensatz auf sich selbst abgebildet wird, sodass im Ergebnis keine Veränderung stattfindet. Dies muss natürlich durch entsprechende Vorkehrungen und geeignete Maßnahmen ausgeschlossen werden.

Beispiel: Die Daten aus Tabelle 1 werden untereinander vermischt, um so die Identifizierung auszuschließen:

Original

Vorname	Nachname	Geb.-Datum	ICD
Käthe	Albers	27.05.1975	C83.0
Frieda	Fischer	15.11.1987	C50.8
Kunigunde	Gewaltig	21.01.1969	C91.10
Franz	Herrlich	17.11.1967	D12.8
Gerfriede	Jensen	23.07.1983	C50.1
Berthold	Koch	28.08.1991	D12.6
Michael	Matuschek	13.04.1968	D46.2
Hugo-Egon	Meyer	27.08.1977	C85.9
Uwe	Müller	31.03.1988	C16.3

Mischung

Vorname	Nachname	Geb.-Datum	ICD
Uwe	Albers	28.08.1991	C16.3
Michael	Fischer	31.03.1988	C18.4
Kunigunde	Gewaltig	15.11.1987	C43.9
Käthe	Herrlich	15.07.1987	C50.1
Jürgen	Jensen	29.11.1985	C50.3
Jan	Koch	23.07.1983	C50.8
Hugo-Egon	Matuschek	11.05.1983	C83.0
Hiltrud	Meyer	23.12.1981	C85.9
Heike	Müller	01.04.1978	C91.10

Original

Vorname	Nachname	Geb.-Datum	ICD
Hiltrud	Niemand	15.07.1987	D46.1
Heike	Richter	11.05.1983	C43.9
Anke	Schmidt	01.04.1978	C50.3
Eckbert	Schneider	23.12.1981	D44.8
Jan	Schröder	03.12.1965	D22.9
Jürgen	Stillstand	29.11.1985	C18.4

Mischung

Vorname	Nachname	Geb.-Datum	ICD
Gerfriede	Niemand	27.08.1977	D12.6
Franz	Richter	27.05.1975	D12.8
Frieda	Schmidt	21.01.1969	D22.9
Eckbert	Schneider	13.04.1968	D44.8
Berthold	Schröder	17.11.1967	D46.1
Anke	Stillstand	03.12.1965	D46.2

Tabelle 6: Vermischung der Datensätze, sodass eine Identifizierung nicht möglich ist

Wie das Ergebnis der Mischung in Tabelle 6 zeigt, bleibt dabei zwar jeder Wert erhalten, sodass sich Mittelwert und Median z. B. des ICD nicht ändern, jedoch können sich durch eine Zufallsverteilung die Zuordnungen und damit Zusammenhänge ändern. Brustkrebs (C50) ist bei Männern eine seltene Erkrankung, durch die Mischung tritt Brustkrebs in der Verteilung jedoch häufiger bei Männern auf. Diese Effekte müssen vor dem Einsatz dieser Technik berücksichtigt werden. In der Medizin ist es aufgrund der geplanten Auswertungen häufig sinnvoller, anstelle einer reinen Mischung „Differential

Privacy“ (siehe Kapitel 8.4.8) zu verwenden, da bei Differential Privacy die statistischen Aussagen des ursprünglichen Datensatzes erhalten bleiben.

8.4.4 Varianzmethode

Bei dieser Methode werden Daten, die auf Zahlen basieren, dadurch verfremdet, dass die Zahlenwerte in festgelegten, zufällig erhöhten oder verringerten Streuungsintervallen verändert werden.

Beispiel: Das Geburtsdatum aus Tabelle 1 wurde mittels der Varianzmethode bearbeitet, wodurch das Geburtsdatum willkürlich verändert wurde, die statistische Aussage der Tabelle jedoch erhalten blieb:

Vorname	Nachname	Geschlecht	Geb.-Datum
Heike	Richter	w	11.05.1983
Jan	Schröder	m	03.12.1965
Hugo-Egon	Meyer	m	27.08.1977
Eckbert	Schneider	m	23.12.1981
Jürgen	Stillstand	m	29.11.1985
Hiltrud	Niemand	w	15.07.1987
Uwe	Müller	m	31.03.1988
Michael	Matuschek	m	13.04.1968
Anke	Schmidt	w	01.04.1978
Kunigunde	Gewaltig	w	21.01.1969
Franz	Herrlich	m	17.11.1967
Berthold	Koch	m	28.08.1991
Frieda	Fischer	w	15.11.1987
Gerfriede	Jensen	w	23.07.1983
Käthe	Albers	w	27.05.1975

Geb.-Datum
14.05.1983
05.12.1965
29.08.1977
21.12.1981
30.11.1985
16.07.1987
03.04.1988
18.04.1968
29.03.1978
17.01.1969
22.11.1967
01.09.1991
18.11.1987
26.07.1983
30.05.1975

Tabelle 7: Anpassung des Geburtsdatums durch die Varianzmethode

8.4.5 Kryptografische Methoden

Hierbei kommen Verschlüsselungs- und/oder Hash-Algorithmen zum Einsatz. Dabei ist zu beachten, dass kryptografische Eigenschaften wie Blocklänge, Ausgabealphabet und Kollisionen der jeweils verwendeten Methoden Auswirkungen auf das Ergebnis der Anonymisierung haben. Weiterhin ist zu berücksichtigen, dass hier kryptografische Methoden im speziellen Kontext der Anonymisierung bzw. Pseudonymisierung betrachtet werden, d. h. einige Betrachtungen im anderen Kontext ggf. zu anderen Ergebnissen führen können.

8.4.5.1 Rahmenbedingungen abklären

Mit dem Verantwortlichen, welcher die pseudonymisierten oder anonymisierten Daten verarbeiten will, müssen die Rahmenbedingungen geklärt werden. So muss z. B. abgesprochen werden, ob

- der Zeichensatz (arabisch, deutsch, …),
- die Zeichenart (numerisch, Buchstabenerhalt, Sonderzeichen bleibt Sonderzeichen),
- Zeichenlänge

bei der Pseudonymisierung/Anonymisierung erhalten bleiben sollen?
Weiterhin können funktionelle Anforderungen wie z. B.

- Kollisionsfreiheit; d. h. unterschiedliche Eingaben führen immer zu unterschiedlichen Ergebnissen, so dass die Unterschiede erhalten bleiben und ggf. Datensätze trotz Pseudonymisierung/Anonymisierung zusammengeführt werden können,
- Eindeutigkeit; gleiche Eingaben führen immer zu gleichen Abbildungen,
- Erhalt der statistischen Verteilung

hinzukommen.

Grundsätzlich gilt: Je mehr Rahmenbedingungen an die Pseudonymisierung bzw. Anonymisierung seitens der für die Weiterverarbeitung verantwortlichen Personen gestellt werden, umso größer wird das Risiko der Re-Identifizierbarkeit durch statistische Analysen.

8.4.5.2 Verschlüsselungsverfahren

Moderne kryptografische Methoden sind nahezu ausschließlich Binärchiffren[81], die sich in Block- und Stromchiffren sowie in symmetrische und asymmetrische Verfahren unterteilen lassen. Hierbei ist Folgendes zu beachten:

1) Stromchiffren müssen zum Erhalt von Eigenschaften mehrfach denselben Schlüsselstrom verwenden, was die kryptografische Stärke der Verfahren abschwächt. Daher sind Stromchiffren für die Pseudonymisierung/Anonymisierung i. d. R. eher ungeeignet.
2) Den Vorteilen im Umgang mit dem Schlüsselmaterial stehen bei asymmetrischen Verfahren sehr hohe Performance-Einbußen und relativ große Chiffrat-Blöcke entgegen.

Binärchiffren erhalten weder den Zeichensatz noch die Zeichenart oder die Zeichenlänge, sind jedoch sowohl kollisionsfrei als auch eindeutig.

Andere Verschlüsselungsverfahren können Anforderungen bzgl. Zeichenart, Zeichensatz und Zeichenlänge ggf. erhalten. Dieses ist z. B. bei entsprechender Implementierung beim symmetrischen Verfahren „One-Time-Pad“ der Fall. Hier wiederum kann ggf. die Anforderung der Eindeutigkeit nicht mehr gegeben sein.

D. h., ob eine Verschlüsselung zur Anonymisierung/Pseudonymisierung genutzt werden kann, ist abhängig von den Anforderungen der Fachabteilung.

8.4.5.3 Hash-Funktionen

Hash-Funktionen (auch: kryptografische Checksumme oder Einwegfunktion genannt) bilden eine beliebig lange Eingabedatenmenge auf einen binären String fester Länge ab. Somit können auch Hash-Funktionen weder den Zeichensatz noch die Zeichenart oder die Zeichenlänge erhalten. Die Möglichkeit der Kollisionsfreiheit ist abhängig von der Ausgabelänge und dem Algorithmus. Bei MD5 beispielsweise ist nachgewiesen, dass Kollisionsfreiheit nicht gegeben ist. Die Anforderung der Eindeutigkeit wird von Hash-Funktionen gewährleistet.

8.4.5.4 Salt

„Salt“ (= „Salz“) bezeichnet in der Kryptografie eine zufällig gewählte Zeichenfolge, die vor der Verwendung als Eingabe einer Hash-Funktion an einen gegebenen Klartext angehängt wird, um die Entropie der Eingabe zu erhöhen, was letztlich zu einer höheren Streuung des Ergebnisses führt.

[81] Eine Einführung in das Thema „Kryptographie“ findet man z. B. im Buch: Ertel W, Löhmann E: Angewandte Kryptographie. Hanser-Verlag, 6. Auflage 2019. ISBN 978-3-446-46313-4. https://doi.org/10.3139/9783446457041

Hierdurch kann z. B. verhindert werden, dass Originaldaten beispielsweise mithilfe von Rainbow-Tabellen identifiziert werden können.

Stand heute wird eine Entropie von 100 Bit als resistent gegen Brute-Force Angriffe mit hohem Angriffspotenzial angesehen. D. h. der Wertebereich muss eine Mindeststreuung von 2^{100} bzw. 10^{30} haben. Zu jedem Datensatz sollte ein eigener Salt existieren, um den größtmöglichen Schutz zu erhalten. Werden alle Datensätze mit ein und derselben Zeichenfolge kombiniert, so wird dies als „Pepper“ bezeichnet.

Bei der Überprüfung eines Datums wird jedoch nicht jedes Mal ein neuer Salt erzeugt, da sich sonst der entstandene Hashwert von dem gespeicherten unterscheidet und somit der Wahrheitsgehalt der Information nicht überprüft werden kann. D. h., die Anforderung der Eindeutigkeit wäre nicht mehr gegeben. Deshalb wird – sofern die Eindeutigkeit eine einzuhaltende Anforderung darstellt – bei der Generierung der zur jeweiligen Information verwendete Salt zusammen mit dem entstandenen Hashwert gespeichert. Dabei müssen Salt und Hashwert natürlich voneinander getrennt aufbewahrt werden, der Salt unbedingt geheim gehalten werden, da ansonsten der Schutz abgeschwächt wird.

8.4.6 Was wird wann mit welcher Methode erreicht?

Methode	Anonymisierung	Pseudonymisierung
Nichtangabe	Nichtangabe sorgt immer für Anonymität bzgl. des betreffenden Datums	Keine Pseudonymisierung möglich
Maskierung/Ersetzung	Bei Vorgehen – Ersetzen mit gleichbleibendem Wert – Ersetzen mit sich erhöhendem Wert – Zufällige Mischung (Initialisierungsschlüssel vernichtet)	Verwendung – eines Schemas – von pseudozufälliger Ersetzung, z. B. schlüsselabhängige Ersetzung
Mischung/Shuffeling	Bei Vorgehen – Zufällige Mischung (Initialisierungsschlüssel vernichtet)	Verwendung von – pseudozufällige Ersetzung, z. B. schlüsselabhängige Ersetzung
Varianzmethode	Bei ausreichend großer Varianz	Vorgehen mit schlüsselabhängiger Abweichung; Schlüssel wird aufbewahrt
Kryptografische Methoden	Bei Vorgehen – Schlüssel wird vernichtet – Verfahren ist nicht invertierbar	Bei Vorgehen – Schlüssel wird sicher aufbewahrt

8.4.7 k-Anonymität

Direkte und indirekte Identifikationsmerkmale werden zu Gruppen mit gleichen Inhalten zusammengefasst, d. h. die Identifikationsmerkmale so verändert, dass die Merkmale zu Gruppen zusammengefasst werden können. Damit sind die hinter den Daten stehenden Individuen nicht mehr unterscheidbar, d. h. eine eindeutige Identifikation ist nicht mehr möglich.

Um k-Anonymität zu erreichen, können alle oben beschriebenen Methoden eingesetzt werden. Dabei gilt: Je größer die Gruppe, desto größer ist das Maß an Anonymität bzw. umso kleiner ist die Wahrscheinlichkeit als Angehöriger einer Gruppe mit bestimmten Merkmalen identifiziert zu werden.

Der Parameter k definiert bei der k-Anonymität die Mindestgröße der Gruppen. Er ist damit gleichzeitig das Maß der Anonymität. In einer Gruppe von k Individuen liegt die Wahrscheinlichkeit bei 1/k ein einzelnes Individuum korrekt zu identifizieren: Ein Datensatz erfüllt formal k-Anonymität, wenn die identifizierenden Merkmale jedes Datums von mindestens k Einträgen im Datensatz erfüllt werden[82].

In der Literatur wird ein Schwellwert von mindestens 5 angegeben[83], d. h.: Bei jeder Auswertung umfasst das Ergebnis zu jedem Zeitpunkt des Auswertungszeitraumes mindestens 5 Betroffene, sodass kein Rückschluss auf Einzelpersonen gegeben sein sollte. Kann eine Rückführbarkeit auf eine Personengruppe unter 5 Personen nicht ausgeschlossen werden, sind sowohl der Schwellwert als auch die Merkmale/Items für die jeweilige Auswertung so zu definieren, dass trotzdem der Identifikationsschutz gewährleistet ist.

Mit festen Werten von k muss jedoch grundsätzlich vorsichtig umgegangen werden. Z. B. muss bedacht werden, dass verschiedene Attribute miteinander in Zusammenhang stehen können und so durch deren Interaktion u. U. Aussagen über eine Einzelperson ermöglicht werden. Eine optimale k-Anonymität stellt ein NP-schweres Problem[84] dar[85], weswegen zur Ermittlung des k-Wertes heuristische Methoden[86] empfohlen werden. In der Literatur[87] wurde die Komplexität der Bestimmung von k durch folgende Funktion dargestellt:

$$k = f(D_A, A_A, C_A, G_A, G_B)$$

mit

D_A: Die Notwendigkeit des Datenschutzes für den Datenbesitzer (Datenschutz-Anforderung)
A_A: Das Erfordernis des Nutzers der anonymisierten Daten (Anwender-Anforderung)
C_A: Die Verpflichtung zur Einhaltung der Vertraulichkeit der Daten (Compliance-Anforderung)
G_A: Der Grad der Abstraktion der Daten
G_B: Der Grad der Beschränkungen, der sich aus der k-Anonymisierung ergibt

wodurch die Komplexität des Findens eines „best k“ anschaulich dargestellt wird.

---

[82] Sweeney L. (2002) -Anonymity: A Model for Protecting Privacy“. International. Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10 (05): 557–70. https://doi.org/10.1142/S0218488502001648

[83] Z. B. zu finden in: El Emam K, Dankar FK. (2008). Protecting privacy using k-anonymity. JAMIA 15(5): 627–637. https://doi.org/10.1197%2Fjamia.M2716

[84] Engl. „NP-hard“, durch fehlerhafte Übersetzung findet sich umgangssprachlich mitunter daher auch „NP-Schwer“. Ein algorithmisches Problem ist NP-schwer, wenn das Problem mindestens so schwer zu lösen ist, wie die Probleme der Komplexitätsklasse NP.

[85] Nachweis des Zutreffens auf -Anonymität zu finden in: Meyerson A, Williams R. (2004) On the complexity of optimal K-anonymity. Proceedings of the twenty-third ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems: 223–228. https://doi.org/10.1145/1055558.1055591

[86] Z. B.

- Bayardo RJ, Agrawal R. (2005) Data privacy through optimal k-anonymization. 21st International Conference on Data Engineering (ICDE'05): 217-228. http://dx.doi.org/10.1109/ICDE.2005.42
- Kenig B, Tassa T. (2012) A practical approximation algorithm for optimal k-anonymity. Data Min Knowl Disc (25): 134–168. https://doi.org/10.1007/s10618-011-0235-9

[87] Venkataramanan N, Shriram A.: Implementing k-Anonymization, Seite 54. In: Venkataramanan N, Shriram A.: Data Privacy. Principles and Practice. verlag Taylor & Francis Group, 1. Auflage 2017. ISBN: 978-1-4987-2104-2

Weiterhin ist zu beachten, dass k-Anonymität auch bei sehr guter Wahl von „k“ alleine noch keine Anonymität gewährleisten muss. Die Möglichkeit, über verknüpfte Aussagen korrespondierender Attribute ggf. Aussagen über Einzelpersonen geben zu können, wurde schon erwähnt. Daneben gibt es noch andere Angriffsmethoden, einige werden in Kapitel 8.5.2 vorgestellt. Daher sollte k-Anonymität immer von flankierenden Maßnahmen wie l-Diversität und t-Closeness (siehe Kapitel 8.5.6) begleitet werden.

8.4.8 Differential Privacy

Differential Privacy wurde eingeführt, um eine De-Anonymisierung von Daten mit Hilfe von „Linkage Attacken“ (siehe Abschnitt 8.5.2.5) zu verhindern.[88]

Bei Differential Privacy wird der Anonymisierungsprozess in einer Art und Weise randomisiert, dass die Änderungen an den jeweiligen einzelnen Daten keinen wesentlichen Einfluss auf statistische Auswertungen haben, d. h. die statistischen Aussagen des ursprünglichen Datensatzes erhalten bleiben.

Cynthia Dwork und Aaron Roth[89] führten ein Beispiel zur Verdeutlichung an:

> „Die Studienteilnehmer werden aufgefordert, anzugeben, ob sie die Fähigkeit P wie folgt besitzen oder nicht:
>
> 1. Werfen Sie eine Münze.
> 2. Wenn Zahl, dann wahrheitsgemäß antworten.
> 3. Bei Kopf wird eine zweite Münze geworfen und bei Kopf mit "Ja" und bei Zahl mit "Nein" beantwortet.
>
> Die „Privatheit“ ergibt sich aus der plausiblen Bestreitbarkeit eines jeden Ergebnisses; selbst wenn der Besitz der Fähigkeit P einem rechtswidrigen Verhalten entspricht, ist auch eine Antwort mit „Ja“ nicht belastend, da diese Antwort mit einer Wahrscheinlichkeit von mindestens 1/4 eintritt, unabhängig davon, ob der Befragte tatsächlich die Fähigkeit P besitzt oder nicht: Die erwartete Anzahl der „Ja“-Antworten ist das 1/4-fache der Anzahl der Teilnehmer, die die Eigenschaft P nicht haben, plus 3/4 der Anzahl, die die Eigenschaft P haben.“

Differential Privacy setzt voraus, dass einerseits eine genügende Datenmenge vorhanden ist, sodass sich auch bei Änderungen einzelner Daten die statistische Aussage nicht ändert, andererseits, dass die Verteilung der Daten selbst entsprechende Annahmen bzgl. der Unveränderheit der statistischen Aussage bei Veränderung einzelner Daten erlaubt. Die Höhe des Schutzes von Differential Privacy wird durch den Parameter ε bestimmt: Je kleiner ε gewählt wird, desto weniger Informationen werden über bestimmbare Personen preisgegeben, d. h., desto höher ist der Grad an Anonymität – aber zugleich wird es schwieriger, die Daten sinnvoll auszuwerten. Der Idealfall bzgl. Anonymität ist dementsprechend bei ε = 0 gegeben, jedoch müssen die Daten dazu so stark verrauscht werden, dass die Aussagekraft der Ergebnisse einer Auswertung nur noch sehr eingeschränkt möglich oder – im schlimmsten Fall – nicht mehr gegeben ist.

---

[88] Hinsichtlich anderer Angriffe siehe z. B.

- Sei Y, Okumura H, Ohsuga A. (2022) Re-Identification in Differentially Private Incomplete Datasets. IEEE Open Journal of the Computer Society (3): 62-42. https://doi.org/10.1109/OJCS.2022.3175999

[89] Dwork D, Roth A. (2014) The Algorithmic Foundations of Differential Privacy. Beispiel auf S. 15. Online, zitiert am 2023-10-14; verfügbar unter https://www.cis.upenn.edu/~aaroth/Papers/privacybook.pdf

Da Differential Privacy bestenfalls die statistischen Aussagen unverändert lässt, sind Aussagen für Randwerte – z. B. Patienten mit seltenen Erkrankungen in einem größeren Patientenkollektiv mit allen möglichen Erkrankungen – ggf. nicht oder nur sehr eingeschränkt möglich.

Soll Differential Privacy eingesetzt werden, müssen daher vorab folgende Punkte betrachtet werden:

1) Welchen Wert soll ε haben? Wird ε zu hoch gewählt, erfolgt die De-Identifikation ggf. unzureichend. Wird ε zu gering gewählt, sind die Ergebnisse einer Auswertung evtl. nicht mehr aussagekräftig.
2) Alle Einträge in der Datenbank müssen unabhängig voneinander sein. Existieren Abhängigkeiten, so werden diese Abhängigkeiten und damit die statistische Aussage dieser Werte beeinflusst, wenn ausgewählte Merkmalsträger zufällig geändert werden. Bestehen daher Abhängigkeiten, so müssen Änderungen diese Abhängigkeiten daher abbilden, um die statistische Aussage zu erhalten, was aber wiederum Angriffe ermöglicht und das Risiko einer Re-Identifikation erhöht.
3) Bei der Implementierung dürfen keine Seitenkanalangriffe ermöglicht werden. Ein Seitenkanalangriff, auch als Seitenkanalattacke bezeichnet, versucht, Korrelationen zwischen den Daten und dem verwendeten kryptografischen Schlüssel zu finden. Gelingt ein Seitenkanalangriff, so kann man Zugriff auf die Ursprungsdaten erhalten und somit ggf. auch eine Re-Identifikation durchführen.

Eine Übersicht und grobe Bewertung von bei Differential Privacy eingesetzten Techniken erstellten Huang[90] et al. 2022.

8.4.9 Beispiele bzgl. Vorgehen

Datentyp	Methode
Zahl	– Neuvergabe der letzten x Stellen (x = abhängig von den Zahlenwerten) – Ersetzen durch Zufallszahlen – Nutzung einer Varianz (z. B. ± x%) – Löschung
String	– Neuvergabe über Tabelle – Ersetzung durch feste Zeichenkette – Ersetzung durch feste Zeichenkette mit laufender Nummer zwecks Beibehaltung der Unterscheidbarkeit
Datum	– Setzen von Tag und Monat auf festen Wert – Setzen des Datums auf einen festen Wert
Postleitzahl	– Neuvergabe von mindestens den letzten 2 Stellen über Umsetzungstabelle – Ersetzen von mindestens den letzten beiden Stellen durch festen Wert – Ersetzen von mindestens den letzten beiden Stellen durch festen Zufallswert
E-Mail-Adresse	– Löschen – Ersetzen durch festen Dummy-Wert
Religion	– Löschen – Ersetzen durch festen Dummy-Wert

[90] Huang et al. (2022) Differential privacy: Review of improving utility through cryptography-based technologies. Concurrency Computat Pract Exper. 2023;35:e7565. https://doi.org/10.1002/ cpe.7565

Datentyp	Methode
Medizinische Code-Systeme wie ICD, OPS, usw.	– Verkürzen der Kodierung – Löschung

Tabelle 8: Beispiel bzgl. Ersetzen von Datentypen

8.4.10 Tool-Unterstützung

8.4.10.1 Software

Es stehen diverse Open Source-Implementierungen von Anonymisierungsalgorithmen zur Verfügung, welche Verantwortliche unterstützen können. Die nachfolgende Darstellung ist nur exemplarisch zu verstehen:

- University of Texas at Dallas: Anonymization ToolBox[91]
  Die „Anonymization ToolBox“ ist eine Java-Implementierung und steht für die Betriebssysteme Microsoft Windows und Linux zum Download zur Verfügung. Die Toolbox unterstützt diverse Methoden, darunter k-Anonymität (Mondrian Multidimensional k-Anonymity[92]), l-Diversität (Incognito[93]) und t-closeness (Incognito).
  Als Eingangs- und Ausgabedaten werden ausschließlich von CSV-Dateien unterstützt, die Verarbeitung der Daten erfolgt sequenziell, was die Anwendbarkeit der Software für bestimmte Projekte ggf. einschränkt.
- Cornell University: Cornell Anonymization Toolkit
  Ursprünglich auf SourceForge gehostet (https://sourceforge.net/projects/anony-toolkit/), inzwischen wie viele andere Projekte auch bei github zu finden (https://github.com/wanghaisheng/Cornell-Anonymization-Toolkit)
  Die Software stammt aus dem Jahr 2009 und wurde seitdem nicht aktualisiert. Das Tool bietet eine grafische Benutzeroberfläche, mit welcher die Daten interaktiv analysiert und anonymisiert werden können.

[91] University of Texas at Dallas: Anonymization ToolBox. Online, zitiert am 2023-11-02; verfügbar unter https://labs.utdallas.edu/dspl/software/anonymization-toolbox/

[92] LeFevre K, DeWitt DJ, Ramakrishnan R. (2006) Mondrian Multidimensional K-Anonymity. 22nd International Conference on Data Engineering: 25. https://doi.org/10.1109/ICDE.2006.101

[93] Siehe

- LeFevre K, DeWitt DJ, Ramakrishnan R. (2005) Incognito: Efficient Full-Domain K-Anonymity. SIGMOD Conference: 49-60. https://dl.acm.org/doi/10.1145/1066157.1066164
- Chakravorty et al. (2017) Efficient Anonymization with INCOGNITO - Framework & Algorithm. IEEE International Congress on Big Data: 39-48. https://doi.org/10.1109/BigDataCongress.2017.15

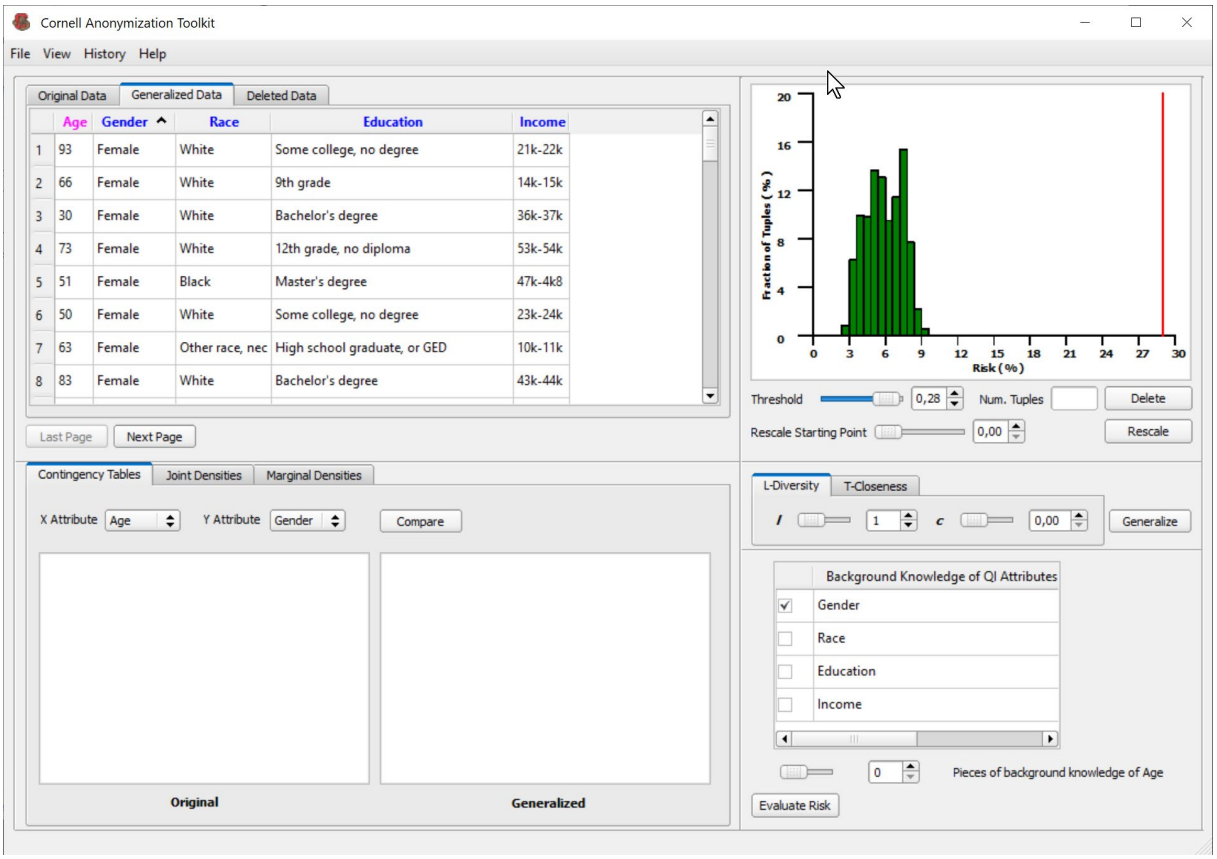

Abbildung 2: Cornell Anonymization Toolkit aus dem Jahr 2009 zur Risiko-Evaluation

- µ-ARGUS[94] wurde für statistische Daten entwickelt und bietet u. a. Perturbation (Randomisierung) sowie Möglichkeiten zur Erzeugung künstlicher Werte (Datensynthese). Die aktuelle Version 5.1.6 aus dem Jahr unterstützt die Betriebssysteme Microsoft Windows und Linux und benötigt das Java Runtime Environment (JRE) 8.
- medGAN: Dieses Tool generiert synthetische Patientendatensätze auf Basis realer Patientendatensätze, sowohl künstliche Texte wie auch künstliche Bilddaten[95] können damit erzeugt werden. Der Code wurde vor 2020 bei github (https://github.com/mp2893/medgan) eingestellt und seitdem nicht mehr aktualisiert.
- ARX – Data Anonymization Tool:[96] das Tool wurde ursprünglich von der Technischen Universität München (TUM) entwickelt, an der Fort- und Weiterentwicklung ist neben der TUM auch die Arbeitsgruppe Medizinische Informatik des Berliner Instituts für Gesundheitsforschung in der Charité beteiligt, die letzte Version stammt von November 2022.

  ARX verfügt über eine einfach zu bedienende Oberfläche, aber die von der Software implementierten Methoden sind komplex. ARX unterstützt u. a. Methoden wie beispielsweise k-anonymity, l-diversity, t-closeness, δ-disclosure privacy, β-likeness und δ-presence, die auch in dieser Praxishilfe erwähnt werden. Daneben werden auch andere Methoden zur Generalisierung wie z. B. Differential privacy angeboten. Was genutzt werden

[94] Statistical Disclosure Control: µ-ARGUS home page. Online, zitiert am 2023-11-02; verfügbar unter https://research.cbs.nl/casc/mu.htm

[95] Armanious et al. (2020) MedGAN: Medical image translation using GANs. Comput Med Imaging Graph: 101684. https://doi.org/10.1016/j.compmedimag.2019.101684

[96] ARX – Data Anonymization Tool. Online, zitiert am 2023-11-02; verfügbar unter https://arx.deidentifier.org/

sollte, ist natürlich kontextabhängig von der jeweils gewünschten späteren Auswertung der anonymisierten bzw. pseudonymisierten Daten, sodass neben einer entsprechenden Einarbeitungszeit auch entsprechende Kenntnisse über die dem Tool zugrundeliegenden statistischen Methoden erforderlich sind.

Das Tool bietet auch diverse Auswertungen an, z. statistische Veränderungen aufgrund der Datenveränderung (siehe Abbildung 3) oder auch eine Darstellung des Re-Identifikationsrisikos (Abbildung 4).

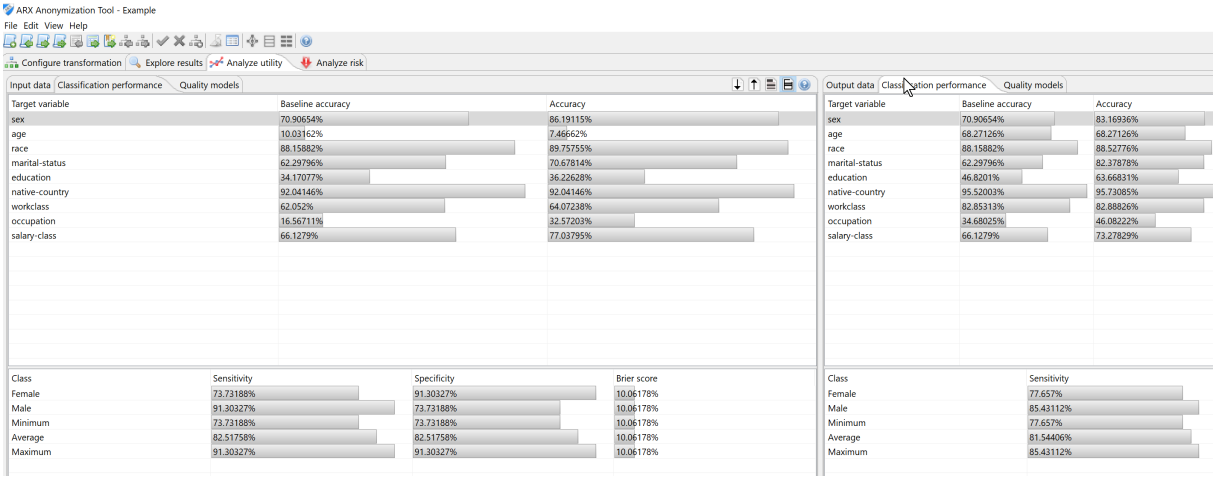

Abbildung 3: ARX Analyze utility: Classification performance

Abbildung 4: ARX Analyze risk: Attacher Models

ARX ist in der Lage, große Datenmengen auf Standard-Hardware zu verarbeiten und bietet eine intuitive, plattformübergreifende grafische Benutzeroberfläche. Die Software steht für die Betriebssysteme Microsoft Windows, Linux und MacOS zum Download zur Verfügung.

Wer das in Java programmierte Tool testen möchte, findet auf der Homepage sowohl einen Beispieldatensatz als auch eine ausführbare jar-Datei; eine Installation ist zum Ausprobieren des Tools nicht erforderlich.

8.4.10.2 Software-Bibliotheken

Zur Nutzung in eigenen Software-Entwicklungen stehen diverse Implementierungen von Algorithmen zur Verfügung. Dabei ist der Unterschied zwischen integrierbarer Bibliothek und integrierbares Tool (in einer Abgrenzung zu den in Kapitel 8.4.10.1 genannten Beispiele) nicht immer eindeutig möglich; die Entscheidung, wo etwas besser hinpasst, ist daher oft eine rein subjektive Entscheidung. Es existieren Implementierungen für diverse Programmiersprachen, nachfolgend nur einige Beispiele:

- anonymizer: Ein Python basiertes Tool, welches Textdateien überarbeitet. Über reguläre Ausdrücke können Masken zur De-Identifizierung beschrieben werden, ein Ersetzen von E-Mail-Adresse, Kreditkarten-Nummer, Telefonnummer ist dabei möglich, sodass Art der Daten in einer Auswertung berücksichtigt werden kann.
  URL: https://github.com/thoughtworks-datakind/anonymizer
- [Data]nymizer: Datanymizer, ein in Rust geschriebenes Projekt, liest Daten direkt aus einer Datenbank und anonymisiert Daten anhand der hinterlegten Regeln. Die Ausgabe ist ein anonymisierter SQL-Dump, der entweder in eine Datei oder direkt in die Standardausgabe geschrieben wird.
  URL: https://github.com/datanymizer/datanymizer
  Beschreibung: https://evrone.com/blog/datanymizer
- DICAT: Dabei handelt es sich um ein Tool, welches die De-Identifizierung von DICOM-Header-Daten ermöglicht. D. h., es geht um die Anonymisierung von Meta-Daten, die eigentlichen Bilddaten bleiben unberührt. Das Tool basiert auf der Python-Bibliothek PyDICOM[97].
  URL: https://github.com/aces/DICAT
- DicomAnonymizer: Das Python basierte Tool überarbeitet die Meta-Daten des DICOM-Formats, sodass in den Meta-Daten eine De-Identifizierung durchgeführt werden kann. Die Bilddaten selbst werden nicht verändert.
  URL: https://github.com/KitwareMedical/dicom-anonymizer
- Google Differential Privacy: google stellt ein Toolkit zur Implementierung von Differential Privacy, aufsetzend auf Apache Beam[98], ist daher in erster Linie für Machine Learning Projekte von Interesse.
  URL: https://github.com/google/differential-privacy
- Incognito-Algorithm: Eine Implementierung des Incognito-Algorithmus zur Umsetzung der k-Anonymität in Python.
  URL: https://github.com/Tolgahan20/Incognito-Algorithm
- Microsoft Tools for Health Data Anonymization: Diese Sammlung stellt die Skripte und Befehlszeilentools für die Anwendung in eigenen Projekten zur Verfügung. Es adressiert die Anonymisierung von Gesundheitsdaten vor Ort oder auch in der Cloud für sekundäre Zwecke wie Forschung, öffentliche Gesundheit usw. Derzeit unterstützt es die Anonymisierung von FHIR-Daten und DICOM-Daten.
  URL: https://github.com/microsoft/Tools-for-Health-Data-Anonymization
- Mondrian: Wie der Name andeutet handelt es sich um eine Implementierung des Algorithmus Mondrian Multidimensional k-Anonymity, Programmiersprache ist Python.
  URL: https://github.com/qiyuangong/Mondrian#mondrian

[97] Pydicom. Online, zitiert am 2023-11-02; verfügbar unter https://pydicom.github.io/
[98] Apache Beam Documentation. Online, zitiert am 2023-11-02; verfügbar unter https://beam.apache.org/documentation/

8.5 Darstellung des Risikos der Re-Identifizierung

Eine Re-Identifizierung kann Teil des geplanten Ablaufes sein, wenn eine Re-Identifikation unter zuvor festgelegten Bedingungen beabsichtigt erfolgt, z. B. für die Kontaktierung des Patienten, da die Verarbeitungsergebnisse Einfluss auf seine Behandlung haben.

Erfolgt eine Re-Identifikation als ungeplantes, insbesondere nicht beabsichtigtes Ereignis, kann eine derartige Re-Identifikation Risiken für die betroffene Person bergen.

8.5.1 Anonymisierung/Pseudonymisierung: Ein Rest-Risiko bleibt immer!?

Letztlich kann ein Verantwortlicher nur die für ihn zugreifbaren Daten auf Anonymität oder Pseudonymität prüfen, denn nur in den seltensten Fällen wird bekannt sein, über welches Zusatzwissen Angreifer verfügen können. Eine Re-Identifikation durch die Hinzuziehung neuer, zusätzlicher Datenquellen wird ein Verantwortlicher nur in den seltensten Fällen sicher ausschließen können.[99]

> Beispiel: Netflix veröffentlichte 2006 anonymisierte Daten seiner Nutzer[100], um einen Algorithmus finden zu lassen, der Netflix-Benutzern bessere Filmvorschläge unterbreitet. Netflix stellte hierfür 100.480.507 Datenbankeinträge mit Filmbewertungen von über 17.770 Filmen von insgesamt 480.189 Benutzern zur Verfügung. 2008 wurde gezeigt, dass diese anonymen Daten re-identifizierbar waren[101]: Die Forscher entwickelten einen Algorithmus, welcher die Daten mit den Bewertungen aus der öffentlichen Filmbewertungsdatenbank „Internet Movie Database“ abglich. Die Forscher konnten zeigen, dass bei acht (8) abgegebenen Bewertungen 99 % der Datenbankeinträge eindeutig zugeordnet werden konnten, lagen lediglich zwei (2) Bewertungen vor, betrug die Re-Identifikationswahrscheinlichkeit 68 %.

Hinsichtlich medizinischer Versorgung gibt es diverse frei verfügbare Quellen, die zusätzliches Wissen zur Re-Identifizierung bieten, beispielsweise

- finden sich in Bewertungsportalen Angaben, welche Patienten in welchem Zeitraum einen Aufenthalt in welchem Krankenhaus hatten,
- in Selbsthilfegruppen können Angaben gefunden werden, welche Personen welche Erkrankungen aufweisen.

Führt man beides zusammen (was durch heutige IT-Lösungen wie z. B. KI-Crawler sehr leicht möglich ist), können darüber Datenbanken zum Abgleich erstellt werden. Einige Firmen bieten die Aufschlüsselung des eigenen genetischen Datensatzes an, die Sequenzierung wird dabei z. T. unter Selbstkostenpreis angeboten: Das Geld wird damit verdient, dass die genetischen Daten Forschern verkauft werden, inklusive der Personendaten zwecks Kontaktaufnahme für Zwecke der

---

[99] So z. B. auch der europäische Datenschutzbeauftragte EDPS und die spanische Datenschutzaufsichtsbehörde Agencia Española de Protección de Datos (AEPD) im Fact-Sheet „10 misunderstandings related to Anonymisation“: Misunderstanding 5. […] Although a 100% anonymisation is the most desirable goal from a personal data protection perspective, in some cases it is not possible and a residual risk of reidentification must be considered.“ Online, zitiert am 2023-10-14; verfügbar unter https://edps.europa.eu/data-protection/our-work/publications/papers/aepd-edps-joint-paper-10-misunderstandings-related_en

[100] Wikipedia: Netflix Prize. Online, zitiert am 2023-11-02; verfügbar unter https://en.wikipedia.org/wiki/Netflix_Prize

[101] Narayanan A, Shmatikov V. (2008) Robust De-anonymization of Large Sparse Datasets. 2008 IEEE Symposium on Security and Privacy: 111-125. Online, zitiert am 2023-11-02; verfügbar unter https://doi.org/10.1109/SP.2008.33 bzw. pdf-Datei unter https://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf

Nacherhebung; der Vertrag, den die betroffenen Personen mit der jeweiligen Firma abschlossen, sieht dies ausdrücklich vor.[102] Auch sind genetische Daten nach Hackerangriffen auf entsprechende Firmen teilweise frei verfügbar im Internet/Darknet abrufbar.[103]

Was im Rahmen einer Risikobetrachtung jedoch häufig übersehen wird, ist, dass auch eine falsche Zuordnung von Daten zu einer Person ein Risiko für diese Person darstellen kann. Werden Daten fälschlicherweise einer Person zugeordnet, so kann ein Angreifer anhand dieser Daten eine Person „erkennen" und davon ausgehen, dass diese Daten zu dieser Person gehören. Diese falsche Zuordnung kann für die Person, deren Daten vielleicht im ursprünglichen Datensatz gar nicht enthalten waren, zu Problemen führen, z. B., weil ihr eine diskriminierende Erkrankung zugeschrieben und dies auch öffentlich bekannt gegeben wird.

Letztlich kann auch bei einer Anonymisierung i. d. R. nur die Wahrscheinlichkeit einer Re-Identifikation angegeben werden, d. h. ein gewisses Rest-Risiko für betroffene Personen, dass sie identifiziert werden können (oder ihnen aus einer Falsch-Erkennung Risiken erwachsen können), wird trotz Anonymisierung häufig bestehen bleiben. Bei der Betrachtung der Wahrscheinlichkeit einer Re-Identifikation muss ein Verantwortlicher jedoch alle Mittel berücksichtigen (siehe auch ErwGr. 26 DS-GVO), die von einer anderen (juristischen oder natürlichen) Person nach allgemeinem Ermessen wahrscheinlich genutzt werden könnte, um eine Re-Identifikation durchzuführen.

Hinsichtlich des Faktors „nach allgemeinem Ermessen wahrscheinlich genutzt werden" wird in der Literatur[104] angeführt, dass man mindestens zwei Aspekte bewerten muss:

1. Ist eine Re-Identifizierung für den potenziellen Angreifer rechtlich zulässig? Die Argumentation läuft dahingehend, dass eine Re-Identifikation nicht „vernünftigerweise"[105] zur Bestimmung einer natürlichen Person eingesetzt wird, wenn die Identifizierung der Person gesetzlich verboten ist.

---

[102] Siehe z. B. Bericht in
- Spektrum der Wissenschaft (2017) Wem gehören meine Gendaten? Online, zitiert am 2023-11-02; verfügbar unter https://www.spektrum.de/news/wem-gehoeren-meine-gendaten/1459381
- Deutschlandfunk (2019) Das Geschäft mit menschlichen Genen. Online, zitiert am 2023-11-02; verfügbar unter https://www.deutschlandfunkkultur.de/gendatenbanken-das-geschaeft-mit-menschlichen-genen-100.html
- Taz (2019) Gefährliches Wissen. Online, zitiert am 2023-11-02; verfügbar unter https://taz.de/Datenschutz-bei-Gentests/!5588812/

[103] Siehe z. B: Berichte aus Oktober 2023:
- Borns IT- und Windows-Blog (2023) Kundendaten von Genom-Analyseanbieter 23andMe im Netz aufgetaucht- Online, zitiert am 2023-11-02; verfügbar unter https://www.borncity.com/blog/2023/10/22/kundendaten-von-genom-analyseanbieter-23andme-im-netz-aufgetaucht/
- TechCrunch (2023) Hackers advertised 23andMe stolen data two months ago. Online, zitiert am 2023-11-02; verfügbar unter https://techcrunch.com/2023/10/10/hackers-advertised-23andme-stolen-data-two-months-ago/
- Heise online (2023) 23andme: Angeblich vier Millionen weitere Genanalyse-Daten geleakt. Online, zitiert am 2023-11-02; verfügbar unter https://www.heise.de/news/23andme-Angeblich-Genanalyse-Daten-aus-Grossbritannien-und-Deutschland-geleakt-9339051.html

[104] Gierschmann S. (2021) Gestaltungsmöglichkeiten durch systematisches und risikobasiertes Vorgehen – Was ist schon anonym? ZD: 482-486

[105] EuGH: Urt. v. 2016-10-19, Az. C-582/14. Online, zitiert am 2023-10-14; verfügbar unter https://dejure.org/2016,33959 bzw. Volltext unter https://curia.europa.eu/juris/document/document.jsf?text=&docid=184668&pageIndex=0&doclang=DE&mode=req&dir=&occ=first&part=1

Hier muss natürlich betrachtet werden, aus welchem Land ein potenzieller Angreifer heraus beabsichtigt, eine Re-Identifikation durchzuführen. Wenn deutsches Recht für den Angreifer nicht anwendbar ist, kann dies für den Angreifer auch keine Schranke darstellen; es muss daher immer das Recht des Landes – oder der Länder - betrachtet werden, aus welchem der Angreifer wahrscheinlich agieren würde.

2. Ist eine Re-Identifikation praktisch nicht durchführbar? Als ein Beispiel wird hier regelmäßig ein „unverhältnismäßiger Aufwand an Zeit, Kosten und Arbeitskräften“ angeführt, der eine praktische Durchführbarkeit einer Re-Identifikation verhindern würde.
   Auch hier muss die Beurteilung einer „Unverhältnismäßigkeit“ aus Sicht des Angreifers erfolgen. Eine persönliche Motivation kann beispielsweise aus Sicht eines Angreifers einen ganz anderen Einsatz an „Zeit, Kosten und Arbeitskräften“ rechtfertigen als eine Gewinnabsicht.

„Vernünftigerweise“ wird dabei aus Rn. 49 des Breyer-Urteils des EuGH abgeleitet, wobei der EuGH die Nutzung von „vernünftigerweise“ wiederum aus ErwGr. 26 Richtlinie 95/46 ableitet. In ErwGr. 26 DS-GVO findet sich statt „vernünftig“ der Begriff „wahrscheinlich"; ob angesichts der heute gängigen Cyber-Angriffe der Bezug von „rechtlich erlaubt“ unter dem Begriff „wahrscheinlich“ seitens EuGH weiterhin Bestand haben würde, ist sicherlich ein Diskussionspunkt.

ErwGr. 26 DS-GVO fordert hinsichtlich der Bewertung des Faktors „nach allgemeinem Ermessen wahrscheinlich genutzt werden“, dass alle „objektiven Faktoren“ herangezogen werden. Als zwei Beispiele für objektive Faktoren nennt ErwGr. 26 DS-GVO die Kosten der Identifizierung und der dafür erforderliche Zeitaufwand.

Letztlich muss also auch unter Berücksichtigung von ErwGr. 26 DS-GVO betrachtet werden, wie wahrscheinlich eine Identifikation betroffener Personen ist, sowohl bei der Bewertung der Wirksamkeit einer Pseudonymisierung als auch der einer Anonymisierung.

Diesen risikobasierten Ansatz vertreten auch die Aufsichtsbehörden in ihrer Leitlinie 04/2020[106] „für die Verwendung von Standortdaten und Tools zur Kontaktnachverfolgung im Zusammenhang mit dem Ausbruch von COVID-19“. In Rn. 15 der Leitlinie 04/2020 findet sich

> „Unter Anonymisierung ist die Verwendung einer Reihe von Techniken zu verstehen, sodass diese Daten nur mit einem unverhältnismäßig hohen Aufwand einer bestimmten oder bestimmbaren natürlichen Person zugeordnet werden können. Diese Verhältnismäßigkeitsprüfung („reasonability test“) muss sowohl objektive Aspekte (Zeit, technische Mittel) als auch kontextuelle Elemente berücksichtigen, die von Fall zu Fall variieren können (Seltenheit eines Phänomens, einschließlich Bevölkerungsdichte, Art und Umfang der Daten). Wenn die Daten diese Prüfung nicht bestehen, so wurden sie nicht anonymisiert und fallen daher in den Anwendungsbereich der DS-GVO.“

Rn. 16 Leitlinie 04/2020 ergänzt:

> „Für die Bewertung der Zuverlässigkeit des Anonymisierungsprozesses sind folgende drei Kriterien maßgeblich: (i) die datenbasierte Isolierung einer Einzelperson aus einer größeren

[106]EDSA: Leitlinien 04/2020 für die Verwendung von Standortdaten und Tools zur Kontaktnachverfolgung im Zusammenhang mit dem Ausbruch von COVID-19. Online, zitiert am 2023-10-14; verfügbar unter https://edpb.europa.eu/our-work-tools/our-documents/guidelines/ guidelines-042020-use-location-data-and-contact-tracing_de

Gruppe, (ii) die Verknüpfbarkeit zweier sich auf ein und dieselbe Person beziehender Datensätze und (iii) die mit hoher Genauigkeit erfolgende Herleitung bisher unbekannter Informationen über eine Einzelperson."

Die Bewertung der Wirksamkeit einer Pseudonymisierung wie auch einer Anonymisierung ist dabei kein einmaliges Ereignis, sondern muss für die gesamte Dauer der Verarbeitung erfolgen und entsprechend wiederholt erfolgen. ErwGr. 26 DS-GVO verlangt, dass „die zum Zeitpunkt der Verarbeitung verfügbare Technologie und technologische Entwicklungen" berücksichtigt werden müssen. Daraus folgt, dass für die gesamte Dauer der Verarbeitung anonymisierter Daten der Nachweis geführt werden muss, dass es sich um anonymisierte Daten handelt.

Auch gemäß den Anforderungen von Art. 25 DS-GVO sind sowohl „zum Zeitpunkt der Festlegung der Mittel für die Verarbeitung [also der Planung der Verarbeitung] als auch zum Zeitpunkt der eigentlichen Verarbeitung geeignete technische und organisatorische Maßnahmen" zu treffen, welche die Rechte der betroffenen Personen schützen. Weiterhin verlangt Art. 32 DS-GVO abhängig vom Risiko für die betroffenen Personen die Ergreifung geeigneter Maßnahmen zum Schutz der betroffenen Person.

D. h., es muss zwingend das Risiko, welches durch eine Re-Identifikation für die betroffene Person existieren kann, für die gesamte Verarbeitungsdauer der anonymisierten Daten beurteilt werden.[107]

8.5.2 Risikodarstellung: Bekannte Angriffsszenarien zur Re-Identifikation

Entscheidend für die Beurteilung des Risikos ist bereits das Vorhandensein der abstrakten Möglichkeit zur Identifikation von Betroffenen. Dabei sind die wesentlichen Risikofaktoren für eine Re-Identifizierung statistische Strukturen und Zusatzwissen. Selbst wenn Datensätze hoch effektiv und nach den aktuellen kryptologischen Methoden geschützt sind, können statistische Auffälligkeiten dazu führen, dass ein Personenbezug - ggf. auch nur teilweise - wiederhergestellt werden kann. Dieses Risiko wird durch Verfügbarkeit von Zusatzwissen erheblich verstärkt.

8.5.2.1 Aussondern („singling out")

Aussondern („singling out") wird in ErwGr. 26 S. 3 DS-GVO als ein Beispiel zur unerwünschten Re-Identifikation genannt, die Verantwortliche bei Pseudonymisierung oder auch Anonymisierung berücksichtigen müssen. Kann eine Person aus einem Datensatz ausgesondert (herausgefiltert) werden, so gilt diese Person als identifizierbar.

Können Einzelpersonen herausgefiltert werden, so kann der Datensatz nur als pseudonym, aber nicht als anonym gelten.

8.5.2.2 Inferenz

Ist eine Inferenz möglich, so können Attributwerte abgeleitet und einer Person zugeordnet werden. Dadurch kann es geschehen, dass durch die Zuordnung mehrerer Attribute letztlich genau eine Person mit diesen Attributen resultiert und somit eine Person identifizierbar ist. Wenn dies möglich ist, kann ein Datensatz bestenfalls nur als pseudonym, aber nicht als anonym im Sinne des Datenschutzes angesehen werden.

---

[107] Wissenschaftliche Ansätze zur berechnung der Wahrscheinlichkeit finden sich beispielsweise in: Rocher L, Hendrickx JM, de Montjoye Y. (2019) Estimating the success of re-identifications in incomplete datasets using generative models. Nat Commun 10: 3069. https://doi.org/10.1038/s41467-019-10933-3

8.5.2.3 Unsorted Matching Angriff

Bei der Durchführung einer Anonymisierung werden zwar einzelne, ausgewählte Werte verarbeitet, jedoch bleibt oftmals die Reihenfolge erhalten. Selbst wenn in einer Tabelle die Reihenfolge verändert wird, bleiben die Referenzwerte zu anderen Tabellen (Fremdschlüssel) unverändert und bieten Möglichkeiten, die ursprüngliche Reihenfolge wieder herzustellen.

Ein Beispiel: Es existiert eine Tabelle mit Patientendaten:

Vorname	Nachname	Geschlecht	Geb.-Datum	PLZ	Diagnose
Heike	Richter	W	11.05.1983	10115	Depression
Jan	Schröder	M	03.12.1965	10115	Delirium
Hugo-Egon	Meyer	M	27.08.1977	10178	Depression
Eckbert	Schneider	M	23.12.1981	10247	Wahnvorstellungen
Frederike		W	03.12.1997	10249	Wahnvorstellungen

Tabelle 9: Patientendaten für Beispiel bei einem Unsorted Matching Angriff

Die Tabelle wird in zwei Tabellen aufgeteilt:

Vorname	Nachname		Geschlecht	Geb.-Datum	PLZ	Diagnose
Heike	Richter	↔	W	11.05.1983	10115	Depression
Jan	Schröder	↔	M	03.12.1965	10115	Delirium
Hugo-Egon	Meyer	↔	M	27.08.1977	10178	Depression
Eckbert	Schneider	↔	M	23.12.1981	10247	Wahnvorstellungen
Frederike		↔	W	03.12.1997	10249	Wahnvorstellungen

Tabelle 10: Re-Identifikation durch einen Unsorted Matching Angriff

Wer Zugriff auf beide Tabellen hat, kann aufgrund der Bekanntheit der Reihenfolge die Personen identifizieren.

8.5.2.4 Komplementärveröffentlichung

Auch aus Komplementärveröffentlichungen, d. h., wenn Daten aus einem Datensatz für zwei oder mehr Veröffentlichungen genutzt werden, können sich Möglichkeiten für einen Angriff auf die Re-Identifikation bieten.

Beispiel: Die im ersten Beispiel dargestellten Patientendaten werden für zwei Publikationen genutzt:

In der ersten Publikation werden folgende Daten verwendet:

Geschlecht	Geb.-Datum	PLZ	Diagnose
W	11.05.1983	10115	Depression
M	03.12.1965	10115	Delirium
M	27.08.1977	10178	Depression
M	23.12.1981	10247	Wahnvorstellungen
W	03.12.1997	10249	Wahnvorstellungen

In der zweiten Publikation werden folgende Daten verwendet:

Alter	PLZ	Diagnose
40	10115	Depression
57	10115	Delirium
45	10178	Depression
41	10247	Wahnvorstellungen
25	10249	Wahnvorstellungen

Auch wenn in Tabelle 2 das Geschlecht nicht angegeben wurde, kann das Geschlecht aus den Daten in der ersten Veröffentlichung abgeleitet werden. Weiterhin lässt sich der Behandlungszeitraum ableiten: In Publikation 1 ist das Geburtsdatum angegeben, in Publikation 2 das Alter zum Zeitpunkt der Behandlung.

Wird in mindestens einer der Komplementärveröffentlichungen noch der Behandlungszeitraum („…wurden in unserer Klinik im Zeitraum … behandelt …“) angegeben, was sehr häufig vorkommt, kann aus dem angegebenen Alter und dem Geburtsdatum auch das Behandlungsdatum abgeleitet werden.

Bei alleiniger Betrachtung jeder einzelnen Tabelle sind die Daten nicht verknüpfbar, aber mit dem Zusatzwissen „identische Quelle“ ergeben sich als Wissen: Behandlung Ende 2022/ Anfang 2023 in Krankenhaus „xy“ mit Diagnose „…“ usw.

Behandelt in	Geschlecht	Geb.-Datum	Datum Behandlung	PLZ	Diagnose
Krankenhaus „xy“	W	11.05.1983	Mai bis August 2023	10115	Depression
Krankenhaus „xy“	M	03.12.1965	Februar bis August 2023	10115	Delirium
Krankenhaus „xy“	M	27.08.1977	Februar bis August 2023	10178	Depression
Krankenhaus „xy“	M	23.12.1981	Februar bis August 2023	10247	Wahnvorstellungen
Krankenhaus „xy“	W	03.12.1997	Februar bis August 2023	10249	Wahnvorstellungen

Tabelle 11: Datenanalyse bei einem Angriff unter Nutzung von Komplementärveröffentlichungen

Wenn verschiedene Veröffentlichungen zu ein und demselben Datensatz existieren, bieten sich daraus derartige Angriffsmöglichkeiten. D. h. bei einer Pseudonymisierung bzw. Anonymisierung muss bei jeder Veröffentlichung bedacht werden, wie sich diese Veröffentlichung und damit die Tatsache, dass diese Informationen zumindest für einen bestimmten Personenkreis als „öffentlich verfügbare“ Informationen anzusehen sind, auf die Bewertung hinsichtlich des Risikos einer Re-Identifikation auswirken können.

Ggf. können somit aus entsprechenden Veröffentlichungen neue (Sekundär-)Identifier entstehen, welche zusammen mit anderen Daten eine Re-Identifikation ermöglichen könnten.

8.5.2.5 „Linkage-Attacke“

Bei der sog. „Linkage-Attacke“ werden vorhandene pseudonymisierte oder anonymisierte Datensätze mit weiteren dem Angreifer bekannten Datensätzen zusammengeführt und existierende Korrelationen gesucht; der Angriff ähnelt darin dem Unsorted-Matching-Angriff. Bei diesem Angriff wird die Möglichkeit der Berechnung von statistischen Zusammenhängen ergänzend genutzt.

Beispiel: Die nachfolgende Tabelle scheint zunächst anonyme Daten zu enthalten.

Geschlecht	Geb.-Datum	PLZ	ICD
w	01.01.1983	1011	C43
m	01.01.1965	1011	D22
m	01.01.1977	1017	C85
m	01.01.1981	1024	D44
m	01.01.1985	1031	C18
w	01.01.1987	1040	D46
m	01.01.1988	1043	C16
m	01.01.1968	1043	D46
w	01.01.1978	1058	C50
w	01.01.1969	1070	C91
m	01.01.1967	1071	D12
m	01.01.1991	1071	D12
w	01.01.1987	1078	C50
w	01.01.1983	1082	C50
w	01.01.1975	1096	C83

Tabelle 12: Auf Anonymität zu prüfendes Ergebnis

Vorname	Nachname	Geschlecht	Geb.-Datum	PLZ
Heike	Richter	w	11.05.1983	10115
Jan	Schröder	m	03.12.1965	10115
Hugo-Egon	Meyer	m	27.08.1977	10178
Eckbert	Schneider	m	23.12.1981	10247
Jürgen	Stillstand	m	29.11.1985	10319
Hiltrud	Niemand	w	15.07.1987	10407
Uwe	Müller	m	31.03.1988	10435
Michael	Matuschek	m	13.04.1968	10439
Anke	Schmidt	w	01.04.1978	10585
Kunigunde	Gewaltig	w	21.01.1969	10707
Franz	Herrlich	m	17.11.1967	10717
Berthold	Koch	m	28.08.1991	10717
Frieda	Fischer	w	15.11.1987	10787
Gerfriede	Jensen	w	23.07.1983	10827
Käthe	Albers	w	27.05.1975	10963

Tabelle 13: Zuordnungsmöglichkeiten durch die Originaldaten

Existiert hingegen auch nur ein teilweise möglicher Zugriff auf die Originaldaten aus Tabelle 13, so ist eine Re-Identifikation der Daten aus Tabelle 12 möglich: Nur Frau Richter und Herr Schröder wohnen in einem Bereich, dessen PLZ mit „1011“ beginnt und durch die Geschlechtsangabe ist eine eindeutige Zuordnung möglich. Die Daten sind also nicht als anonyme Daten anzusehen, sondern als pseudonyme Daten.

8.5.2.6 Homogenitätsangriff

Der Homogenitätsangriff gehört zu den Methoden, wo verfügbares Zusatzwissen für einen Angriff auf die De-Identifikation genutzt wird.

Beispiel: Max Mustermann ist 1983 geboren. Ist bekannt, dass er zu einem bestimmten Zeitpunkt in einer bestimmten Klinik in Behandlung war, so ist er mit diesem Zusatzwissen bei Zugriff auf (ggf. öffentlich gemachte) Daten ggf. re-identifizierbar.

Geschlecht	Geb.-Jahr	PLZ	Diagnose
W	1983	101*	Depression
M	1981	101*	Wahnvorstellungen
M	1983	101*	Depression
W	1981	101*	Wahnvorstellungen

In obiger Tabelle ist Max Mustermann trotz der k-Anonymität von k=2 sofort zu identifizieren und damit ist seine Diagnose bekannt.

In größeren Tabellen werden Werte ggf. nicht nur auf eine Person zutreffen. Aber selbst, wenn aufgrund der großen Datenmenge sechs Personen die gleichen Werte aufweisen, ist die Diagnose für Max Mustermann immer noch eindeutig – auch wenn man nicht weiß, welcher der sechs Einträge für Max Mustermann richtig ist:

Geschlecht	Geb.-Jahr	PLZ	Diagnose
M	1983	101*	Depression
M	1983	101*	Depression
M	1983	101*	Depression
M	1983	101*	Depression
M	1983	101*	Depression
M	1983	101*	Depression

8.5.3 Grundbedingungen für eine Prüfung

Bei einer Prüfung des Ergebnisses einer Pseudonymisierung/Anonymisierung muss sowohl der Prozess der Verarbeitung, d. h. die Methodik und die Umsetzung der Methodik, wie auch das vorhandene Ergebnis beurteilt werden. Voraussetzung dafür ist eine Dokumentation, die in ihrer Gesamtheit nachvollziehbar ist. Die/der Prüfende benötigt:

- eine ausführliche Dokumentation der Methodik,
- eine detaillierte Darstellung der Umsetzung der Methodik,
- die Ergebnistabellen.

Bei den Ergebnissen ist darauf zu achten, dass auch statistische Kennzahlen Informationen zu einzelnen Individuen beinhalten können:

- Ein Mittelwert, basierend auf wenigen Beobachtungen, kann ggf. Rückschlüsse bzgl. gering besetzter Gruppen beinhalten. Um diese Randgruppen identifizieren und diese bzgl. einer Möglichkeit des Rückschlusses auf einzelne Individuen prüfen zu können, muss neben der Fallzahl immer auch Minimum, Maximum und Standardabweichung angegeben werden.
- Die Angabe von Perzentilen (Prozenträngе) bei einer geringen Fallzahl beinhaltet nahezu immer die Möglichkeit von Rückschlüssen auf einzelne Individuen.

Hochfellner [108] empfiehlt als Mindestgrößen:

- Mindestens 20 Beobachtungen für die Ausgabe von Mittelwerten
- Mindestens 40 Beobachtungen für die Ausgabe von 50%-Perzentilen
- Mindestens 80 Beobachtungen für die Ausgabe von 25%- oder 75%-Perzentilen
- Mindestens 200 Beobachtungen für die Ausgabe von 10%- oder 90%-Perzentilen
- Mindestens 400 Beobachtungen für die Ausgabe von 5%- oder 95%-Perzentilen
- Mindestens 2.000 Beobachtungen für die Ausgabe von 1%- oder 99%-Perzentilen

8.5.4 Risikobewertung ist erforderlich

Aus den in den vorhergehenden Kapiteln dargestellten Gründen muss also eine Risikobewertung erfolgen. Diese setzt zweierlei voraus: Einerseits die Abschätzung der Wahrscheinlichkeit einer Re-Identifikation, andererseits die Betrachtung der Gefahren, die aus einer Re-Identifikation erwachsen könnten.

Dazu werden Risiken am besten einerseits in Kategorien eingeteilt, welche eine Zuordnung der Risiken in individuelle und Individuen übergreifende Risiken erlaubt, z. B.[109]:

- Strukturelle Risiken, beispielsweise gesellschaftlich-politische Risiken (wie z. B. die Informationsmacht, die gegenüber einem Individuum gewonnen wird) oder wirtschaftliche Risiken;
- Individuelle Risiken, wie z. B. die Erhöhung individueller Verletzlichkeit für Straftaten, da jemand erfährt, wo betroffene Personen angreifbar sind;
- Risiken für Gesellschaft und Individuum, z. B. durch Bildung von Persönlichkeitsprofilen oder Fremdbestimmung oder auch die Enttäuschung von Vertraulichkeitserwartungen.

Andererseits müssen Risiken hinsichtlich der Bedeutung erfasst werden, d. h. eine Quantifizierung vorgenommen werden.[110] Es liegt in der Natur der Sache, dass das jeweilige Risiko nur abgeschätzt werden kann, sodass eine Einstufung des Risikos entsprechend einer zuvor definierten Skala möglich ist, somit ein Skalenniveau für das vorhandene Risiko einer Re-Identifizierung angegeben wird, z. B.:

Bewertung	Kriterien	Geschätzte Kosten
Katastrophal	Keine Kontrolle möglich	> 1 Mill. €
Kritisch	Gravierende Mängel / Schäden	≤ 1 Mill. €
Mittelmäßige Auswirkungen	Beträchtliche Abweichungen vom Soll	50 – 100.000 €
Geringe Auswirkungen	Geringe Folgen	< 50.000 €
Vernachlässigbare Auswirkungen	Unbedeutende Folgen	Keine

[108] Hochfellner et al. (2012) FDZ-Methodenreporte: Datenschutz am Forschungsdatenzentrum. (Hrsg.: Bundesagentur für Arbeit). Online, zitiert am 2023-10-14; verfügbar unter http://doku.iab.de/fdz/reporte/2012/MR_06-12.pdf

[109] Stefan Drackert (2014) Die Risiken der Verarbeitung personenbezogener Daten - Eine Untersuchung zu den Grundlagen des Datenschutzrechts. Duncker & Humblot GmbH. ISBN '978-3-428-1 4730-4

[110] Siehe z. B. auch Personal Data Protection CommsissionSingapur (2018) Guide to basic data anonymisation techniques. S. 30.ff. Online, zitiert am 2023-10-14; verfügbar unter https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Other-Guides/Guide-to-Anonymisation_v1-(250118).pdf?la=en

Die DSK veröffentlichte ein Kurzpapier Nr. 18 „Risiko für die Rechte und Freiheiten natürlicher Personen“[111], welches auch Hinweise bzgl. der Schweregradbeurteilung aus Sicht der deutschen Aufsichtsbehörden enthält.

8.5.5 Angreifermodell

Verantwortliche, welche eine Anonymisierung oder Pseudonymisierung durchführen, müssen also letztlich eine Wahrscheinlichkeit für das Risiko einer Re-Identifikation angeben. Dabei ist es hilfreich, die Wahrscheinlichkeit für verschiedene potenzielle Angreifer darzustellen. Einige Angreifer verfügen über zusätzliches Wissen, andere Angreifer über besondere Ressourcen wie Rechenkapazitäten usw.

Für jeden potenziellen Angreifertyp wird vor Beginn der Anonymisierung festgelegt, welches Risiko einer Re-Identifikation seitens des Verantwortlichen als tragbar beurteilt wird; dieses Risiko muss insbesondere bei einer Anonymisierung auch aus Sicht der betroffenen Personen als angemessen bewertet werden können, denn im Falle einer Re-Identifikation und einer daraus resultierenden Klage wird im Rahmen eines Gerichtsprozesses die Höhe des vom Verantwortlichen akzeptierten Risikos letztlich mit entscheiden, ob die Daten überhaupt als „anonym“ anzusehen waren und somit ohne datenschutzrechtliche Rahmenbedingungen verarbeitet werden durften.

Typische Kennzahlen werden im folgenden Kapitel besprochen.

8.5.6 Kennzahlen zur Beurteilung der Güte einer Pseudonymisierung/Anonymisierung

Zur Bestimmung des Erfolgs von Pseudonymisierungs- bzw. Anonymisierungsmaßnahmen gibt es eine Reihe von Kennzahlen. Zu den bekanntesten Kennzahlen zählen:

- k-Anonymität: Die Kennzahl k stellt eine Untergrenze für die Anzahl der Personen mit der gleichen Wertekombination dar. Je höher der Wert k ist, desto größer sind die Gruppen der gemeinsam betrachteten Personen und umso geringer ist die Wahrscheinlichkeit einer Re-Identifizierung einer Person.
- l-Diversität: Eine Äquivalenzklasse heißt l-divers, wenn für jedes sensitive Merkmal mindestens l „gut repräsentierte“ Ausprägungen in der mittels k-Anonymität gebildeten Klasse enthalten sind. Gilt dies für alle im Datensatz enthaltenen Klassen, so ist der Datensatz „l-divers“. Damit ergänzt die l-Diversität die k-Anonymität und verhindert, dass durch in Ausprägungen indirekt enthaltenes Wissen die k-Anonymität durchbrochen wird.
  Ein Beispiel:

Alter	PLZ	Geschlecht	Diagnose
40-50	405**	-	Bluthochdruck
50-60	405**	-	Schwangerschaft
40-50	405**	-	Bluthochdruck
40-50	405**	-	Diab. Mellitus
50-60	405**	-	Diab. Mellitus

  Obwohl die Tabelle k-anonym ist, kann das Geschlecht aus der Diagnose „Schwangerschaft“ abgeleitet und bei allen Personen mit dieser Diagnose ergänzt werden. Die Anforderung der l-Diversität verlangt, dass mindestens „l“ verschiedene Werte für als „sensitiv“ bewertete Felder existieren.

[111] Konferenz der unabhängigen Datenschutzbehörden des Bundes und der Länder (Datenschutzkonferenz - DSK): Kurzpapier 18 „Risiko für die Rechte und Freiheiten natürlicher Personen“. Online, zitiert am 2023-10-14; verfügbar unter https://www.datenschutzkonferenz-online.de/media/kp/dsk_kpnr_18.pdf

„Gut repräsentierte“ Ausprägungen werden in der Literatur[112] wie folgt beschrieben:

- Distinkte l-Diversität: Die einfachste Festlegung stellt sicher, dass mindestens l verschiedene Werte für das sensitive Feld in jeder Äquivalenzklasse existieren.
- Entropische l-Diversität: Die komplexeste Festlegung definiert die Entropie einer Äquivalenzklasse E als die negative Summe von s über den Bereich des sensiblen Attributs. Eine Tabelle hat Entropie l-Diversität, wenn für jede Äquivalenzklasse E gilt: Entropie(E) ≥ log(l).
- Rekursive (c-l) Diversität: Eine Kompromissdefinition, welche sicherstellt, dass der häufigste Wert nicht zu oft vorkommt während weniger häufige Werte nicht zu selten vorkommen.

- t-Closeness: Bei der t-Closeness wird ergänzend noch die statistische Verteilung der sensitiven Merkmale berücksichtigt, d. h. die Verteilung soll nicht zu stark von der Verteilung im Original-Datenbestand abweichen. Dies soll gewährleisten, dass aus der Zugehörigkeit zu einer bestimmten Äquivalenzklasse möglichst keine Rückschlüsse gezogen werden können, die nicht auch aus der gesamten Tabelle hätten gezogen werden können. Eine mittels k-Anonymität gebildete Äquivalenzklasse besitzt die t-Closeness-Eigenschaft, wenn die Verteilung eines sensitiven Merkmals höchstens den Abstand t von der Verteilung des sensitiven Merkmals im ursprünglichen Datenbestand hat. I. d. R. kann davon ausgegangen werden, dass kleinere t-Werte eine größere Ähnlichkeit der Verteilungen in der ursprünglichen Datenbasis bedeutet, was wiederum auf einen höheren Grad der Anonymisierung hindeutet.

  k-Anonymität, l-Diversität und t-Closeness ergänzen sich: Wird alles drei angewendet, so „müssen nicht nur mindestens l verschiedene Werte in jeder Äquivalenzklasse vertreten sein, es ist auch erforderlich, dass jeder Wert so oft vertreten ist, dass die ursprüngliche Verteilung für jedes einzelne Merkmal abgebildet wird“[113].
- δ-Presence: Ist die Ableitung von Mitgliedern einer Gruppe (Membership Disclosure) möglich[114], so kann die Tatsache, dass sich eine bestimmte/bestimmbare Person in einem Datensatz befindet, aus der Tatsache abgeleitet werden, dass eine bestimmte Gruppe von Personen im Datensatz enthalten sind.

  Um vor derartigen Angriffen zu schützen, wurde das Kriterium der δ-Presence entwickelt. δ-Presence ist ein Messwert zur Quantifizierung der Wahrscheinlichkeit, dass eine Person zu einem analysierten Datensatz gehört.

  Bei der δ-Presence werden zwei unterschiedliche Datenbanken benutzt: Im ersten Datensatz befinden sich öffentlich verfügbare Informationen, während im zweiten Datensatz sich die Informationen befinden, welche Angriffspotenzial beinhalten. Daten aus dem zweiten Datensatz werden nur aggregiert/generalisiert veröffentlicht, der Parameter δ gibt dabei die

[112] Li N, Li T, Venkatasubramanian S. (2007) t-Closeness: Privacy Beyond k-Anonymity and l-Diversity. IEEE 23rd International Conference on Data Engineering. https://doi.org/10.1109%2FICDE.2007.367856

[113] Artikel-29-Datenschutzgruppe (2014) Stellungnahme 5/2014 zu Anonymisierungstechniken. S. 22. Online, zitiert am 2023-10-14; verfügbar unter https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/index_en.htm#maincontentSec4

[114] Ausführliche Betrachtung, wenn auch noch unter der Definition des Personenbezugs unter Geltung des BDSG aus dem Jahr 2016, in der Dissertation: Küpper J. „Personenbezug von Gruppendaten? Eine Untersuchung am Beispiel von Scoring und Geo-Gruppendaten“. Herbert Utz Verlag, 1. Auflage 2016. ISBN 978-3-8316-4597-8

Wahrscheinlichkeit an, dass ein Angreifer trotz der Generalisierung auf die geschützten Merkmale der zweiten Datenbank zugreifen kann.

Alternativ werden Gruppenzugehörigkeitsattribute verwendet. Die sensiblen Daten der zweiten Gruppe erhalten Gruppen-IDs, in der öffentlichen Tabelle werden Daten die den Gruppen zugehörige IDs zugeordnet. Bei Nutzung der Gruppen-IDs muss die k-Anonymität gewährleistet werden, d. h. in der öffentlichen Tabelle muss jede Gruppen-ID mindestens k-mal vorhanden sein. Dieses Verfahren ist in der Literatur auch unter dem Begriff „Anatomy“[115] zu finden.

Grundvoraussetzung für den Einsatz der δ-Presence ist somit, dass das Angriffspotenzial bekannt ist.

- (n,k)-Dominanzregel: Die Regel besagt, dass bei einer gegebenen Summe die größten n Beiträge nicht mehr als k Prozent der Summe ausmachen dürfen.

8.6 Aufbau und Struktur einer Verfahrensbeschreibung

Gemäß Art. 30 DS-GVO sind Verarbeitungstätigkeiten in einem Verzeichnis zu führen. Dies gilt selbstverständlich auch für die Verarbeitung im Rahmen einer Pseudonymisierung oder Anonymisierung. Darüber hinaus enthält Art. 32 Abs. 3 DS-GVO eine implizite Anforderung, dass die ergriffenen technischen und organisatorischen Maßnahmen, zu denen sowohl die Pseudonymisierung als auch die Anonymisierung gehören, nachweisbar sein müssen (siehe auch Kapitel 5.2). Es ist daher erforderlich, dass die Verfahren hinreichend genau beschrieben werden.

Eine entsprechende Beschreibung der Durchführung einer Pseudonymisierung bzw. Anonymisierung sollte folgende Informationen enthalten[116, 117]:

- ID des für die Verarbeitung der personenbezogenen Daten Verantwortlichen, wobei sowohl an dieser Stelle wie auch an folgenden Punkten der Aufzählung ID für „Identifikationsdatum“ steht, d. h. ein Datum wie beispielsweise eine Personalnummer, anhand dessen eine Person identifiziert werden kann;
- ID des für die Verarbeitung der anonymisierten oder pseudonymisierten Daten Verantwortlichen;
- Beschreibung der Verarbeitung (Zwecke der Verarbeitung, Art der verarbeiteten Daten, Verarbeitungsmethoden, usw.), für welche die Daten ursprünglich erhoben wurden;
- Beschreibung der Verarbeitung (Zwecke der Verarbeitung, Art der verarbeiteten Daten, Verarbeitungsmethoden, usw.), für welche die Anonymisierung oder Pseudonymisierung benötigt wird;

---

[115] Xia X, Tao Y. (2006) Anatomy: simple and effective privacy preservation. Proceedings of the 32nd international conference on Very large data bases: 139–150. https://dl.acm.org/doi/10.5555/1182635.1164141

[116] Siehe auch DIN EN ISO 25237:2017: Medizinische Informatik – Pseudonymisierung. Erhältlich z. B. online beim Beuth-Verlag unter https://www.beuth.de/de/norm/din-en-iso-25237/258588981

[117] Beispiele bzgl. Policy findet man z. B.

- NHS Business Services Authority: Pseudonymisation and anonymisation of data .policy. Online, zitiert am 2023-10-14; verfügbar unter https://studylib.net/doc/8201785/pseudonymisation-and-anonymisation-of-data-policy
- Mayo Clinic: De-identification and Re-identification of Protected Health Information. Online, zitiert am 2023-10-14; verfügbar unter https://www.mayoclinic.org/documents/deidentification-jax-pdf/doc-20079518
- Sanofi: Clinical Trial Data Sharing Data De-Identification Guidelines. Online, zitiert am 2023-10-14; verfügbar unter https://www.clinicalstudydatarequest.com/Documents/Sanofi-DeIdentification-Guide.pdf

- ggf. Darstellung der Vereinbarkeit von ursprünglichem und neuen Zweck;
- Darstellung der Rechtsgrundlage für die Anonymisierung oder Pseudonymisierung;
- Beschreibung des Verfahrens, mit welchem eine Anonymisierung oder Pseudonymisierung durchgeführt wird, u. a. beinhaltend
    - Kontext der Anonymisierung bzw. Pseudonymisierung,
    - Methode/Verfahren der Anonymisierung bzw. Pseudonymisierung,
    - Beschreibung, welche Datenarten/-kategorien für die Anonymisierung oder Pseudonymisierung ausgewählt wurden sowie eine Begründung, warum diese Daten relevant bzgl. einer Identifikationsmöglichkeit waren, andere nicht,
    - je nach gewähltem Verfahren sind zu allen Datenarten/-kategorien alle Ersetzungen, Zusammenfassungen (Aggregationen), Löschungen usw. festzuhalten, z. B.:
        - Alle männlichen Vornamen wurden durch „Adam“ ersetzt, alle weiblichen Vornamen durch „Eva“.
        - Bildaufnahmen liegen im medizinischen Kontext häufig im DICOM-Format[118] vor. Im DICOM-Format liegen einerseits die Bilddaten selbst vor, andererseits auch Metadaten wie Untersucher, untersuchende Einrichtung, Patientenname usw. Auch ist es möglich, Befunde mit abzuspeichern. Metadaten wie auch Befunde können, wie andere Texte auch, pseudonymisiert oder auch anonymisiert werden. In DICOM Supplement 55[119] findet sich eine Beschreibung zur De-Identifizierung, allerdings stammt das Supplement aus dem Jahr 2002, sodass nicht zwingend alle Identifier, insbesondere indirekte Identifier, gelistet sind.
          Diese Möglichkeit zur Anonymisierung/Pseudonymisierung gilt für die Bilddaten nur eingeschränkt, denn ggf. ist aus den Bilddaten eine Identifizierung möglich, z. B. durch eine 3D-Rekonstruktion des Kopfes. Somit müssen evtl. die Bilddaten selbst auch geändert werden, um eine Pseudonymisierung/Anonymisierung zu erzielen.
        - In Audioaufnahmen wurden alle Namen durch einen Piepton ersetzt, einer Stimmenidentifizierung durch Hinzufügung von Rauschsignalen sowie Tonhöhenveränderungen begegnet.
          Hinweise:
            - Die Stimme eines Menschen ist individuell, Personen können allein anhand ihrer Stimmer identifiziert werden. Somit sind in Tonaufnahmen Menschen, welche sprechen, in der Regel als identifizierbar anzusehen. Mittels moderner Methoden der Mustererkennung können Stimmen ausgetauscht werden und so auch dieses Risiko einer ungewollten Identifikation minimiert werden.
            - In einer aktuellen Studie fanden Forscher heraus, dass sich Personen auch anhand des Häsitationsverhaltens, also der Verwendung von Füllwörtern und anderer Verzögerungsphänomene, identifizieren

[118] Digital Imaging and Communications in Medicine (DICOM. Online, zitiert am 2023-11-02; verfügbar unter https://www.dicomstandard.org/ bzw. der Standard selbst unter https://www.dicomstandard.org/current
[119] DICOM Supplement 55 (2002) Attribute Level Confidentiality (including De-identification). Online, zitiert am 2023-11-02; verfügbar unter https://www.dicomstandard.org/News-dir/ftsup/docs/sups/sup55.pdf

lassen.[120] Selbst wenn man also die Stimme austauscht, kann es sein, dass Personen anhand der verwendeten Sprachmuster erkennbar sind. Um das Risiko dieser Identifikationsmöglichkeit ebenfalls zu minimieren, muss daher auch die Wortwahl angepasst und bspw. Füllwörter entfernt werden, Pausen zwischen Wörtern und Sätzen geändert werden, usw.

- In Videoaufnahmen wurden Gesichter verpixelt, besondere Merkmale wie spezielle Gangarten oder körperliche Auffälligkeiten ebenfalls verpixelt.
  Alternativ: Eine KI, die mittels maschinellem Lernen Gesichter zu erkennen lernte[121], kann automatisiert die Elemente des Gesichts lokalisieren, deren Ausprägung erfassen und dann diese verfremden (ähnlich den bekannten Deepfake-Mechanismen), z. B., indem Gesichter dergestalt gemittelt werden, dass Menschen gleichen Alters, Hautfarbe und Geschlecht das gleiche Gesicht bekommen. Dieses Vorgehen hätte den Vorteil, dass ggf. Mimik/Ausdruck erhalten bleibt, individuelle Gesichtszüge jedoch nicht mehr erkannt werden können.
- Usw.

- ID der Person oder des automatisierten IT-Systems, welches die Anonymisierung oder Pseudonymisierung durchführt.
- Bei einer Anonymisierung der Nachweis der Anonymität, d. h. der Nachweis der Nicht-Beziehbarkeit der verarbeiteten Daten auf eine identifizierte oder identifizierbare natürliche Person einschließlich den zur Anonymisierung verwendeten Kontrollmechanismen sowie die Darstellung des Ergebnisses der Anonymisierung wie beispielsweise die Datenstruktur vorher/nachher;

- Umgang mit ggf. zur Anonymisierung oder Pseudonymisierung genutzten kryptografischen Schlüssel oder einer entsprechenden Verknüpfungstabelle; hierzu gehört insbesondere auch
  - eine Beschreibung dessen, was geschieht, wenn die Organisation ihren Betrieb hinsichtlich der Anonymisierungs- oder Pseudonymisierungs-Aktivitäten einstellt,
  - eine Beschreibung, in welchen Bereichen und für welche Anwendungen die kryptografischen Schlüssel oder die entsprechende Verknüpfungstabelle verwendet werden
  - eine Beschreibung des Gültigkeitszeitraums (aus welchem sich letztlich auch der späteste Zeitpunkt zur Validierung der durchgeführten Pseudonymisierung oder Anonymisierung ergibt),
  - eine Beschreibung der Möglichkeiten und Verfahren zur Verknüpfung mit Alt-Daten oder neu hinzugekommenen Daten, sofern die Möglichkeit vorhanden ist;

[120] Braun A, Elsässer N, Willems L. (2023) Disfluencies Revisited - Are They Speaker-Specific? Languages 8(3): 155. https://doi.org/10.3390/languages8030155

[121] Z. B. Beispiel:
- DeepPrivacy2 - A Toolbox for Realistic Image Anonymization. Online, zitiert am 2023-11-02; verfügbar unter https://github.com/hukkelas/deep_privacy2
- Deface: Video anonymization by face detection. Online, zitiert am 2023-11-02; verfügbar unter https://github.com/ORB-HD/deface

- Bewertung der Anonymisierung, beinhaltend insbesondere
    - Statistische Beurteilung der Anonymisierung,
    - Robustheit der Anonymisierung bzw. Darstellung von vorhandenem Potenzial einer Re-Identifizierung,
    - Risikobewertung;
- Festlegung Zeitpunkt Neubewertung, ob anonyme Daten vorliegen oder nicht, beinhaltend
    - Festlegung eines Datums, wann spätestens eine Neubewertung erfolgt,
    - Darstellung des Prozesses, wie regelmäßig überprüft wird, ob sich Rahmenbedingungen ergaben oder vorhandene Rahmenbedingungen sich änderten, welche eine Neubewertung erfordern,
    - Vorgehen bei Feststellung neu auftretender Risiken (z. B. Angriff auf eine verwendete kryptografische Methode ist bekannt geworden)
- Ausführliche Beschreibung, unter welchen Umständen die Pseudonymisierung durch wen auf welche Art umkehrbar ist und welche Berechtigung hierzu von wem erforderlich ist;
- Festlegung der Beschränkungen, denen der Empfänger der anonymisierten oder pseudonymisierten Daten unterliegt, z. B. vertragliche Regelungen oder die vereinbarten Verarbeitungsgrundsätze zu informationsbezogenen Aktionen mit diesen Daten, insbesondere bzgl. Weiterleitung und Aufbewahrung wie beispielsweise:
    - Der Empfänger darf die Daten nicht öffentlich zugänglich machen.
    - Der Empfänger muss die Daten vor unberechtigtem Zugriff schützen.
    - Der Empfänger darf die Daten nur intern nutzen, um entpersonalisierte Daten zu erzeugen, und erst diese dürfen öffentlich zugänglich gemacht oder an Kunden veräußert werden.
    - Der Empfänger muss die Daten zerstören, wenn die Verarbeitung hinsichtlich der vereinbarten Zwecke beendet wurde und kein weiterer rechtlicher Aufbewahrungsgrund für die Daten mehr existiert.

9 Frequently Asked Questions (FAQ)

9.1 Kann eine juristische Person mehrere Verantwortliche haben?

In dieser Praxishilfe wird immer von „Verantwortlichem“ gesprochen. Der Begriff „Verantwortlicher“ wird in Art. 4 Ziff. 7 DS-GVO wie folgt definiert:

> „‘Verantwortlicher‘ die natürliche oder juristische Person, Behörde, Einrichtung oder andere Stelle, die allein oder gemeinsam mit anderen über die Zwecke und Mittel der Verarbeitung von personenbezogenen Daten entscheidet; sind die Zwecke und Mittel dieser Verarbeitung durch das Unionsrecht oder das Recht der Mitgliedstaaten vorgegeben, so kann der Verantwortliche beziehungsweise können die bestimmten Kriterien seiner Benennung nach dem Unionsrecht oder dem Recht der Mitgliedstaaten vorgesehen werden“.

In den meisten Fällen wird der „Verantwortliche“ durch die Unternehmensleitung repräsentiert, da nur diese „über die Zwecke und Mittel der Verarbeitung von personenbezogenen Daten“ entscheiden darf.

Im medizinischen Kontext sind jedoch mitunter weisungsfreie Tätigkeiten zu berücksichtigen, wodurch auch andere Personen als eine Unternehmensleitung als „Verantwortliche“ anzusehen sind. Forschung beispielsweise ist in den meisten Universitätskliniken arbeitsrechtlicher Bestandteil der Tätigkeit von Ärztinnen und Ärzten. Im Kontext der Forschung sind forschende Mediziner i. d. R. weisungsfrei, d. h. welche personenbezogenen Daten zu welchen Forschungen wie verarbeitet werden, entscheiden selbstständig Mediziner ohne Mitspracherecht der Leitung des Universitätsklinikums. In diesen Fällen sind die forschenden Ärzte als Verantwortliche für die Verarbeitung im Kontext des jeweiligen Forschungsprojektes anzusehen, wobei die oder der Datenschutzbeauftragte des Universitätsklinikums natürlich auch in diesem Kontext seine Zuständigkeit behält.

Ggf. trifft der Verantwortliche die Festlegung der Mittel und Zwecke der Verarbeitung nicht allein, sondern gemeinsam mit einem oder mehreren anderen Verantwortlichen. Die Rechtsfolgen ergeben sich aus Art. 26 DS-GVO. In derartigen Fällen sollte auch für den Vorgang der Anonymisierung bzw. Pseudonymisierung in der zu treffenden Vereinbarung festgelegt werden, wer von den gemeinsam Verantwortlichen welche Verpflichtung gemäß der DS-GVO erfüllt, insbesondere was die Wahrnehmung der Betroffenenrechte (siehe Kapitel 5.3) angeht, und wer welchen Informationspflichten gemäß den Art. 13 und 14 nachkommt. Zu beachten ist auch, dass „Verantwortlicher“ i. S. v. Art. 4 Ziff. 7 DS-GVO die gemeinsam Verantwortlichen sind, somit allen gemeinsam auch die aus der DS-GVO resultierenden Nachweispflichten (siehe Kapitel 5.2) obliegen. Zudem benötigt jeder der gemeinsam Verantwortlichen einen eigenen Erlaubnistatbestand (siehe Kapitel 5.1) für seinen Teil der Verarbeitung.

9.2 Muss ich anonyme Daten, die ich bekomme, auf Anonymität prüfen?

Erhält man Daten, die grundsätzlich einer Person zugeordnet werden könnten, wie dies bei Gesundheitsdaten regelhaft anzunehmen ist, so wird man ein datenschutzrechtlich Verantwortlicher, wenn die Daten nicht als anonym, sondern als personenbezogen oder personenbeziehbar anzusehen sind. Die (juristische oder natürliche) Person, von der man die Daten erhielt, kann bestenfalls grob abschätzen, welche technischen Möglichkeiten, Zusatzwissen usw. beim Empfänger vorhanden sind, die Einstufung als „anonym“ kann daher falsch sein.

Ist man rechtlich ein Verantwortlicher, weil die erhaltenen Daten als personenbezogen oder personenbeziehbar anzusehen sind, so unterliegt man allen Verpflichtungen des Datenschutzrechts[122]. Es ist nicht möglich, sich diesen Pflichten zu entziehen, indem man sich darauf beruft, dass man dem Übermittler der Daten hinsichtlich der Anonymität der Daten vertraut habe. Die ständige Rechtsprechung des EuGH ist in dieser Hinsicht eindeutig: Der für die Verarbeitung Verantwortliche ist „verantwortlich".

Empfängern von Daten ist daher zu raten, diese auf Anonymität zu testen und den Test wie auch Testergebnis nachweisbar zu dokumentieren, sodass auch vor einem oder mehreren Gerichten nachgewiesen werden kann, dass die erhaltenen Daten als anonym anzusehen waren.

9.3 Darf eine Anonymisierung umkehrbar sein?

Nach Art. 2 Ziff. 7 der Richtlinie (EU) 2019/1024 ist „Anonymisierung" ein Prozess, dessen Ergebnis darin besteht, dass Daten nicht oder nicht mehr einer betroffenen Person zugeordnet werden können. Eine Anonymisierung muss daher grundsätzlich unumkehrbar sein.

Ist eine Re-Identifizierung möglich, handelt es sich um eine Pseudonymisierung.

9.4 Absolute oder relative Anonymisierung

Insbesondere in Deutschland wird häufig die Unterscheidung zwischen absoluter und relativer Anonymisierung diskutiert, also die Frage, ob ein Personenbezug, wie er in Art. 4 Ziff. 1 DS-GVO definiert ist (direkt oder indirekt identifizierbare Person), relativ gesehen werden kann. Die Unterscheidung kann wie folgt beschrieben werden:

- Bei absolut anonymen Daten ist eine Wiederherstellung des Personenbezugs unmöglich, auch mit dem größtmöglichen Aufwand und mittels bei anderen vorhandenen Zusatzinformationen ist eine Re-Identifizierung ausgeschlossen.
- Bei relativ anonymen Daten (oft auch als „faktisch anonyme Daten" bezeichnet) ist eine Identifizierung nicht gänzlich ausgeschlossen, aber eine Re-Identifizierung kann nur mit einem unverhältnismäßigen Aufwand erfolgen.

Zu beachten ist, dass bei der Weitergabe anonymisierter Daten, die sich im Nachhinein als personenbezogen oder personenbeziehbar herausstellen, der Übermittler dieser Daten datenschutzrechtlich verantwortlich ist. D. h., er haftet im Falle einer nicht vorhandenen Anonymität für eine unrechtmäßige Weitergabe personenbezogener Daten und Aufsichtsbehörden könnten diese Tat mit einem Bußgeld ahnden. Weiterhin könnten betroffene Personen Schadensersatzansprüche geltend machen. (Siehe auch Kapitel 9.12)

Auch ein Empfänger kann bei der Verarbeitung von anonymen Daten, die sich als personenbezogen i. S. v. Art. 4 Ziff. 1 DS-GVO herausstellen, ggf. für die Verarbeitung haften. (Siehe auch Kapitel 9.2)

Unabhängig davon sollte ein Verantwortlicher beachten, dass eine absolute Sichtweise eine deutliche Einschränkung bedeutet, sodass die Anwendbarkeit des Werkzeugs „Anonymisierung" gerade im

[122]EuGH Urt. v. 2020-07-09, Az. C-272/19, Leitsatz: Art. 4 Nr. 7 DS-GVO ist dahin auszulegen, dass wer allein oder gemeinsam mit anderen über die Zwecke und Mittel der Verarbeitung entscheidet, als „Verantwortlicher" im Sinne dieser Bestimmung einzustufen ist. Online, zitiert am 2023-12-12; verfügbar unter https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A62019CJ0272&qid=1702720547526

medizinischen Kontext aufgrund des kaum abschätzbaren Zusatzwissens, wie es z. B. in sozialen Netzwerken usw. zu finden ist, entsprechend eingeschränkt werden könnte.

Für beide Auffassungen gibt es Argumente, von denen einige im Folgenden dargestellt werden. Beim EuGH ist eine Rechtssache[123] anhängig, die vermutlich klären wird, welche Sichtweise richtig ist. Bis dahin müssen Verantwortliche selbst entscheiden, welcher Sichtweise sie folgen wollen.

9.4.1 Argumentation bzgl. absoluter Anonymität

Für die Argumentation der absoluten Anonymität spricht in erster Linie der Wortlaut von Art. 4 Ziff. 1 DS-GVO:

> „[…] als identifizierbar wird eine natürliche Person angesehen, die direkt oder indirekt […] identifiziert werden kann".

Im Wortlaut des Gesetzestextes selbst findet sich keine Relativierung.

In ErwGr. 26 S. 3, 4 DS-GVO findet sich (Hervorhebung durch die Autoren):

> „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 technologische Entwicklungen zu berücksichtigen sind."

Der Erwägungsgrund, den man zur Interpretation von Art. 4 Ziff. 1 DS-GVO heranziehen muss, beinhaltet hingegen eine Relativierung hinsichtlich dessen, was berücksichtigt werden sollte, um eine Identifizierbarkeit zu beurteilen.

Jedoch sind nach ständiger Rechtsprechung des EuGH[124] „Erwägungsgründe eines Unionsrechtsakts rechtlich nicht verbindlich" und können weder herangezogen werden, um von den Bestimmungen des betreffenden Rechtsakts abzuweichen, noch, um diese Bestimmungen in einem Sinne auszulegen, der ihrem Wortlaut offensichtlich widerspricht.

Somit kann ErwGr. 26 DS-GVO eine eindeutige Auslegbarkeit von Art. 4 Ziff. 12 DS-GVO nicht verändern.

9.4.2 Argumentation bzgl. relativer Anonymität

Für einen relativen Ansatz spricht zunächst, dass aufgrund der vielfältigen digital verfügbaren Datenquellen das verfügbare Zusatzwissen für niemanden abschätzbar ist, eine Anonymisierung

---

[123] EuGH Az. C-413/23 P. Online, zitiert am 2023-12-16; verfügbar unter
- Verfahrensmitteilung: https://curia.europa.eu/juris/document/document.jsf?text=&docid=276483&pageIndex=0&doclang=DE&mode=req&dir=&occ=first&part=1
- Verfahrensdokumentation: https://curia.europa.eu/juris/document/document.jsf?text=&docid=276483&pageIndex=0&doclang=DE&mode=req&dir=&occ=first&part=1

[124] EuGH Urt. v- 2023-10-26, Az. C-307/22. Rn. 44. Online, zitiert am 2023-12-16; verfügbar unter https://eur-lex.europa.eu/legal-content/de/TXT/?uri=CELEX:62022CJ0307

somit (fast) unmöglich wäre. Weiterhin enthält ErwGr. 26 DS-GVO Vorgaben, anhand dessen eine Identifizierbarkeit beurteilt werden muss, und beschreibt eine Beurteilung eines Re-Identifikationsrisikos anhand von Wahrscheinlichkeiten („[…] alle Mittel zu berücksichtigen […] die nach allgemeinem Ermessen wahrscheinlich zur Identifizierung […] genutzt werden") – also einer relativen Anonymisierung.

Auch der EDSA beschreibt eine relative Anonymisierung[125]: „Unter Anonymisierung ist die Verwendung einer Reihe von Techniken zu verstehen, sodass diese Daten nur mit einem unverhältnismäßig hohen Aufwand einer bestimmten oder bestimmbaren natürlichen Person zugeordnet werden können."

Seitens EuGH existiert noch kein abschließendes Urteil. Vereinzelt wird in diesem Zusammenhang das Scania-Urteil des EuGH[126] herangezogen und als Argumente z. B. genannt:

- Rn. 45, S. 2: „Bei der Entscheidung, ob eine natürliche Person unmittelbar oder mittelbar identifizierbar ist, sollten alle Mittel berücksichtigt werden, die vernünftigerweise eingesetzt werden könnten, um die betreffende Person zu bestimmen, ohne dass es jedoch erforderlich ist, dass sich alle zur Identifizierung dieser Person erforderlichen Informationen in den Händen einer einzigen Einrichtung befinden."
- Rn. 46: „[…] werden Daten wie die FIN – die gemäß Art. 2 Nr. 2 der Verordnung Nr. 19/2011 als alphanumerischer Code, den der Hersteller einem Fahrzeug zu dem Zweck zuweist, dass es einwandfrei identifiziert werden kann, definiert sind, und die als solche keine „personenbezogenen" Daten darstellen – für denjenigen, der bei vernünftiger Betrachtung über Mittel verfügt, die es ermöglichen, sie einer bestimmten Person zuzuordnen, zu personenbezogenen Daten."
- Rn. 48: Unter diesen Umständen handelt es sich bei der FIN um ein personenbezogenes Datum im Sinne von Art. 4 Nr. 1 DSGVO der in der Zulassungsbescheinigung ausgewiesenen Person, sofern derjenige, der Zugang zur FIN hat, über Mittel verfügen könnte, die es ihm ermöglichen, die FIN zur Identifizierung des Halters des Fahrzeugs, auf das sich die FIN bezieht, oder zur Identifizierung der Person, die aufgrund eines anderen Rechtstitels denn als Halter über das betreffende Fahrzeug verfügen kann, zu nutzen.

Letztlich geht es in diesem Urteil aber nicht um eine Entscheidung zwischen absolut und relativ im eigentlichen Sinne, sondern um die Klarstellung, dass es anonyme Daten gibt, also Daten, die bei Erzeugung anonym sind. Wenn ein Auto produziert wird, so bekommt es eine FIN. Die FIN ist zu diesem Zeitpunkt nicht personenbezogen. Wenn eine natürliche Person das Auto kauft, wird diese FIN dieser natürlichen Person zugeordnet und stellt ein personenbezogenes Datum für jeden dar, der diese Zuordnung, z. B. durch eine Auskunft bei einer Behörde durchführen kann, vornehmen kann. Dies findet sich so im Urteil wieder:

- Rn. 49: „[…] stellt die FIN, wenn die unabhängigen Wirtschaftsakteure bei vernünftiger Betrachtung über Mittel verfügen können, die es ermöglichen, die FIN einer identifizierten

[125] EDSA: Leitlinien 04/2020 für die Verwendung von Standortdaten und Tools zur Kontaktnachverfolgung im Zusammenhang mit dem Ausbruch von COVID-19. Rn. 15. Online, zitiert am 2023-12-16; verfügbar unter https://edpb.europa.eu/our-work-tools/our-documents/guidelines/_guidelines-042020-use-location-data-and-contact-tracing_de

[126] EuGH Urt. v. 2023-11-09, Az. C-319/22. Online, zitiert am 2023-12-16; verfügbar unter https://eur-lex.europa.eu/legal-content/de/TXT/?uri=CELEX:62022CJ0319

> oder identifizierbaren natürlichen Person zuzuordnen […] ein personenbezogenes Datum im Sinne von Art. 4 Nr. 1 DSGVO dar, selbst wenn die FIN für sich genommen für die Fahrzeughersteller kein persönliches Datum darstellt, insbesondere dann nicht, wenn das Fahrzeug, dem sie zugewiesen wurde, nicht einer natürlichen Person gehört."

Letztlich bestätigt der EuGH hier seine Rechtsprechung im sog. „Breyer Urteil" von 2016[127]. Im Breyer-Urteil findet sich:

- Rn. 45: „Zu prüfen ist jedoch, ob die Möglichkeit, eine dynamische IP-Adresse mit den Zusatzinformationen zu verknüpfen, über die der Internetzugangsanbieter verfügt, ein Mittel darstellt, das vernünftigerweise zur Bestimmung der betreffenden Person eingesetzt werden kann."
- Rn. 46: „Wie der Generalanwalt in Nr. 68 seiner Schlussanträge im Wesentlichen ausgeführt hat, wäre dies nicht der Fall, wenn die Identifizierung der betreffenden Person gesetzlich verboten oder praktisch nicht durchführbar wäre, z. B. weil sie einen unverhältnismäßigen Aufwand an Zeit, Kosten und Arbeitskräften erfordern würde, sodass das Risiko einer Identifizierung *de facto* vernachlässigbar erschiene.

Entsprechend dem Breyer-Urteil des EuGH kommt es hinsichtlich der Beurteilung der Identifizierbarkeit darauf an, ob

a) die Identifizierung der betreffenden Person gesetzlich verboten (wobei man hier ggfs. auch das Recht im Empfängerland prüfen muss) oder
b) die Identifizierung der betreffenden Person praktisch nicht durchführbar wäre.

Obwohl das Breyer-Urteil auf eine relative Anonymisierung hindeutet, ist gerade das Kriterium „praktisch nicht durchführbar" rechtssicher nur schwer abbildbar.

Im Rahmen der Nachweispflicht bei einer Anonymisierung müsste, wenn man sich auf die relative Anonymisierung bezieht, entweder das gesetzliche Verbot oder die praktische Undurchführbarkeit nachgewiesen werden.

9.5 Muss ich bei pseudonymen Daten die Vorgaben der DS-GVO beachten?

Entsprechend der Begriffsbestimmung in Art. 4 Ziff. 5 DS-GVO existieren für pseudonyme Daten zusätzliche Informationen, mit welchen die Daten einer betroffenen Person zugeordnet werden können. Pseudonyme Daten stellen somit personenbezogene Daten i. S. v. Art 4 Ziff. 1 DS-GVO dar.

Somit müssen auch bei der Verarbeitung von pseudonymen Daten die Vorgaben der DS-GVO vollumfänglich eingehalten werden.

9.6 Wieso stellt eine Anonymisierung eine Verarbeitung dar?

Art. 4 Ziff. 2 definiert den Begriff der „Verarbeitung" sehr weitreichend, die darin enthaltenen Begriffe wie „Erheben" oder „Erfassen" stellen lediglich eine nicht abschließende Aufzählung von Beispielen dar. In Art. 4 Ziff. 2 DS-GVO heißt es:

> „Verarbeitung ist jeder mit oder ohne Hilfe automatisierter Verfahren ausgeführte Vorgang oder jede solche Vorgangsreihe im Zusammenhang mit personenbezogenen Daten […}".

[127] EuGH Urt. v. 2016-10-19, Az. C-582/14. Online, zitiert am 2023-12-16; verfügbar unter https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A62014CJ0582&qid=1702730387859

Jeder Vorgang oder Vorgangsreihe im Zusammenhang mit personenbezogenen Daten stellt laut der DS-GVO eine Verarbeitung dar.

Damit personenbezogene Daten zu pseudonymen oder anonymen Daten werden, müssen die Daten entsprechend verändert werden, also „ein Vorgang oder eine Vorgangsreihe“ durchgeführt werden.

Somit stellen sowohl die Anonymisierung als auch die Pseudonymisierung eine Verarbeitung i. S. v. Art. 4 Ziff. 2 DS-GVO dar.[128]

9.7 Braucht man für Pseudonymisierung oder Anonymisierung eine Rechtsgrundlage?

Art. 5 Abs. 1 lit. a DS-GVO gibt vor, dass personenbezogene Daten auf rechtmäßige Weise verarbeitet werden müssen. Nach Art. 6 Abs. 1 DS-GVO erfolgt eine Verarbeitung personenbezogener Daten nur rechtmäßig, wenn mindestens eine der in Art. 6 Abs. 1 lit. a-f DS-GVO genannten Bedingungen erfüllt ist. In Bezug auf die in Art. 9 Abs. 1 DS-GVO genannten Datenkategorien, zu denen Gesundheitsdaten und genetische Daten gehören, ist eine Verarbeitung sogar grundsätzlich verboten, wenn nicht zusätzlich zu den in Art. 6 Abs. 1 DS-GVO enthaltenen Vorgaben mindestens auch eine der in Art. 9 Abs. 2 DS-GVO enthaltenen Bedingungen erfüllt wird.

Somit muss für eine Anonymisierung oder Pseudonymisierung personenbezogener Daten stets ein Erlaubnistatbestand vorliegen.

(Zum Thema siehe auch Kapitel 5.1)

9.8 Muss die Pseudonymisierung extern durchgeführt werden oder kann ein Verantwortlicher diese selbst durchführen?

Entsprechend ErwGr. 29 DS-GVO ist eine Pseudonymisierung bei demselben Verantwortlichen möglich,

- wenn dieser die erforderlichen **technischen und organisatorischen Maßnahmen** getroffen hat,
- um für die jeweilige Verarbeitung zu **gewährleisten**,
- dass **zusätzliche Informationen**, mit denen die personenbezogenen Daten einer speziellen betroffenen Person zugeordnet werden können,
- seitens des Verantwortlichen im Rahmen der Verarbeitung **nicht zugreifbar sind**.

(siehe auch Kapitel 8.1.2)

9.9 VIP-Patient & Co.: Maskierung Name = Pseudonymisierung

In versorgenden Einrichtungen wie z. B. Krankenhäusern gibt es in vielen Informationssystemen die Möglichkeit, die Daten besonderer Patienten – i. d. R. prominente Personen des öffentlichen Lebens oder Beschäftigte der eigenen Institution, sog. VIPs („very important person“) – vor unberechtigten Zugriffen zu schützen, indem eine Maskierung des Namens (und oftmals auch der Anschrift) erfolgt.

[128]Vgl.: Bundesbeauftragte für den Datenschutz und die Informationsfreiheit (BfDI): Positionspapier zur Anonymisierung unter der DSGVO unter besonderer Berücksichtigung der TK-Branche, S. 5, Pkt. 3. Stand: 2020-06-29. Online, zitiert am 2023-12-12; verfügbar unter https://www.bfdi.bund.de/SharedDocs/Downloads/DE/Konsultationsverfahren/1_Anonymisierung/Positionspapier-Anonymisierung.html bzw. pdf-Datei unter https://www.bfdi.bund.de/SharedDocs/Downloads/DE/Konsultationsverfahren/1_Anonymisierung/Positionspapier-Anonymisierung.pdf?__blob=publicationFile&v=6

In der Suchmaske würde man bspw. „Elvis Presley“ nicht finden, da statt des richtigen Namens etwas anderes angezeigt wird, z. B. „John Lennon“.

Bewertung: Dies stellt keine Pseudonymisierung dar (und natürlich erst recht keine Anonymisierung).

Laut der Definition der Pseudonymisierung in Art. 4 Ziff. 5 DS-GVO gelten nur Daten als Pseudonym, wenn

a) die personenbezogenen Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können **und**
b) technische und organisatorische Maßnahmen gewährleisten, dass während der Verarbeitung kein Zugriff auf diese zusätzlichen Informationen erfolgen kann.

In der Einrichtung können die Daten jederzeit einer „spezifischen betroffenen Person“, nämlich Patienten xy auf Station z, zugeordnet werden, wer auf die Station geht, kann i. d. R. dann auch problemlos die Identität feststellen, gerade im Falle von VIP.

Daher stellt diese Maskierung eine Schutzmaßnahme i. S. v. Art. 32 DS-GVO dar, nicht aber eine Pseudonymisierung i. S. v. Art. 4 Ziff. 5 DS-GVO.

9.10 Stellt eine Anonymisierung eine Löschung dar?

Anonyme Daten fallen nicht unter die Regelungshoheit der DS-GVO, somit gelten auch aus der DS-GVO resultierende Löschpflichten nicht für anonyme Daten. Liegen jedoch personenbezogene Daten vor, so gelten die entsprechenden Regelungen. Hierbei stellt sich mitunter die Frage, ob eine gegebene Löschpflicht durch die Anonymisierung der personenbezogenen Daten erfüllt werden kann und ebenso, ob die Löschpflicht in diesem Fall einen Erlaubnistatbestand für die Anonymisierung darstellt.

Löschen wird in der DS-GVO nicht definiert. In ErwGr. 65 S. 2 DS-GVO findet sich: „[…] dass ihre personenbezogenen Daten gelöscht und nicht mehr verarbeitet werden […]“. Ziel einer Löschung besteht also darin, dass personenbezogene Daten nicht mehr verarbeitet werden können.

Eine Löschung stellt also einen Vorgang dar, der mit hundertprozentiger Sicherheit gewährleistet, dass personenbezogene Daten nach einer Löschung nicht wiederhergestellt und verarbeitet werden können.[129] Eine Löschung kann in einer (ebenfalls in Art. 4 Ziff. 2 DS-GVO erwähnten) Vernichtung der Daten bestehen, allerdings verlangt eine Löschung nicht grundsätzlich eine Vernichtung, sonst wären nicht beide Begriffe unabhängig voneinander als Beispiele für eine Verarbeitung in Art. 4 Ziff. 2 DS-GVO genannt worden.

Ob eine Anonymisierung eine Löschung darstellt, wird in der Rechtsliteratur unterschiedlich bewertet:[130]

---

[129] Roßnagel A. (2021) Datenlöschung und Anonymisierung. Verhältnis der beiden Datenschutzinstrumente nach DS-GVO. ZD: 188-192

[130] Beispiele:

- Anonymisierung = Löschung
  - Österreichische Datenschutzbehörde, Beschluss vom 5. Dezember 2018, Geschäftszahl DSB-D123.270/0009-DSB/2018. Online, zitiert am 2023-11-02; verfügbar unter https://www.ris.bka.gv.at/Dokumente/Dsk/DSBT_20181205_DSB_D123_270_0009_DSB_2018_00/DSBT_20181205_DSB_D123_270_0009_DSB_2018_00.html
  - Hornung G, Wagner B. (2020) Anonymisierung als datenschutzrelevante Verarbeitung? Rechtliche Anforderungen und Grenzen für die Anonymisierung personenbezogener Daten. ZD: 223-228

- Argumentation „Anonymisierung = Löschung“:
  Eine Anonymisierung entfernt den Personenbezug. D. h., eine Anonymisierung kann verhindern, dass die personenbezogenen Daten weiter genutzt werden können. Daher kann auch eine Anonymisierung eine Löschung der Daten darstellen - sofern die Anonymisierung nachweisbar irreversibel ist.
  Es muss durch die Anonymisierung sicher gewährleistet werden, dass auch mit künftigen Methoden, leistungsfähigeren Computern, verbesserten Algorithmen usw. eine Re-Identifikation anhand der anonymisierten Daten absolut sicher ausgeschlossen werden kann.
- Argumentation „Anonymisierung ≠ Löschung“:
  Löschen soll entsprechend den Vorgaben von Art. 17 Abs. 1 DS-GVO jedoch dazu führen, dass „gespeicherte personenbezogene Daten vollständig unkenntlich gemacht werden, sodass sie nicht mehr verarbeitet, ausgelesen oder wahrgenommen werden können“[129].
  Somit existiert ein grundlegender Unterschied zwischen einer Löschung und Anonymisierung: Nach einer Anonymisierung können Daten (ohne Personenbezug) weiterverarbeitet werden, nach einer Löschung hingegen nicht. Da eine Anonymisierung und Löschung von Daten zu unterschiedlichen Ergebnissen führen, kann eine Anonymisierung keine Löschung darstellen.

Die Argumentation, dass eine Anonymisierung keine Löschung darstellt, ist aus der Zielrichtung der DS-GVO eher nachvollziehbar: Art. 1 Abs. 2 DS-GVO definiert als Ziel, die Grundrechte und Grundfreiheiten natürlicher Personen und insbesondere deren Recht auf Schutz personenbezogener Daten zu schützen. Würde eine Anonymisierung eine Löschung darstellen, gelten für die anonymisierten Daten die Regelungen der DS-GVO nicht mehr, Verantwortliche könnten mit diesen Daten machen, was sie wollen – was der Zielsetzung der Löschvorgabe von Art. 17 Abs. 1 DS-GVO widerspricht und weiterhin betroffene Personen die Möglichkeit nimmt, über die Verarbeitung ihrer Daten zu entscheiden.

Art. 6 Abs. 1 Richtlinie 2002/58/EG[131] unterscheidet zwischen den Begriffen „löschen“ und „anonymisieren“ („[…] sind […] zu löschen oder zu anonymisieren […]“), der europäische Gesetzgeber schien zumindest 2002 die Begriffe nicht gleichzusetzen, was dafürspricht, dass „Löschung“ und „Anonymisierung“ aus europarechtlicher Sichtweise nicht dasselbe bedeuten; was letztlich aber auch nicht zwingend bedeuten muss, dass eine Anonymisierung *keine* Löschung i. S. d. DS-GVO darstellen könnte.

- Stürmer V. (2020) Löschen durch Anonymisieren? Mögliche Erfüllung der Löschpflicht nach Art. 17 DS-GVO. ZD: 626-631
- Gierschmann S. (2021)
- Gestaltungsmöglichkeiten durch systematisches und risikobasiertes Vorgehen – Was ist schon anonym? Planung und Bewertung der Risiken der Anonymisierung. ZD: 482-486

- Anonymisierung ≠ Löschung
  - Roßnagel A. (2021) Datenlöschung und Anonymisierung. Verhältnis der beiden Datenschutzinstrumente nach DS-GVO. ZD: 188-192
  - Thüsing G, Rombey S. (2021) Anonymisierung an sich ist keine rechtfertigungsbedürftige Datenverarbeitung. Eine Auslegung von Art. 4 Nr. 2 DS-GVO nach den Methoden des EuGH. ZD: 548-553

[131] Richtlinie 2002/58/EG des Europäischen Parlaments und des Rates vom 12. Juli 2002 über die Verarbeitung personenbezogener Daten und den Schutz der Privatsphäre in der elektronischen Kommunikation, Art. 6 Abs. 1. Online, zitiert am 2023-11-02; verfügbar unter https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX:32002L0058

Da die Frage bisher jedoch nicht höchstrichterlich entschieden wurde, kann nicht sicher entschieden werden, welche Sichtweise richtig ist und jeder Rechtsanwender muss sich hier aktuell noch eine eigene Meinung bilden.

9.11 Meine Daten sind nicht länger anonym: Was tun?

Durch Veränderungen in der Technik, neu gefundene mathematische Verfahren usw. kann es im Laufe der Zeit vorkommen, dass als anonym angesehene Daten nicht länger als „anonym“ eingestuft werden können, sondern als pseudonyme Daten angesehen werden müssen.

In diesem Fall sind alle Anforderungen der Datenschutzgesetze anzuwenden, wie beispielsweise:

- Darstellung der Rechtsgrundlage der Verarbeitung;
- Gewährleistung der Betroffenenrechte wie die Information der betroffenen Personen oder Löschen unzulässig verarbeiteter Daten sowie Information von Empfängern der Daten hinsichtlich Erforderlichkeit der Löschung;
- Durchführung einer Datenschutz-Folgenabschätzung;
- Gewährleistung der Sicherheit der Verarbeitung;
- Usw.

9.12 Daten wurden „anonym“ weitergegeben, jetzt erfolgte eine Re-Identifizierung: wer ist verantwortlich?

Der EuGH[132] urteilte im Leitsatz 3, dass der Verantwortliche die Beweislast dafür trägt, dass die getroffenen Schutzmaßnahmen geeignet waren. D. h., der Verantwortliche, der anonymisierte Daten weitergab, welche im Nachhinein re-identifizierbar waren, trägt die Beweislast dafür, dass die erfolgte Anonymisierung als Ergebnis auch den Vorgaben von Art. 2 Ziff. 7 der Richtlinie (EU) 2019/1024 entsprechende Daten lieferte.

Sollte ein Gericht die getroffenen Maßnahmen prüfen, so muss dieses Gericht entsprechend dem erwähnten EuGH-Urteil (Rn. 45) eine materielle Prüfung der Maßnahmen anhand aller in Art. 32 genannten Kriterien sowie der Umstände des Einzelfalls und der dem Gericht dafür zur Verfügung stehenden Beweismittel vornehmen, wobei die mit der betreffenden Verarbeitung verbundenen Risiken zu berücksichtigen sind und zu beurteilen ist, ob Art, Inhalt und Umsetzung dieser Maßnahmen diesen Risiken angemessen sind (Rn. 47).

Weiterhin kann im Fall der unbefugten Offenlegung von personenbezogenen Daten durch Dritte, wie sie eine unbefugte Re-Identifizierung i. d. R. darstellen wird, der Verantwortliche gegenüber den Personen, denen ein Schaden entstanden ist, schadensersatzpflichtig sein (siehe zitiertes Urteil, Rn. 86), es sei denn, der Verantwortliche weist nach, dass er in keinerlei Hinsicht für den Schaden verantwortlich ist.

9.13 Anonymisierung und „Big Data“: Geht das?

Methoden zur Wahrung der Privatsphäre in der Big-Data-Analytik befinden sich noch in der Anfangsphase, und diese Methoden können die Privatsphäre der Benutzer aufgrund von Betriebs-

[132] EuGH Urt. v. 2023-12-14, Az. C-340/21. Online, zitiert am 2023-12-14; verfügbar unter https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A62021CJ0340

und Effizienzproblemen nicht gewährleisten.[133] Dies trifft insbesondere dann zu, wenn nicht absehbar ist, welche Daten von wem zusammengeführt werden. In diesen Fällen kommt durch die Zusammenführung ein immer kaum abschätzbares bzw. bewertbares Zusatzwissen zum eigentlich betrachteten Datensatz zusammen, sodass letztendlich nur die Möglichkeit bleibt, rechtlich, d. h. vertraglich, sicherzustellen, dass eine Zusammenführung des als „anonym" betrachteten Datensatzes mit anderen, unbekannten Datensätzen untersagt wird.

Pramanik et al. bewerteten 2021[133] verschiedene Methoden zur Wahrung der Privatsphäre im Kontext von Big Data Anwendungen und versuchten, deren Vor- und Nachteile darzustellen. Dabei identifizieren die Autoren u. a. die Hauptschwächen bestehender Ansätze zur Wahrung der Privatsphäre in der Big-Data-Analytik und gaben Wirtschaftsunternehmen in diesem Kontext fünf Empfehlungen[133]:

1) Entscheidungsträger in einer Organisation bzw. einem Unternehmen müssen umfassende, anpassungsfähige organisatorische und technologische Rahmenbedingungen entwickeln und umsetzen, um das Vertrauen der Verbraucher aufzubauen und zu erhalten. Es müssen Mechanismen vorgesehen werden, um Einrichtungen, die Daten sammeln und analysieren, für die Einhaltung der Regeln und bewährten Verfahren zur Rechenschaft zu ziehen.
2) Organisationen/Unternehmen sollten ein Team ernennen, welches für drei Schlüsselbereiche verantwortlich ist:
    a. Analyse der Vorgaben des Datenschutzrechts, insbesondere daraus resultierender Verbraucherrechte, und der diesen Gesetzen zugrunde liegenden Prinzipien
    b. Gewinnung neuer Ideen, um diese Vorgaben in eigene Prozesse und Techniken einzubeziehen und zu berücksichtigen
    c. Erstellung eines konkreten Berichts und von Empfehlungen, die an den oder die Entscheidungsträger zur endgültigen Beschlussfassung vorgelegt werden.
3) Organisationen/Unternehmen sollten die Umsetzung von Rechten betroffener Personen in den drei folgenden Punkten umsetzen:
    a. Individuelle Kontrolle, d. h. die betroffenen Personen können selbst bestimmen, welche Daten ein Unternehmen von ihnen erhebt und wie es sie verwendet.
    b. Transparenz, d. h. betroffene Personen können die Datenschutz- und Sicherheitspraktiken des Unternehmens selbst beurteilen.
    c. Kontrolle der Richtigkeit, d. h. betroffene Personen erhalten die Möglichkeit, die sie betreffenden personenbezogene Daten zu korrigieren, um das Risiko nachteiliger Folgen für unrichtige Daten zu vermeiden.
4) Um die Sicherheit von Big Data zu gewährleisten, sollten sich Organisationen/Unternehmen auf die Anwerbung, Einstellung und Schulung von Datensicherheitsexperten sowie auf Datenschutz und Datensicherheit spezialisierten Softwareanalysten und -architekten konzentrieren, welche in der Lage sind, Anwendungen für die Datenexploration und Datenanalyse auf sichere Weise zu entwickeln. Übergreifende Schulungen in verschiedenen Datendisziplinen wie Data Warehousing, Datenintegration, Datenqualität, Content Management und Datenbankverwaltung sind für die meisten Business Intelligence-/Data Warehousing-Experten erforderlich.

[133] Pramanik et al. (2021) Privacy preserving big data analytics: A critical analysis of state-of-the-art. WIREs: e1387-e1392. https://doi.org/10.1002/widm.1387

5) Organisationen/Unternehmen sollten Strategien für eine gute Zusammenarbeit entwickeln, da Big Data eine große Bandbreite beinhaltet und somit verschiedenartige Teams mit unterschiedlichen Kompetenzen erfordert. Big Data sollte mit weitreichendem Zugang gemanagt und von verschiedenen Geschäftseinheiten und Interessengruppen optimal verwendet werden. Eine umfangreiche und reibungslose Zusammenarbeit zwischen allen Geschäftsbereichen ist erforderlich, um alle Vorgänge sicher zu verwalten. Neue Formen der Zusammenarbeit sind erforderlich, um den Datenschutz im Zusammenhang mit Big Data zu kontrollieren und die Verwendung personenbezogener Daten im Geschäftsprozess zu regeln. Es muss ein starker Rahmen für den Datenschutz von Big Data eingeführt werden, um die Zufriedenheit der Verbraucher mit dem Datenschutz zu gewährleisten und wettbewerbsfähige Marktbedingungen zu schaffen.

Keine der in der Arbeit von Pramanik et al. betrachteten Methoden bezieht sich direkt auf das Thema der Anonymisierung oder Pseudonymisierung, jedoch wird eine Befolgung der Empfehlungen – deren Inhalte allesamt in der DS-GVO gesetzlich vorgeschrieben sind und somit schon aus rechtlicher Sicht umgesetzt werden müssen – dazu führen, dass gerade bei Big Data Anwendungen eine Anonymisierung nur in seltenen Fällen sinnvoll umgesetzt werden kann.

9.14 Stand der Technik und Anonymisierung

Im sog. „Kalkar"-Urteil[134] hielt das BVerG schon 1978 fest:

- Im Kalkar Urteil schrieb das BVerfG zum „Stand der Technik" (Rn. 100), dass dieser als der „rechtliche Maßstab für das Erlaubte oder Gebotene hierdurch an die Front der technischen Entwicklung verlagert" gilt. Die „allgemeine Anerkennung und die praktische Bewährung allein" sind für den Stand der Technik nicht ausschlaggebend.
- Der „Stand der Wissenschaft" repräsentiert hingegen den aktuellen Forschungsstand im jeweiligen Thema, umfasst also die neuesten technischen und wissenschaftlichen Erkenntnisse. Wird hierauf Bezug genommen, dass „mit der wissenschaftlichen und technischen Entwicklung Schritt gehalten wird" (Rn. 101 Kalkar-Urteil). Es muss diejenige Vorsorge gegen Schäden getroffen werden, welche „nach den neuesten wissenschaftlichen Erkenntnissen für erforderlich gehalten wird" (Rn. 101 Kalkar-Urteil).

Daraus entwickelte sich die sog. „3-Stufen-Theorie", die sich u. a. im 2008 herausgegebenen „Handbuch der Rechtsförmlichkeit"[135] des Bundesjustizministeriums findet (Rn. 253):

> „Im Interesse der Verständlichkeit der Vorschriften und einer einheitlichen Rechtsanwendung sollten nur folgende Generalklauseln für technische Regeln verwendet werden:
>
> - allgemein anerkannte Regeln der Technik,
> - Stand der Technik und
> - Stand von Wissenschaft und Technik.

[134] BVerfG, Urt. v. 1978-08-08, Az. 2 BvL 8/77. Online, zitiert am 2023-12-29; verfügbar unter https://dejure.org/1978,1, Volltext unter https://openjur.de/u/166332.html

[135] Bundesministerium der Justiz (BMJ): Handbuch der Rechtsförmlichkeit. Stand 2008-09-22. Online, zitiert am 2023-12-29; verfügbar unter https://www.bmj.de/SharedDocs/Publikationen/DE/Fachpublikationen/Handbuch_der_Rechtsfoermlichkeit.html?nn=17134

Welche der drei Grundformen zu wählen ist, richtet sich nach dem Gefährdungspotenzial der Materie, die geregelt werden soll, und nach der technischen Beherrschbarkeit dieses Gefährdungspotenzials."

In den folgenden Randnummern finden sich Vorgaben, wann welche Regel verwendet werden soll:

- Rn: 255: „Die Generalklausel „allgemein anerkannte Regeln der Technik" wird für Fälle mit vergleichsweise geringem Gefährdungspotenzial oder für Fälle verwendet, die auf Grund gesicherter Erfahrungen technisch beherrschbar sind. Allgemein anerkannte Regeln der Technik sind schriftlich fixierte oder mündlich überlieferte technische Festlegungen für Verfahren, Einrichtungen und Betriebsweisen, die nach herrschender Auffassung der beteiligten Kreise (Fachleute, Anwender, Verbraucherinnen und Verbraucher und öffentliche Hand) geeignet sind, das gesetzlich vorgegebene Ziel zu erreichen und die sich in der Praxis allgemein bewährt haben oder deren Bewährung nach herrschender Auffassung in überschaubarer Zeit bevorsteht."
- Rn. 256: „Das Anforderungsniveau bei der Generalklausel „Stand der Technik" liegt zwischen dem Anforderungsniveau der Generalklausel „allgemein anerkannte Regeln der Technik" und dem Anforderungsniveau der Generalklausel „Stand von Wissenschaft und Technik". Stand der Technik ist der Entwicklungsstand fortschrittlicher Verfahren, Einrichtungen und Betriebsweisen, der nach herrschender Auffassung führender Fachleute das Erreichen des gesetzlich vorgegebenen Zieles gesichert erscheinen lässt. Verfahren, Einrichtungen und Betriebsweisen oder vergleichbare Verfahren, Einrichtungen und Betriebsweisen müssen sich in der Praxis bewährt haben oder sollten – wenn dies noch nicht der Fall ist – möglichst im Betrieb mit Erfolg erprobt worden sein."
- Rn. 257: „Die Generalklausel „Stand von Wissenschaft und Technik" umschreibt das höchste Anforderungsniveau und wird daher in Fällen mit sehr hohem Gefährdungspotenzial verwendet. Stand von Wissenschaft und Technik ist der Entwicklungsstand fortschrittlichster Verfahren, Einrichtungen und Betriebsweisen, die nach Auffassung führender Fachleute aus Wissenschaft und Technik auf der Grundlage neuester wissenschaftlich vertretbarer Erkenntnisse im Hinblick auf das gesetzlich vorgegebene Ziel für erforderlich gehalten werden und das Erreichen dieses Ziels gesichert erscheinen lassen."

„Stand von Wissenschaft und Technik" wird beispielsweise in § 7 Abs. 2 Nr. 3 Atomgesetz[136] gefordert, aber insbesondere der Stand der Technik wird von verschiedenen Gesetzen adressiert. Auch Begriffsbestimmungen zum Stand der Technik finden sich in verschiedenen Gesetzen, z. B. in § 3Nr. 6 BImSchG[137], § 3 Nr. 11 WHG[138], § 3 Nr. 28 KrWG[139]. Aus europäischer Sicht entspricht am

[136] Gesetz über die friedliche Verwendung der Kernenergie und den Schutz gegen ihre Gefahren (Atomgesetz): § 7 Genehmigung von Anlagen. Online, zitiert am 2023-12-29; verfügbar unter https://www.gesetze-im-internet.de/atg/__7.html

[137] Gesetz zum Schutz vor schädlichen Umwelteinwirkungen durch Luftverunreinigungen, Geräusche, Erschütterungen und ähnliche Vorgänge (Bundes-Immissionsschutzgesetz - BImSchG): § 3 Begriffsbestimmungen. Online, zitiert am 2023-12-29; verfügbar unter https://www.gesetze-im-internet.de/bimschg/__3.html

[138] Gesetz zur Ordnung des Wasserhaushalts 1) 2) (Wasserhaushaltsgesetz – WHG): § 3 Begriffsbestimmungen. Online, zitiert am 2023-12-29; verfügbar unter https://www.gesetze-im-internet.de/whg_2009/__3.html

[139] Gesetz zur Förderung der Kreislaufwirtschaft und Sicherung der umweltverträglichen Bewirtschaftung von Abfällen (Kreislaufwirtschaftsgesetz – KrWG): § 3 Begriffsbestimmungen. Online, zitiert am 2023-12-29; verfügbar unter https://www.gesetze-im-internet.de/krwg/__3.html

ehesten der Terminus „beste verfügbare Technik“ entsprechend Art. 3 Ziff. 10 Industrieemissions-Richtlinie[140] dem deutschen Begriff „Stand der Technik“.

Die niedrigste der drei Stufen ist somit die Stufe, bei deren Einsatz die meiste Erfahrung im Umgang mit der Technik existiert, die höchste Stufe hingegen diejenige, welche den besten Schutz und das höchste Sicherheitsniveau verspricht (siehe Abbildung 5).

Abbildung 5: 3-Stufen-Modell des eingesetzten Technik-Standes

Art. 2 Ziff. 7 der Richtlinie (EU) 2019/1024 verlangt für eine Anonymisierung, dass „die betroffene Person nicht oder nicht mehr identifiziert werden kann“, eine Anonymisierung also unumkehrbar ist. Somit kann es erforderlich sein, dass bei der Beurteilung, ob Daten als anonym anzusehen sind oder nicht, der „Stand der Wissenschaft“ resp. „Stand von Wissenschaft und Technik“ angewendet werden muss, nicht nur der „Stand der Technik“. „Stand der Wissenschaft“ resp. „Stand von Wissenschaft und Technik“ ist jedenfalls immer dann zu berücksichtigen, wenn dieser Stand in der Realität auch tatsächlich und nicht nur theoretisch anwendbar ist.

Letztlich wird man auch den Schutzbedarf der Daten berücksichtigen müssen. Wie es schon im Handbuch der Rechtsförmlichkeit des BMJ heißt, richtet sich die Auswahl der zu wählenden Stufe einerseits nach dem Gefährdungspotenzial und andererseits nach der technischen Beherrschbarkeit dieses Gefährdungspotenzials. Somit kann man festhalten: Je höher der Schutzbedarf der zu

[140] Richtlinie 2010/75/EU des Europäischen Parlaments und des Rates vom 24. November 2010 über Industrieemissionen (integrierte Vermeidung und Verminderung der Umweltverschmutzung). Online, zitiert am 2017-09-04; Verfügbar unter https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:32010L0075#d1e600-17-1

anonymisierenden Daten eingestuft werden muss, desto eher ist bei der Bewertung der Anonymität wahrscheinlich auch „Stand der Wissenschaft“ resp. „Stand von Wissenschaft und Technik“ zu berücksichtigen.

9.15 Werden Quantencomputer und Quantenkryptographie eine Bewertung hinsichtlich Anonymisierung oder Pseudonymisierung beeinflussen?

Dies hängt von den verwendeten Anonymisierungs- bzw. Pseudonymisierungsmethoden ab. Ein Quantencomputer erhöht zunächst „nur“ die Rechenleistung deutlich. Insbesondere rechenintensive Schutzmaßnahmen werden also von der Tatsache, dass Quantencomputer eingesetzt werden, beeinträchtigt.

Dies trifft auch auf kryptographische Verfahren zu. Die gerade online sehr stark eingesetzten Public-Key-Verfahren[141] sind Methoden der asymmetrischen Kryptografie. Asymmetrische Methoden beruhen wesentlich auf der angenommenen Schwierigkeit bestimmter mathematischer Probleme, wie beispielsweise dem Problem, eine natürliche Zahl in ihre Primfaktoren zu zerlegen. Die Sicherheit dieser Verfahren wird von der Verfügbarkeit von Quantencomputern entsprechend beeinträchtigt. Symmetrische kryptografische Algorithmen wiederum sind aufgrund ihrer Nichtlinearität nur mit hohem Aufwand in einem Quantencomputerschaltkreis abbildbar, sie sind also von der Verfügbarkeit von Quantencomputern deutlich weniger betroffen.

Letztlich kann es hier keine allgemeingültige Antwort auf diese Frage geben, sondern es müssen die jeweils eingesetzten Verfahren betrachtet werden. Das Bundesamt für Sicherheit in der Informationstechnik (BSI) veröffentlichte 2021 einen Leitfaden[142] zum damaligen aktuellen Stand der quantensicheren Kryptografie und im Kapitel 6 werden auch Handlungsempfehlungen gegeben; bei entsprechenden Fragestellungen, wie sie beispielsweise auch bei der Betrachtung langjähriger Forschungsprojekte gestellt werden müssen, wird empfohlen, diesen Leitfaden des BSI als Einstieg in das Thema „quantensichere Kryptographie“ zu nutzen.

---

[141] Ein Beispiel ist die TI-Infrastruktur der gematik, in der ein Großteil der Sicherheit während der Kommunikation durch den Einsatz von Transport Layer Security (TLS) abgesichert wird. Siehe die diversen Implementierungsleitfäden unter https://fachportal.gematik.de/schnelleinstieg/downloadcenter/implementierungsleitfaeden oder auch die übergreifende Spezifikation Verwendung kryptographischer Algorithmen in der Telematikinfrastruktur, abrufbar unter https://fachportal.gematik.de/fachportal-import/files/gemSpec_Krypt_V2.28.0.pdf

[142] Bundesamt für Sicherheit in der Informationstechnik (BSI): Kryptografie quantensicher gestalten - Grundlagen, Entwicklungen, Empfehlungen. Stand 2021-12-16. Online, zitiert am 2023-12-01; verfügbar unter https://www.bsi.bund.de/DE/Service-Navi/Presse/Alle-Meldungen-News/Meldungen/Leitfaden_quanten-sichere_Kryptografie_211216.html bzw. pdf-Datei unter https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/Broschueren/Kryptografie-quantensicher-gestalten.pdf?__blob=publicationFile&v=5

10 Checkliste

10.1 Rechtliche Anforderungen

Anforderung	Erfüllt	Nicht erfüllt
Der Verarbeitungszweck, für welchen die Anonymisierung erfolgen soll, wurde festgelegt		
Darstellung der Rechtsgrundlage für die Anonymisierung/Pseudonymisierung		
Gewährleistung der Informationspflichten nach Art. 13/14 DS-GVO unter Beachtung des Transparenzgebotes von Art. 12 DS-GVO		
Bei Zweckänderung: Information betroffener Personen gemäß Art. 13 Abs. 3 bzw. Art. 14 Abs. 4 DS-GVO vor Beginn der Verarbeitung		
Prüfung, ob Datenschutz-Folgenabschätzung erforderlich ist		
Datenschutz-Folgenabschätzung durchgeführt		

10.2 Inhaltliche Anforderungen

Anforderung	Erfüllt	Nicht erfüllt
Auswahl eines geeigneten Anonymisierungsverfahrens unter Berücksichtigung des Verwendungszweckes der anonymen Daten, z. B. ob statistische Aussagen nach Anonymisierung erhalten bleiben.		
Dokumentation zur Darstellung und zum Nachweis eines ordnungsgemäßen Verfahrens der Anonymisierung veranlasst und begonnen.		
Es wurden Kennzahlen zur Bestimmung der Güte des Ergebnisses ausgewählt und einzuhaltende Mindestvorgaben für die Kennzahlen festgelegt.		
In den Originaldaten werden alle benötigten statistischen Werte bestimmt, die für die Anonymisierung/Pseudonymisierung sowie die Bewertung des Verfahrens benötigt werden.		
Es sind alle Datenarten/-kategorien bestimmt worden, die überarbeitet werden müssen.		

10.3 Organisatorische Anforderungen

Anforderung	Erfüllt	Nicht erfüllt
Es werden alle Daten pseudonymisiert/anonymisiert verarbeitet oder es existiert eine Begründung, warum eine pseudonyme Verarbeitung nicht möglich ist		
Es ist gewährleistet, dass eine dezidierte Analyse zur Beurteilung der Kritikalität der Daten und des Aufwandes zur Anonymisierung bzw. Pseudonymisierung durchgeführt wird und die Ergebnisse dokumentiert sind, sodass jederzeit eine Überprüfung der Ergebnisse erfolgen kann		
Die Pseudonymisierung/Anonymisierung erfolgt im jeweiligen Quellsystem		
Es ist gewährleistet, dass personenbezogenen Daten vor einer Pseudonymisierung/Anonymisierung ausschließlich zu Zwecken der Prüfung auf Inplausibilitäten oder Doppelungen zwischengespeichert werden		
Es ist gewährleistet, dass die Prüfung auf Inplausibilitäten und Dopplungen im Vorfeld einer Pseudonymisierung/ Anonymisierung grundsätzlich automatisiert erfolgt		

10.4 Vorgaben für das Verfahren

Anforderung	Erfüllt	Nicht erfüllt
Es ist sichergestellt, dass das eingesetzte Identifikationsschutzverfahren auch Freitextfelder, Kommentarfelder, Anlagen, etc. im Hinblick auf die Ersetzung von personenbezogenen Daten berücksichtigt		
Es ist gewährleistet, dass der Aufwand zur De-Pseudonymisierung bzw. De-Anonymisierung für jedermann unverhältnismäßig hoch (d. h. faktisch ausgeschlossen) ist		
Es ist gewährleistet, dass der eingesetzte Verfremdungsprozess so gestaltet ist, dass er das notwendige Identifikations-Schutzverfahren wirksam umsetzt		
Es ist beim eingesetzten Verfremdungsprozess sichergestellt, dass mittels der verfügbaren Datenfelder eine Rückführbarkeit nur auf Gruppen mit mindestens 5 Personen möglich ist		
Es ist gewährleistet, dass die verwendete Verfremdungsmethode vertrauenswürdig, sicher implementiert und dokumentiert sowie allen beteiligten Parteien bekannt und funktional überprüfbar ist		

10.5 Nichtangabe

Anforderung	Erfüllt	Nicht erfüllt
Daten, die einen Personenbezug ermöglichen, werden gelöscht		

10.6 Maskierung/Ersetzung

Anforderung	Erfüllt	Nicht erfüllt
Daten, die einen Personenbezug ermöglichen, werden durch andere konstante oder sich ändernde Werte ersetzt		

10.7 Mischung/Shuffeling

Anforderung	Erfüllt	Nicht erfüllt
Es ist sichergestellt, dass die Mischung unter Änderung aller Datensätze und Datenfelder erfolgt		
Daten, die schon für sich eindeutig personenbeziehbar sind und durch eine Mischung nicht veränderbar sind, werden zusätzlich mit einer anderen Verfremdungsmethode bearbeitet		

10.8 Varianzmethode

Anforderung	Erfüllt	Nicht erfüllt
Die jeweilige Erhöhung oder Verringerung basiert auf Zufallswerten		

10.9 Kryptografische Methoden

10.9.1 Verschlüsselung

Anforderung	Erfüllt	Nicht erfüllt
Es ist gewährleistet, dass die Erzeugung des Schlüssels bzw. Schlüsselmaterials ein sicherer Prozess ist.		
Es ist gewährleistet, dass der Erzeugung des Schlüssels bzw. Schlüsselmaterials eine qualitativ hochwertige Zufallszahlenquelle zugrunde liegt.		
Es ist sichergestellt, dass der Schlüssel bzw. das Schlüsselmaterial derart erzeugt wird, dass diese weder vorhersagbar sind, noch erraten werden können.		
Es ist gewährleistet, dass die Vertraulichkeit des Schlüssels bzw. des Schlüsselmaterials während des vollständigen Lebenszyklus der verarbeiteten personenbezogenen Daten gewährleistet ist.		
Es ist sichergestellt, dass der Zugriff auf den Schlüssel bzw. das Schlüsselmaterial auf ein absolutes Minimum vertrauenswürdiger Anwender beschränkt ist.		
Es werden ausschließlich Standard-Verschlüsselungs-Algorithmen entsprechend den Empfehlungen des BSI bzw. der Bundesnetzagentur verwendet.		
Es ist sichergestellt, dass das verwendete Verfahren eine hinreichende Stärke sowie keinerlei bekannte Schwächen aufweist.		
Es ist sichergestellt, dass der für die Verschlüsselungsalgorithmen verwendete Schlüssel von ausreichender Qualität ist.		
Es ist gewährleistet, dass der Schlüssel geheim gehalten wird.		
Es liegt ein Konzept zum Schlüsselmanagement vor und dieses enthält Informationen zum Schlüsseltausch, zur Feststellung von und Vorgehensweisen bei Kompromittierung.		

10.9.2 Hash-Funktionen

Anforderung	Erfüllt	Nicht erfüllt
Es ist sichergestellt, dass ausschließlich Standard-Hash-Funktionen verwendet werden, für die es keine bekannten Schwachstellen gibt.		
Es ist sichergestellt, dass bei Verwendung von Hash-Funktionen ein Salt benutzt wird.		
Es ist sichergestellt, dass der Salt derart erzeugt wird, dass dieser weder vorhersagbar ist noch erraten werden kann.		
Es ist sichergestellt, dass der Zugriff auf den Salt auf ein absolutes Minimum vertrauenswürdiger Anwender beschränkt ist.		
Es ist sichergestellt, dass der Salt von ausreichender Qualität ist (= Mindestentropie von 100 Bit).		
Es ist gewährleistet, dass der Salt geheim gehalten wird.		

10.10 Risikobewertung

Anforderung	Erfüllt	Nicht erfüllt
Analyse, welche Risiken zur Re-Identifizierung bestehen (könnten)		
Bewertung, ob Risiken der Re-Identifizierung akzeptabel sind oder nicht		

11 Gerichtsurteile

LG Berlin (27. Zivilkammer), Urteil vom 09.05.2023, Az. 27 O 140/23

- Titel: Erwähnung von identifizierbarem früheren Partner in juristischer Fachzeitschrift
- Leitsätze:
    1. Für die Frage, ob er – der Antragsteller – auch ohne Namensnennung identifizierbar sei, sei nicht der Durchschnittsleser entscheidend, sondern das Umfeld mit Vorwissen. (Rn. 13)
    2. Wie der Bundesgerichtshof entschieden hat, reicht es aber aus, dass die Information an solche Personen gerät, die aufgrund ihrer sonstigen Kenntnisse in der Lage sind, den Antragsteller zu identifizieren (BGH, Urteil vom 6. Dezember 2022 – VI ZR 237/21)
- Dejure: https://dejure.org/2023,13104
- Entscheidungsdatenbank Berlin: https://gesetze.berlin.de/bsbe/document/KORE521262023

EuG, Urteil vom 26.4.2023 – T-557/20

- Titel: Bestimmung des Personenbezugs von Daten
- Leitsätze:
    1. Der Ausdruck „alle Informationen“ im Zusammenhang mit der Bestimmung des Begriffs „personenbezogene Daten“ in Art. 3 Nr. 1 der Verordnung 2018/1725 ist weit auszulegen. Er ist nicht auf sensible oder private Informationen beschränkt, sondern umfasst potenziell alle Arten von Informationen sowohl objektiver als auch subjektiver Natur in Form von Stellungnahmen oder Beurteilungen, unter der Voraussetzung, dass es sich um Informationen „über“ die in Rede stehende Person handelt. (Rn. 68)
    2. Die letztgenannte Voraussetzung ist erfüllt, wenn die Information aufgrund ihres Inhalts, ihres Zwecks oder ihrer Auswirkungen mit einer bestimmten Person verknüpft ist. (Rn. 69)
- Dejure: https://dejure.org/2023,8600
- EUR-Lex: https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX:62020TJ0557

BGH, Urteil vom 06.12.2022, Az. VI ZR 237/21

- Titel: Zur Personenbestimmbarkeit reicht es aus, wenn nicht jeder die Person identifizieren kann, sondern Informationen an solche Personen geraten können, die aufgrund ihrer sonstigen Kenntnisse in der Lage sind, die betroffene Person zu identifizieren
- Leitsatz
    1. Eine Berichterstattung über eine nicht öffentlich gemachte Liebesbeziehung und ihr Ende sind Teil der Privatsphäre beider daran beteiligter Partner. Sie berührt damit die Privatsphäre beider Partner, soweit diese für potenzielle Leser identifizierbar sind. Dabei ist nicht entscheidend, ob alle oder ein erheblicher Teil der Adressaten der Berichterstattung oder gar der "Durchschnittsleser" die betroffene Person identifizieren können. Es reicht vielmehr aus, dass über die Berichterstattung Informationen über den Betroffenen an solche Personen geraten, die aufgrund ihrer sonstigen Kenntnisse in der Lage sind, die betroffene Person zu identifizieren.
    2.
- Dejure: https://dejure.org/2022,40374
- openJur: https://openjur.de/u/2461988.html

VG Hamburg, Urteil vom 28.7.2022, Az. 21 K 1802/21

- Titel: Bestimmung des Personenbezugs von Daten
- Leitsätze:
    1. Personenbezogene Daten können auch dann vorliegen, wenn diese mittels eines Hash-Verfahrens und anschließender Überschlüsselung verarbeitet worden sind.
    2. Eine unmögliche Zuordnung zu einer Person setzt voraus, dass eine Zuordnung endgültig nicht durchführbar ist.
    3. Eine Formulierung wie „es ist sicherzustellen, dass [...] eine Reidentifizierung nicht erfolgt“ ist kein Verbot der Reidentifizierung, daher ist diese auch nicht rechtlich unmöglich.
- Dejure: https://dejure.org/2022,36335
- openJur: https://openjur.de/u/2459330.html

LG München I (7. Zivilkammer), Beschluss vom 01.06.2021, Az. 7 O 14276/20

- Titel: Anonymisierung von Gerichtsentscheidungen in Patentstreitsachen
- Leitsatz:
    1. Zu den Voraussetzungen und Maßstäben bei der Anonymisierung eines Urteils in einer Patentstreitsache betreffend eine Anti-Anti-Suit-Injunction
- Dejure: https://dejure.org/2021,22969
- openJur: https://openjur.de/u/2347437.html

OLG Köln (7. Zivilsenat), Beschluss vom 13.04.2021, Az. 7 VA 9/21

- Titel: Anonymisierung von Gerichtsentscheidungen
- Leitsatz:
    1. Ob einzelne Dritte, die die Prozessparteien näher kennen, den Sachverhalt trotz der Anonymisierung einer veröffentlichten Gerichtsentscheidung unter Umständen „wiedererkennen“ könnten, ist für die Frage, ob eine ausreichende Anonymisierung vorliegt, nicht ausschlaggebend.
- Dejure: https://dejure.org/2021,38668
- openJur: https://openjur.de/u/2360820.html

OLG Karlsruhe (6. Zivilsenat), Beschluss vom 22.12.2020, Az. 6 VA 25/20

- Titel: Anforderungen an die Anonymisierung bei Urteilsveröffentlichungen
- Leitsätze
    1. Die Entscheidung der Gerichtsverwaltung darüber, ob eine Gerichtsentscheidung als veröffentlichungswürdig publiziert wird, unterliegt ihrem pflichtgemäßen Ermessen. Dies erfordert eine Abwägung der widerstreitenden Interessen, vorliegend des Informationsinteresses der Öffentlichkeit und gegebenenfalls der Fachöffentlichkeit an der Veröffentlichung der Gerichtsentscheidung sowie des Geheimhaltungsinteresses des Betroffenen als Ausfluss des allgemeinen Persönlichkeitsrechts.
    2. Ein Verfahrensbeteiligter kann grundsätzlich nicht ausschließen, dass die ihn betreffende Entscheidung veröffentlicht wird, auch wenn die Prozessparteien der Öffentlichkeit oder einzelnen Dritten trotz Anonymisierung bekannt werden. Allerdings ist bei der Ermessensentscheidung zu beachten, dass der Antragsteller grundsätzlich ein berücksichtigungsfähiges und berechtigtes Interesse daran hat, dass seine persönlichen Angaben und Umstände in der veröffentlichten Entscheidung anonymisiert sind.
- Dejure: https://dejure.org/2020,48121
- openJur: https://openjur.de/u/2352730.html

OLG Köln (15. Zivilsenat), Beschluss vom 14.05.2020, Az. 15 W 10/20

- Titel: Unzureichende Anonymisierung bei Verdachtsberichterstattung
- Leitsätze:
    1. Bei einer identifizierenden Bildberichterstattung über einen strafrechtlichen Verdacht (hier: Spionageverdacht gegen einen Berater der Bundeswehr) muss zusätzlich zu den in der Rechtsprechung anerkannten Grundvoraussetzungen einer identifizierenden Verdachtsberichterstattung im Rahmen des § 23 Abs. 1 Nr. 1 KUG ein sog. „qualifiziertes öffentliches Interesse“ vorliegen.
    2. Bis zu einem erstinstanzlichen Schuldspruch soll "oftmals" das Recht des Beschuldigten auf Schutz der Persönlichkeit das Interesse an einer identifizierenden Bildberichterstattung überwiegen; das gilt - wie hier - jedenfalls dann, wenn die Gefahr der Stigmatisierung aufgrund des Deliktstyps besonders hoch ist und der Beschuldigte bislang nicht im Blickfeld der Öffentlichkeit stand.
- Dejure: https://dejure.org/2020,22065
- openJur: https://openjur.de/u/2343388.html

VG Bayreuth, Beschluss vom 8.5.2018, Az. B 1 S 18.105

- Titel: Personenidentifizierbarkeit gehashten E-Mail-Adressen
- Leitsätze:
    1. Durch den Vorgang des „Hashens“ werden die Daten auch nicht anonymisiert, da der Personenbezug hierdurch nicht völlig aufgehoben wird.
    2. Vielmehr ist es weiterhin mit nicht nur unverhältnismäßigem Aufwand möglich, sie einer bestimmten oder bestimmbaren Person zuzuordnen.
- Dejure: https://dejure.org/2018,13905
- openJur: https://openjur.de/u/2280784.html

EuGH, Urteil vom 19. Oktober 2016, Az. C-582/14

- Titel: Personenbezug dynamischer Internetprotokoll-Adressen
- Leitsätze:
    1. Art. 2 Buchst. a der Richtlinie 95/46/EG des Europäischen Parlaments und des Rates vom 24. Oktober 1995 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Datenverkehr ist dahin auszulegen, dass eine dynamische Internetprotokoll-Adresse, die von einem Anbieter von Online-Mediendiensten beim Zugriff einer Person auf eine Website, die dieser Anbieter allgemein zugänglich macht, gespeichert wird, für den Anbieter ein personenbezogenes Datum im Sinne der genannten Bestimmung darstellt, wenn er über rechtliche Mittel verfügt, die es ihm erlauben, die betreffende Person anhand der Zusatzinformationen, über die der Internetzugangsanbieter dieser Person verfügt, bestimmen zu lassen.
    2. Art. 7 Buchst. f der Richtlinie 95/46 ist dahin auszulegen, dass er einer Regelung eines Mitgliedstaats entgegensteht, nach der ein Anbieter von Online-Mediendiensten personenbezogene Daten eines Nutzers dieser Dienste ohne dessen Einwilligung nur erheben und verwenden darf, soweit ihre Erhebung und ihre Verwendung erforderlich sind, um die konkrete Inanspruchnahme der Dienste durch den betreffenden Nutzer zu ermöglichen und abzurechnen, ohne dass der Zweck, die generelle Funktionsfähigkeit der Dienste zu gewährleisten, die Verwendung der Daten über das Ende eines Nutzungsvorgangs hinaus rechtfertigen kann.
- Dejure: https://dejure.org/dienste/vernetzung/rechtsprechung?Text=C-582/14
- openJur: https://openjur.de/u/2392731.html
- EUR-Lex: https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A62014CJ0582

VG Leipzig, Urteil vom 18.05.2016, Az. 1 K 1720/14

- Titel: Anforderungen an die Anonymisierung bei Urteilsveröffentlichungen
- Leitsatz:
    1. Die Veröffentlichung von Urteilen eines Finanzgerichts ist rechtswidrig, wenn die Urteile vor Weitergabe nicht anonymisiert sind. Eine Anonymisierung erfordert, dass die Person des Klägers auch dann nicht mehr erkennbar ist, wenn diese persönlichen Daten (Beruf, Geburtsdaten, Kinder) an Dritte (sozialüblich) bekannt gegeben hat. Eine vorbeugende Unterlassungsklage gegen eine zu befürchtende unzureichend anonymisierte Veröffentlichung ist zulässig und begründet, wenn konkrete Anhaltspunkte für eine künftig zu erwartende Rechtsverletzung durch das Finanzgericht vorliegen.
- Dejure: https://dejure.org/2016,22574

12 Abkürzungen

Abs.	Absatz
AEUV	Vertrag über die Arbeitsweise der Europäischen Union
Art.	Artikel
BDSG	Bundesdatenschutzgesetz
BfDI	Bundesbeauftragte für den Datenschutz und die Informationsfreiheit
BvD	Berufsverband der Datenschutzbeauftragten Deutschlands (BvD) e. V.
BVerfG	Bundesverfassungsgericht
Bvitg	Bundesverband Gesundheits-IT – bvitg e. V.
DICOM	Digital Imaging and Communications in Medicine
DIN	Deutsche Institut für Normung e. V.
DSFA	Datenschutz-Folgenabschätzung
DSG	Datenschutzgesetz
DS-GVO	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 (Datenschutz-Grundverordnung)
DSK	Datenschutzkonferenz
EDPB	European Data Protection Board
EDSA	Europäischer Datenschutzausschuss
EDV	Elektronische Datenverarbeitung
EN	Europäische Norm
ErwGr.	Erwägungsgrund/Erwägungsgründe
EU	Europäische Union
EuG	Gericht der Europäischen Union; in europäischen Verträgen oft nur kurz als „Gericht“ bezeichnet, ein dem EuGH nachgeordnetes Gericht, welches ursprünglich geschaffen wurde, um den EuGH zu entlasten
EuGH	Europäischer Gerichtshof
EWR	Europäischer Wirtschaftsraum
GDD	Gesellschaft für Datenschutz und Datensicherheit (GDD) e. V.
GMDS	Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e.V.
HIPAA	Health Insurance Portability and Accountability Act
Hs.	Halbsatz
i. d. R.	in der Regel
i. S.	im Sinne
i. S. d.	Im Sinne des/der
i. S. d.	im Sinne der/des
i. V. m.	in Verbindung mit
i.S.v.	im Sinne von
ICD	Eigentlich ICD-10: Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme (**I**nternational Statistical **C**lassification of **D**iseases and Related Health Problems)
ID	Identifikator/Identifier, Identifikationsnummer
ISO	International Organization for Standardization
IT	Informationstechnik, informationstechnisches…
Kap.	Kapitel
LDSG	Landesdatenschutzgesetz
lit.	littera (lat. „Buchstabe“)
Nr.	Nummer

OLG	Oberlandesgericht
pdf	Portable Document Format
PIN	Persönliche Identifikationsnummer
RL	Richtlinie
Rn.	Randnummer
S.	Satz
SGB	Sozialgesetzbuch
StGB	Strafgesetzbuch
StPO	Strafprozeßordnung
StudyUID	Study Unique Identifiers (Untersuchungs-UID)
TAN	Transaktionsnummer
Unterabs.	Unterabsatz
Urt.	Urteil
u. U.	unter Umständen
vgl.	vergleiche
VO	Verordnung
Ziff.	Ziffer

13 Glossar

Aggregierte Daten	Zusammenfassung von Einzelwerten zu größeren Einheiten; ein Rückschluss auf die Einzeldaten ist i. d. R. nicht mehr möglich
Anonymisierte Daten	Daten, die als Ergebnis eines Anonymisierungsprozesses erzeugt wurden (Quelle: DIN EN ISO 25237)
Anonymisierung	Prozess, in dessen Verlauf Dokumente in anonyme Dokumente umgewandelt werden, die sich nicht auf eine identifizierte oder identifizierbare natürliche Person beziehen, oder personenbezogene Daten so anonym gemacht werden, dass die betroffene Person nicht oder nicht mehr identifiziert werden kann (Quelle: Art. 2 Ziff. 7 der Richtlinie (EU) 2019/1024)
Anonymität	Merkmale von Informationen, die eine direkte oder indirekte Identifizierung des Betroffenen nicht zulassen (Quelle: ISO/IEC 29100:2011)
Auftragsverarbeiter	„'Auftragsverarbeiter' eine natürliche oder juristische Person, Behörde, Einrichtung oder andere Stelle, die personenbezogene Daten im Auftrag des Verantwortlichen verarbeitet" (Quelle: Art. 4 Ziff. 8 DS-GVO)
Automatische Verarbeitung	Verarbeitung unter Nutzung von EDV; also z. B. Word- oder Excel-Datei, aber auch KIS, RIS, PACS, unabhängig ob Client-Server-Lösung oder Stand-alone PC, Tablet oder anderweitige Hardware genutzt wird
Betroffener/betroffene Person	Genau genommen „betroffene Person", in der Literatur aber häufig als "Betroffener" aufgeführt; „'Personenbezogene Daten' alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person (im Folgenden „betroffene Person") beziehen; als identifizierbar wird eine natürliche Person angesehen, die direkt oder indirekt, insbesondere mittels Zuordnung zu einer Kennung wie einem Namen, zu einer Kennnummer, zu Standortdaten, zu einer Online-Kennung oder zu einem oder mehreren besonderen Merkmalen, die Ausdruck der physischen, physiologischen, genetischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität dieser natürlichen Person sind, identifiziert werden kann" (Quelle: Art. 4 Ziff. 1 DS-GVO)
Data linkage	Abgleich und Zusammenführung von Daten aus mehreren Datenbanken (Quelle: DIN EN ISO 25237)
Datenverknüpfung	Siehe „Data linkage"
De-Anonymisierung	(Gezielte) Aufhebung einer zuvor durchgeführten Anonymisierung von Daten
De-Identifizierung	Ein Prozess, bei dem Informationen, die zur Identifizierung einer Person verwendet werden könnten, aus einem Datensatz entfernt werden.
Einwilligung	„'Einwilligung' der betroffenen Person jede freiwillig für den bestimmten Fall, in informierter Weise und unmissverständlich abgegebene Willensbekundung in Form einer Erklärung oder einer sonstigen eindeutigen bestätigenden Handlung, mit der die betroffene Person zu verstehen gibt, dass sie mit der Verarbeitung der sie betreffenden personenbezogenen Daten einverstanden ist" (Quelle: Art. 4 Ziff. 11 DS-GVO)
Entpersonalisierung	Allgemeine Benennung für jeden Prozess der Reduktion der Zuordnung zwischen einer Menge von zur Identifizierung geeigneten Daten und der betroffenen Person (Quelle: DIN EN ISO 25237)

Genetische Daten	„'Genetische Daten' personenbezogene Daten zu den ererbten oder erworbenen genetischen Eigenschaften einer natürlichen Person, die eindeutige Informationen über die Physiologie oder die Gesundheit dieser natürlichen Person liefern und insbesondere aus der Analyse einer biologischen Probe der betreffenden natürlichen Person gewonnen wurden“ (Quelle: Art. 4 Ziff. 13 DS-GVO)
Gesundheitsdaten	„'Gesundheitsdaten' personenbezogene Daten, die sich auf die körperliche oder geistige Gesundheit einer natürlichen Person, einschließlich der Erbringung von Gesundheitsdienstleistungen, beziehen und aus denen Informationen über deren Gesundheitszustand hervorgehen“ (Quelle: Art. 4 Ziff. 15 DS-GVO)
Hashfunktion	Eindeutige Prüfsumme mit festgelegter Länge (Hash-Wert) einer Datei oder eines Datensatzes
ICD	Der ICD ist ein Klassifikationssystem für medizinische Diagnosen, wobei sich die Zuordnung eines Codes zu einer Krankheit jährlich ändern kann. D. h., ein Code aus dem Jahr 2004 kann eine andere Krankheit beschreiben als der identische Code aus dem Jahr 2011. Daher muss zur eindeutigen Darstellung der Erkrankung immer die Jahreszahl angegeben werden, wobei sich in der Praxis die Jahreszahl häufig nur indirekt z. B. aus dem Behandlungsdatum ergibt.
Identifizierbare Person	Jemand, der direkt oder indirekt identifiziert werden kann, insbesondere über die Referenz zu einer Identifizierungsnummer oder zu einem oder mehreren Kennzeichen, die bezüglich seiner körperlichen, physiologischen, geistigen, ökonomischen, kulturellen oder sozialen Identität spezifisch sind (Quelle: DIN EN ISO 25237)
Identifizierung	Prozess der Nutzung von behaupteten oder beobachteten Attributen einer juristischen Person mit dem Ziel, diese aus den anderen juristischen Personen in einer Reihe von Identitäten herauszufinden (Quelle: DIN EN ISO 25237)
Normadressat	Rechtssubjekt (z. B. natürliche Person, juristische Person, Personenvereinigung), an die sich die Regelung eines Gesetzes (= einer Norm) richtet
Personenbezeichner	Informationen, deren Zweck darin besteht, eine Person innerhalb eines bestimmten Kontexts eindeutig zu identifizieren (Quelle: DIN EN ISO 25237)
Personenidentifizierung	Prozess der Aufstellung einer Verbindung zwischen einem Informationsobjekt und einer physischen Person (Quelle: DIN EN ISO 25237)
Pseudonym	Personenbezeichner, der sich vom üblicherweise verwendeten Personenbezeichner unterscheidet und mit pseudonymisierten Daten verwendet wird, um für Kohärenz innerhalb des Datensatzes zu sorgen, die alle Informationen über die betroffene Person miteinander verknüpft, ohne die Identität dieser Person in der realen Welt offenzulegen (Quelle: DIN EN ISO 25237)
Pseudonymisierung	„'Pseudonymisierung' 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“ (Quelle: Art. 4 Ziff. 5 DS-GVO)

Re-Identifikation	(Gezielte) Aufhebung einer zuvor durchgeführten Pseudonymisierung oder Anonymisierung von Daten
Salt	Zufällig gewählte Zeichenfolge, die an einen gegebenen Klartext vor der Verwendung als Eingabe einer Hashfunktion angehängt wird
Unumkehrbarkeit	Situation, in der es für einen beliebigen Übergang von identifizierbar zu pseudonym rechentechnisch unmöglich ist, vom Pseudonym auf den ursprünglichen Bezeichner zu schließen (Quelle: DIN EN ISO 25237)
Verantwortlicher	„'Verantwortlicher' die natürliche oder juristische Person, Behörde, Einrichtung oder andere Stelle, die allein oder gemeinsam mit anderen über die Zwecke und Mittel der Verarbeitung von personenbezogenen Daten entscheidet; sind die Zwecke und Mittel dieser Verarbeitung durch das Unionsrecht oder das Recht der Mitgliedstaaten vorgegeben, so kann der Verantwortliche beziehungsweise können die bestimmten Kriterien seiner Benennung nach dem Unionsrecht oder dem Recht der Mitgliedstaaten vorgesehen werden" (Quelle: Art. 4 Ziff. 7 DS-GVO)
Verarbeitung	„'Verarbeitung' jeden mit oder ohne Hilfe automatisierter Verfahren ausgeführten Vorgang oder jede solche Vorgangsreihe im Zusammenhang mit personenbezogenen Daten wie das Erheben, das Erfassen, die Organisation, das Ordnen, die Speicherung, die Anpassung oder Veränderung, das Auslesen, das Abfragen, die Verwendung, die Offenlegung durch Übermittlung, Verbreitung oder eine andere Form der Bereitstellung, den Abgleich oder die Verknüpfung, die Einschränkung, das Löschen oder die Vernichtung" (Quelle: Art. 4 Ziff. 2 DS-GVO)
Verknüpfung von Informationsobjekten	Prozess, der die Aufstellung einer logischen Verbindung zwischen verschiedenen Informationsobjekten ermöglicht (Quelle: DIN EN ISO 25237)

14 Literatur

14.1 Bücher

- Berberich O.: Trusted Web 4.0 – Konzepte einer digitalen Gesellschaft: Konzepte der Dezentralisierung und Anonymisierung. Springer Vieweg, 1. Auflage 2016. ISBN 978-3-662-49189-8
- Haase MS: Datenschutzrechtliche Fragen des Personenbezugs. Mohr Siebeck, 1. Auflage 2015. ISBN 978-3-16-153799-8
- El Emam K, Arbuckle L.: Anonymizing Health Data. O'Reilly Media, Inc., 1. Auflage 2014. ISBN 9781449363079
- Krebs HA, Hagenweiler: Datenanonymisierung im Kontext von Künstlicher Intelligenz und Big Data. Springer Vieweg, 1. Auflage 2022. ISBN 978-3-658-37587-4
- Küpper J.: Personenbezug von Gruppendaten? Eine Untersuchung am Beispiel von Scoring- und Geo- Gruppendaten. Herbert Utz Verlag, 1. Auflage 2016. ISBN 978-3-8316-4597-8
- Mühlenbeck, RL.: Anonyme und pseudonyme Daten. Nomos Verlagsgesellschaft, 1. Auflage 2023. ISBN 978-3-7560-0225-2

14.2 Online

- Artikel-29-Datenschutzgruppe. WP 136 „Stellungnahme 4/2007 zum Begriff 'personenbezogene Daten'". Online, zitiert am 2023-10-14; verfügbar unter http://ec.europa.eu/justice/article-29/documentation/opinion-recommen dation/files/2007/wp136_en.pdf
- Artikel-29-Datenschutzgruppe (2014) Stellungnahme 5/2014 zu Anonymisierungstechniken. Online, zitiert am 2023-11-02; verfügbar unter https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/index_en.htm#maincontentSec4
- Bitkom (2020) Anonymisierung und Pseudonymisierung von Daten für Projekte des maschinellen Lernens. Online, zitiert am 2023-10-14; verfügbar unter https://www.bitkom.org/Bitkom/Publikationen/Anonymisierung-und-Pseudonymisierung-von-Daten-fuer-Projekte-des-maschinellen-Lernens
- Bundesbeauftragte für den Datenschutz und die Informationsfreiheit (2020) Anonymisierung - Eine Standortbestimmung zwischen der DSGVO und dem TKG. Online, zitiert am 2023-10-14; verfügbar unter https://www.bfdi.bund.de/DE/Fachthemen/Inhalte/Telefon-Internet/Positionen/Positionspapier-Anonymisierung-DSGVO-TKG.html
- Bundesverband der Deutschen Industrie e. V. (BDI) (2020) Anonymisierung personenbezogener Daten. Online, zitiert am 2023-10-14; verfügbar unter https://bdi.eu/publikation/news/anonymisierung-personenbezogener-daten/
- DICOM Standard – Supplement 142 Clinical Trial De-identification Profiles (Stand 2009). Online, zitiert am 2023-10-14; verfügbar unter ftp://medical.nema.org/medical/dicom/final/sup142_ft.pdf
- EDPS (2019) Introduction to the hash function as a personal data pseudonymisation technique. Online, zitiert am 2023-10-14; verfügbar unter https://edps.europa.eu/data-protection/our-work/publications/papers/introduction-hash-function-personal-data_en
- EDPS/ AEPD (2021) 10 misunderstandings related to Anonymisation. Online, zitiert am 2023-10-14; verfügbar unter https://edps.europa.eu/data-protection/our-work/publications/papers/aepd-edps-joint-paper-10-misunderstandings-related_en

- ENISA
  - Pseudonymisation techniques and best practices (2019) Online, zitiert am 2023-10-14; verfügbar unter https://www.enisa.europa.eu/publications/pseudonymisation-techniques-and-best-practices
  - Data Pseudonymisation: Advanced Techniques and Use Cases (2021) Online, zitiert am 2023-10-14; verfügbar unter https://www.enisa.europa.eu/publications/data-pseudonymisation-advanced-techniques-and-use-cases
  - Deploying Pseudonymisation Techniques (2022) Online, zitiert am 2023-10-14; verfügbar unter https://www.enisa.europa.eu/publications/deploying-pseudonymisation-techniques
  - Data Protection Engineering (2022) Online, zitiert am 2023-10-14; verfügbar unter https://www.enisa.europa.eu/publications/data-protection-engineering
- Garfinkel S. (2015) De-Identification of Personal Information. Online, zitiert am 2023-10-14; verfügbar unter https://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8053.pdf
- Hochfellner et al. (2012) FDZ-Methodenreporte: Datenschutz am Forschungsdatenzentrum. (Hrsg.: Bundesagentur für Arbeit). Online, zitiert am 2023-10-14; verfügbar unter http://doku.iab.de/fdz/reporte/ 2012/MR_06-12.pdf
- Information Commissioner's Office (ICO) (2012) Anonymisation: managing data protection risk code of practice. Online, zitiert am 2023-10-14; verfügbar unter https://ico.org.uk/media/1061/anonymisation-code.pdf
  Hinweis: Überarbeitung vorhanden, Draft von einzelnen Kapitel ist online verfügbar: https://ico.org.uk/about-the-ico/ico-and-stakeholder-consultations/ico-call-for-views-anonymisation-pseudonymisation-and-privacy-enhancing-technologies-guidance/
- Integrating the Healthcare Enterprise (IHE), Domain „IT Infrastructure“ (ITI): nalysis of Optimal De-Identification Algorithms for Family Planning Data Elements (Stand 2016-12-02). Online, zitiert am 2023-10-14; verfügbar unter http://www.ihe.net/uploadedFiles/Documents/ITI/IHE_ITI_WP_Analysis-of-DeID-Algorithms-for-FP-Data_Elements.pdf
- Integrating the Healthcare Enterprise (IHE), Domain „IT Infrastructure“ (ITI): Algorithm Mapping Spreadsheet (for use with De-Identification Handbook) (Stand 2014-06-06). Online, zitiert am 2018-10-14; verfügbar unter http://www.ihe.net/uploadedFiles/Documents/ITI/IHE_ITI_Handbook_De-Identification-Mapping_Rev1.1_2014-06-06.xlsx
- Integrating the Healthcare Enterprise (IHE), Domain „Quality, Research and Public Health“ (QRPH): Pseudonymization White Paper (Stand: 2008-12-22). Online, zitiert am 2023-10-14; verfügbar unter ftp://ftp.ihe.net/Quality/2009_2010_YR_3/Planning/White%20papers%20yr%203/Pseudonymisation-WP.doc
- Leibniz-Institut für Bildungsforschung und Bildungsinformation: Anonymisieren und Pseudonymisieren. Online, zitiert am 2018-10-14; verfügbar unter https://www.forschungsdaten-bildung.de/anonymisieren-pseudonymisieren
  - Hinweise zur Anonymisierung von qualitativen Daten (2014) Online, zitiert am 2018-10-14; verfügbar unter https://www.pedocs.de/volltexte/2022/21968/pdf/fdb-informiert_1_Meyermann_ua_Hinweise_zur_Anonymisierung_von_qualitativen_Daten_2014_v1-1_A.pdf

    - Hinweise zur Anonymisierung von quantitativen Daten (2015) Online, zitiert am 2018-10-14; verfügbar unter https://www.pedocs.de/volltexte/2022/21970/pdf/fdb-informiert_3_Ebel_ua_Hinweise_zur_Anonymisierung_von_quantitativen_Daten_2015_v1-2_A.pdf
- Oracle. (2013) Data Masking Best Practice. Online, zitiert am 2023-10-22; verfügbar unter https://www.oracle.com/technetwork/database/manageability/data-masking-wp-12c-1964774.pdf
- Personal Data Protection Commsission Singapur (2018) Guide to basic data anonymisation techniques. Online, zitiert am 2023-10-14; verfügbar unter https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Other-Guides/Guide-to-Anonymisation_v1-(250118).pdf?la=en
- Schröder, Dominique (2022) Sachverständigengutachten zum Schutz medizinischer Daten im Rahmen der Vorgaben des § 303b SGB V. Online, zitiert am 2023-11-02; verfügbar unter https://freiheitsrechte.org/uploads/documents/Freiheit-im-digitalen-Zeitalter/Gesundheitsdaten/2022-04-25-Gutachten_Schroeder-Gesundheitsdaten-Gesellschaft_fuer_Freiheitsrechte.pdf
- Schwartmann R, Weiß S. (Hrsg.) Entwurf für einen Code of Conduct zum Einsatz DS-GVO konformer Pseudonymisierung. Stand Oktober 2019. Online, zitiert am 2024-01-25; verfügbar unter https://www.de.digital/DIGITAL/Redaktion/DE/Digital-Gipfel/Download/2019/p9-code-of-conduct.pdf?blob=publicationFile&v=1
- Stiftung Datenschutz (2022) Grundsatzregeln für die Anonymisierung personenbezogener Daten. Online, zitiert am 2023-10-22; verfügbar unter https://stiftungdatenschutz.org/fileadmin/Redaktion/Dokumente/Anonymisierung_personenbezogener_Daten/SDS_Studie_Grundsatzregeln_Web_01.pdf
- Stiftung Datenschutz (2022) Praxisleitfaden zur Anonymisierung. Online, zitiert am 2023-10-22; verfügbar unter https://stiftungdatenschutz.org/fileadmin/Redaktion/Dokumente/Anonymisierung_personenbezogener_Daten/SDS_Studie_Praxisleitfaden-Anonymisieren-Web_01.pdf
- Statistisches Bundesamt (Destatis)
    - Handbuch zur Anonymisierung wirtschaftsstatistischer Mikrodaten (Band 4 der Reihe Statistik und Wissenschaft, 2005). Online, zitiert am 2023-10-14; verfügbar unter https://www.statistischebibliothek.de/mir/receive/DEMonografie_mods_00000267https://www.destatis.de/DE/Publikationen/StatistikWissenschaft/Band4_AnonymisierungMikrodaten.html
    - Verfahren zur Anonymisierung von Einzeldaten (Band 16 der Reihe Statistik und Wissenschaft, 2010). Online, zitiert am 2023-10-14]; verfügbar unter https://www.statistischebibliothek.de/mir/receive/DEMonografie_mods_00000242
    - Methoden der Geheimhaltung wirtschaftsstatistischer Einzeldaten und ihre Schutzwirkung (Band 18 der Reihe Statistik und Wissenschaft, 2010). Online, zitiert am 2023-10-14; verfügbar unter https://www.statistischebibliothek.de/mir/receive/DEMonografie_mods_00000245
- TMF - Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V.
    - Anonymisierung. Online, zitiert am 2023-10-14; verfügbar unter https://www.toolpool-gesundheitsforschung.de/produkte/?term=anonymisierung
    - Pseudonymisierung. Online, zitiert am 2023-10-14; verfügbar unter https://www.toolpool-gesundheitsforschung.de/produkte/?term=pseudonymisierung

- U.S. Department of Health & Human Services: Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. Online, zitiert am 2023-10-14; verfügbar unter https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html

14.3 Zeitschriften

- Abbasi A, Mohammadi B. (2021) A clustering-based anonymization approach for privacy-preserving in the healthcare cloud. Concurrency Computat Pract Exper. (34): e6487-e6491. https://doi.org/10.1002/cpe.6487
- Adrian et al. (2022) Entwicklung und Evaluation automatischer Verfahren zur Anonymisierung von Gerichtsentscheidungen. LTZ: 233-238
- Akgün et al. (2015) Privacy preserving processing of genomic data: A survey. Journal of Biomedical Informatics, http://dx.doi.org/10.1016/j.jbi.2015.05.022
- Amiri et al. (2019) Bayesian-based Anonymization Framework against Background Knowledge Attack in Continuous Data Publishing. Transactions On data Privacy (12): 197-225
- Babu et al. (2013) Achieving k-anonymity Using Improved Greedy Heuristics for Very Large Relational Databases. Transactions on Data Privacy: 1-7
- Bandara K, Bandara D, Fernando S. (2021) Evaluation of Re-identification Risks in Data Anonymization Techniques Based on Population Uniqueness. IEEE 2021. https://ieeexplore.ieee.org/abstract/document/9310884
- Bayardo RJ, Agrawal R. (200) Data Privacy Through Optimal k-Anonymization. 21st International Conference on Data Engineering (ICDE'05), Tokyo, Japan: 217-228, https://doi.org/10.1109/ICDE.2005.42
- Bischoff, C. (2020) Pseudonymisierung und Anonymisierung von personenbezogenen Forschungsdaten im Rahmen klinischer Prüfungen von Arzneimitteln (Teil I) – Gesetzliche Anforderungen. PharmR: 309-315
- Bischoff C, Drechsler J. (2020) Pseudonymisierung und Anonymisierung im Rahmen klinischer Prüfungen von Arzneimitteln (Teil II). PharmR: 389-396
- Branson et al. (2020) Evaluating the re-identification risk of a clinical study report anonymized under EMA Policy 0070 and Health Canada Regulations. Trials: https://doi.org/10.1186/s13063-020-4120-y
- Brisch K, Pieper F. (2015) Das Kriterium der "Bestimmbarkeit" bei Big Data-Analyseverfahren - Anonymisierung, Vernunft und rechtliche Absicherung bei Datenübermittlungen. CR: 724-729
- Chuanlu et al. (2021) Utility Preserved Facial Image De-identification Using Appearance Subspace Decomposition. Chinese Institute of Electronics (30): 413-418. https://doi.org/10.1049/cje.2021.03.004
- Chevrier et al. (2019) Use and Understanding of Anonymization and De-Identification in the Biomedical Literature: Scoping Review. J Med Internet Res): e13484. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6658290/?report=printable#
- De Capitani di Vimercati et al. (2023) k-Anonymity: From Theory to Applications. Transactions On data Privacy (16): 25-49
- El Emam, et al. (2011) A Systematic Review of Re-Identification Attacks on Health Data. PLoS One 6(12): e28071, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3229505/

- Elliot et al. (2018) Functional anonymisation: Personal data and the data environment. Computer Law and Security Review (34): 204-221
- Emam KE, Image O, Bass J. (2020) Evaluating Identity Disclosure Risk in Fully Synthetic Health Data: Model Development and Validation. J Med Internet Res 22(11):e23139. https://doi.org/10.2196/23139
- Gal et al. (2014) A data recipient centered de-identification method to retain statistical Attributes. J Biomed Inform, http://dx.doi.org/10.1016/j.jbi.2014.01.001
- Geschonneck A, Meyer J, Scheben B. (2011) Anonymisierung im Rahmen der forensischen Datenanalyse. BB:2677-2680
- Gille F, Brall C. (2021) Limits of data anonymity: lack of public awareness risks trust in health system activities. Life Sciences, Society and Policy: https://doi.org/10.1186/s40504-021-00115-9
- Hammer V, Knopp M. (2015) Datenschutzinstrumente Anonymisierung, Pseudonyme und Verschlüsselung. DuD: 503509
- Hansson et al. (2016) The risk of re-identification versus the need to identify individuals in rare disease research. Eur J Hum Genet 24(11): 1553–1558. Online, zitiert am 2023-10-14; verfügbar unter https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5110051/
- Härting N. (2013) Anonymität und Pseudonymität im Datenschutzrecht. NJW: 2065-2071
- Hauswaldt et al. (2021) Das Risiko von Re-Identifizierung bei der Auswertung medizinischerRoutinedaten – Kritische Bewertung und Lösungsansätze. ZEFQ (149): 22-31 https://doi.org/10.1016/j.zefq.2020.01.002
- He et al. (2015) CRFs based de-identification of medical records. J Biomed Inform: S39-S46, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4988860/pdf/nihms807198.pdf
- He Z. (2023) From Privacy-Enhancing to Health Data Utilisation: The Traces of Anonymisation and Pseudonymisation in EU Data Protection Law. Digital Society: https://doi.org/10.1007/s44206-023-00043-5
- Höhne H, Strum R, Vorgrimler D. (2003) Konzept zur Beurteilung der Schutzwirkung von faktischer Anonymisierung. Wirtschaft und Statistik (4): 287-292
- Höhne H. (2008) Anonymisierungsverfahren für Paneldaten. Wirt Sozialstat Archiv: 259–275
- Hornby R, Hu RJ. (2021) Identification Risks Evaluation of Partially Synthetic Data with the IdentificationRiskCalculation R Package. Transactions On data Privacy (14): 37-52
- Hornung G, Wagner B. (2020) Anonymisierung als datenschutzrelevante Verarbeitung? Rechtliche Anforderungen und Grenzen für die Anonymisierung personenbezogener Daten. ZD: 223-228
- Huang et al. (2022) Differential privacy: Review of improving utility through cryptography-based technologies. Concurrency Computat Pract Exper. 2023;35:e7565. https://doi.org/10.1002/cpe.7565
- Hu J, Savitsky TD (2023) Bayesian Data Synthesis and Disclosure Risk Quantification: An Application to the Consumer Expenditure Surveys. Transactions On data Privacy (16): 83-121
- Jadhav PS, Borkar GM (2023) Quasi-identifier recognition with echo chamber optimization-based anonymization for privacy preservation of cloud storage. Concurrency Computat Pract Exper. 2023;e7906. https://doi.org/10.1002/cpe.7906
- Jiang et al. (2017) De-identification of medical records using conditional random fields and long short-term memory networks. J Biomed Inform: S43-S53, https://www.sciencedirect.com/science/article/pii/S1532046417302228

- Johannes PC, Geminn CL (2023) Anonymisierung von Patientendaten durch Fremdlabore für Dritte. MedR (41): 368–372
- Karg M. (2015) Anonymität, Pseudonyme und Personenbezug revisited. DuD: 520-526
- Keerie et al. (2018) Data sharing in clinical trials – practical guidance on anonymising trial datasets. Trials: https://DOI.org/10.1186/s13063-017-2382-9
- Kim S, Lee H, Chung YD (2016) Privacy-preserving Data Cube for Electronic Medical Records: An Experimental Evaluation. International Journal of Medical Informatics, http://dx.doi.org/10.1016/j.ijmedinf.2016.09.008
- Knopp M. (2015) Pseudonym – Grauzone zwischen Anonymisierung und Personenbezug. DuD: 527-530
- Kushida, et al. (2012) trategies for De-identification and Anonymization of Electronic Health Record Data for Use in Multicenter Research Studies. Med Care: S82–S101
- Liu et al. (2022) Quantum-resistant anonymous identity-based encryption with trable identities. IET Inf. Secur. (16): 111–126. https://doi.org/10.1049/ise2.12049
- Loukides et al. (2014) Disassociation for electronic health record privacy. Journal of Biomedical Informatics: 46-61
- Lu Y, Sinnott RO, Verspoor K. (2017) A Semantic- ased K- nonymity Scheme for Health Record Linkage. Integrating and Connecting Care: 84-90
- Malin B, Sweeney L. (2004) How (not) to protect genomic data privacy in a distributed network: using trail re-identification to evaluate and design anonymity protection systems. Journal of Biomedical Informatics: 179-192
- Malin B. (2010) Secure construction of k-unlinkable patient records from distributed providers. Artificial Intelligence in Medicine: 29–41
- Malin et al. (2011) Identifiability in biobanks: models, measures, and mitigation strategies. Hum Genet: 383–392
- Mao et al. (2019) Understanding structure-based social network de-anonymization techniques via empirical analysis. Journal on Wireless Communications and Networking https://doi.org/10.1186/s13638-018-1291-2
- Marnau N. (2016) Anonymisierung, Pseudonymisierung und Transparenz für Big Data - Technische Herausforderungen und Regelungen in der Datenschutz-Grundverordnung. DuD: 428-433
- Marnau N, Berrang P, Humbert M. (2018) Anonymisierungsverfahren für genetische Daten. DuD: 83-88
- Mauger C, Mahec GL, Dequen G. (2023) Optimizing Privacy and Data Utility: Metrics and Strategies. Transactions on Data Privacy: 153-189
- Mauw et al. (2019) Conditional adjacency anonymity in social graphs under active attacks. Knowledge and Information Systems (61): 485-511. https://doi.org/10.1007/s10115-018-1283-x
- Meyer S. (2021) Landesrechtliche Legaldefinitionen der „Anonymisierung“ im Anwendungsbereich der DS-GVO. ZD: 669-674
- Murakami T, Takahashi K. (2021) Toward Evaluating Re-identification Risks in the Local Privacy Model. Transactions On data Privacy (14): 79-116
- Neamatullah et al. (2008) Automated de-identification of free-text medical records. BMC Med Inform Decis Mak, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2526997/pdf/1472-6947-8-32.pdf

- Neubauer et al. (2010) Pseudonymisierung für die datenschutzkonforme Speicherung medizinischer Daten. Elektrotechnik & Informationstechnik: 135–142
- Ni et al. (2019) An anonymous entropy-based location privacy protection scheme in mobile social networks. Journal on Wireless Communications and Networking: https://doi.org/10.1186/s13638-019-1406-4
- Pilán et al. (2022) The Text Anonymization Benchmark (TAB): A Dedicated Corpus and Evaluation Framework for Text Anonymization. https://arxiv.org/abs/2202.00443
- Poulis et al. (2017) Anonymizing datasets with demographics and diagnosis codes in the presence of utility constraints. Journal of Biomedical Informatics: 76-96
- Pramanik et al. (2021) Privacy preserving big data analytics: A critical analysis of state-of-the-art. WIREs: e1387-e1392. https://doi.org/10.1002/widm.1387
- Prasser et al. (2020) Flexible data anonymization using ARX—Current status and challenges ahead. Softw: Pract Exper. (50): 1277–1304. https://doi.org/10.1002/spe.2812
- Puri V, Kaur P, Sachdeva S. (2022) (k,m, t)-anonymity: Enhanced privacy for transactional data. Concurrency Computat Pract Exper. (34): e7020. https://doi.org/10.1002/cpe.7020
- Raisaro et al. (2017) Addressing Beacon re-identification attacks: quantification and mitigation of privacy risks. JAMIA: 1-8, doi: 10.1093/jamia/ocw167
- Rocher L, Hendrickx JM, de Montjoye Y. (2019) Estimating the success of re-identifications in incomplete datasets using generative models. Nat Commun 10: 3069. https://doi.org/10.1038/s41467-019-10933-3
- Rosemann M, Vorgrimler D, Lenz R. (2004) Erste Ergebnisse faktischer Anonymisierung wirtschaftsstatistischer Einzeldaten. Allgemeines Statistisches Archiv: 73–99
- Roßnagel A. (2018) Pseudonymisierung personenbezogener Daten. ZD: 243-247
- Roßnagel A. (2021) Datenlöschung und Anonymisierung. Verhältnis der beiden Datenschutzinstrumente nach DS-GVO. ZD: 188-192
- Sei Y, Okumura H, Ohsuga A. (2022) Re-Identification in Differentially Private Incomplete Datasets. IEEE Open Journal of the Computer Society (3): 62-42. https://doi.org/10.1109/OJCS.2022.3175999
- Shao et al. (2019) Fast De-anonymization of Social Networks with Structural Information. Data Science and Engineering (4): 76-92. https://doi.org/10.1007/s41019-019-0086-8
- Shringarpure S, Bustamante C. (2015) Privacy Risks from Genomic Data-Sharing Beacons. The American Journal of Human Genetics: 631–646
- Slijepčevič et al. (2021) k-Anonymity in practice: How generalisation and suppression affect machine learning classifiers. Computers & Security: 102488. https://www.sciencedirect.com/science/article/pii/S0167404821003126
- Steinebach M, Vogel I. (2022) IT and more: Anonymisierung unstrukturierter Daten. WPg: 1230-1232
- Stürmer V. (2020) Löschen durch Anonymisieren? Mögliche Erfüllung der Löschpflicht nach Art. 17 DS-GVO. ZD: 626-631
- Stummer S. (2023) Identifizierbarkeit und Anonymität im Internet. Metriken zur Verifikation des Anonymitätsgrads im Rahmen der Internetnutzung. ZfDR: 263-280
- Thüsing G, Rombey S. (2021) Anonymisierung an sich ist keine rechtfertigungsbedürftige Datenverarbeitung. ZD: 548-553

- Torra V, Navarro-Arribas G. (2023) Attribute disclosure risk for k-anonymity: the case of numerical data. International Journal of Information Security (22): 2015-2024. https://doi.org/10.1007/s10207-023-00730-x
- Tsou et al. (2021) (k, ε, δ)-Anonymization: privacy-preserving data release based on k-anonymity and differential privacy. Service Oriented Computing and Applications (15): 175-185. https://doi.org/10.1007/s11761-021-00324-2
- Wallace SE. (2016) What Does Anonymization Mean? DataSHIELD and the Need for Consensus on Anonymization Terminology. Biopreservation and Biobanking: 224-230
- Wang J, Kwan M (2020) Daily activity locations k-anonymity for the evaluation of disclosure risk of individual GPS datasets. J Health Geogr: https://doi.org/10.1186/s12942-020-00201-9
- Wang et al. (2020) A Comprehensive Overview of Person Re-Identification Approaches. IEEE 2021. https://ieeexplore.ieee.org/document/9023965
- Weitzenboeck etr al. (2022) The GDPR and unstructured data: is anonymization possible? International Data Privacy Law (12): 184-206
- Wójtowicz M, Cebulla M. (2017) Anonymisierung nach der DSGVO. PinG: 186-192
- Xia X, Tao Y. (2006) Anatomy: simple and effective privacy preservation. Proceedings 101ft he 32nd international conference on Very large data bases: 139–150. https://dl.acm.org/doi/10.5555/1182635.1164141
- Zou et al. (2021) Person re-identification based on metric learning: a survey. Multimedia Tools and Applications (80): 26855-26888
- Zouinina et al. (2020) Data Anonymization through Collaborative Multi-view Microaggregation. J. Intell. Syst. (30): 327-45. https://doi.org/10.1515/jisys-2020-0026