Datenschutz in der medizinischen Forschung

Leitfaden für Forschende in der Medizin

Erstellt durch den Hessischen Beauftragten für Datenschutz und Informationsfreiheit (HBDI) und die Deutsche Gesellschaft für Innere Medizin e.V. (DGIM)

Herausgeber:

Prof. Alexander Rossnagel
Hessischer Beauftragter für Datenschutz und Informationsfreiheit

Univ.-Prof. Dr. med. Georg Ertl
Generalsekretär der Deutschen Gesellschaft für Innere Medizin

Stand: 28.10.2025

DGIM Deutsche Gesellschaft für Innere Medizin

Inhaltsverzeichnis

Dieser Text verwendet aus Gründen der Lesbarkeit kontextbezogen ein generisches Maskulinum. Generell sind an diesen Stellen geschlechtsunabhängig alle potenziell betroffenen Personengruppen impliziert.

1. Autorenverzeichnis

Prof. Dr. Alexander Rossnagel	Hessischer Beauftragter für Datenschutz und Informationsfreiheit Wiesbaden
Dr. Nils Gaebel	Hessischer Beauftragter für Datenschutz und Informationsfreiheit, Referat 2.4, Gesundheit und Pflege, Wissenschaft und Forschung, Statistik, Wiesbaden
Tobias Schäfer	Hessischer Beauftragter für Datenschutz und Informationsfreiheit, Referat 2.4, Gesundheit und Pflege, Wissenschaft und Forschung, Statistik Wiesbaden
Univ.-Prof. Dr. med. Georg Ertl	Deutsche Gesellschaft für Innere Medizin Wiesbaden
Dr. med. Thomas Gamstätter	Deutsche Gesellschaft für Innere Medizin Wiesbaden
Dr. med. Philipp Stachwitz	Deutsche Gesellschaft für Innere Medizin Wiesbaden

2. Abkürzungsverzeichnis

AEUV	Vertrag über die Arbeitsweise der Europäischen Union
AMG	Arzneimittelgesetz
AWS	Amazon Web Services
BDSG	Bundesdatenschutzgesetz
DGIM	Deutsche Gesellschaft für Innere Medizin
DKG	Deutsche Krankenhausgesellschaft
DSFA	Datenschutzfolgenabschätzung
DS-GVO	Datenschutzgrundverordnung
DSK	Konferenz der unabhängigen Datenschutzbehörden des Bundes und der Länder
EDSA	Europäischer Datenschutzausschuss
EHDS	European Health Data Space
ePA	elektronische Patientenakte
ErwG	Erwägungsgrund
EuGH	Europäischer Gerichtshof
FL	föderiertes Lernen
G-BA	Gemeinsamer Bundesausschuss
GDNG	Gesundheitsdatennutzungsgesetz
HBDI	Hessischer Beauftragter für Datenschutz und Informations-freiheit
HDSIG	Hessisches Datenschutz- und Informationsfreiheitsgesetz
HIPAA	Health Insurance Portability and Accountability Act

HKHG	Hessisches Krankenhausgesetz
IIT	Investigator Initiated Trials
KI	Künstliche Intelligenz
MPG	Medizinproduktegesetz
MSI	Mikrosatelliteninstabilität
PET	Privacy Enhancing Technology
SGB V	Sozialgesetzbuch V
StPO	Strafprozessordnung
VPN	Virtual Private Network

3. Einleitung

Die Deutsche Gesellschaft für Innere Medizin (DGIM) nahm im Jahr 2022 öffentlich Stellung zum Thema Datenschutz bei der Nutzung von Gesundheitsdaten zu Forschungs- und Versorgungszwecken.[1] Hierbei sah sie sich von der Prämisse geleitet, Gesundheitsdaten im Sinne des Grundrechts auf körperliche Unversehrtheit zur gesundheitlichen Wohlfahrt der Patientinnen und Patienten in Deutschland einzusetzen und bestehende Regelungen und Auslegungen in diesem Sinne zu hinterfragen. Daraufhin nahm der Hessische Beauftragte für Datenschutz und Informationsfreiheit (HBDI) Kontakt mit der Fachgesellschaft auf und schlug einen inhaltlichen Austausch vor. Es entwickelte sich ein Dialog zwischen den DGIM und HBDI, der sich vertieft mit den Positionen der DGIM[2] sowie den Ausführungen der „Petersberger Erklärung“[3] zum Datenschutz im Bereich der wissenschaftlichen Forschung befasste. Übereinstimmend waren beide Organisationen der Ansicht, dass die Fachwelt und die medizinisch-wissenschaftliche Öffentlichkeit von der Erarbeitung einer konkreten, praxisnahen Handreichung zu kritischen und zeitaktuellen Fragen zum Schutz von Gesundheitsdaten und den datenschutzrechtlichen Rahmenbedingungen profitieren könne.

Hierbei sollten konkrete Fallbeispiele aus Themenbereichen, die Forschung und Versorgung innerhalb der Inneren Medizin aktuell maßgeblich bestimmen, adressiert werden. Hierzu zählen u.a. Fragen zur Anonymisierung und Personenbeziehbarkeit von Gesundheitsdaten, zur Nutzung von Gesundheitsdaten aus der Versorgungsroutine zu Forschungszwecken, zum Umgang mit genomischen Daten oder die Nutzung von künstlicher Intelligenz im Umgang mit Gesundheitsdaten.

In der Folge stellte die DGIM dem HBDI eine Sammlung mehrerer realer Anwendungsfälle zur Verfügung, bei denen Mediziner oder medizinische Forschende auf Probleme stießen, die sie zum großen Teil als datenschutzrechtliche Probleme einordneten. DGIM und HBDI haben diese Anwendungsfälle genauer untersucht und möchten die Ergebnisse der Untersuchung nunmehr der medizinischen Fachwelt mit diesem Leitfaden zur Verfügung stellen.

---

1 https://www.dgim.de/fileadmin/user_upload/PDF/Pressemeldungen/PM_DGIM_Datenschutz_April_2022_korr_28.04.2022_F.pdf

2 https://www.dgim.de/fileadmin/user_upload/PDF/UEber_uns/Gremien/Digitale_Transformation/Anforderungen_an_ein_Gesundheitsdatennutzungsgesetz.pdf

3 https://www.bfdi.bund.de/SharedDocs/Downloads/DE/DSK/DSKEntschliessungen/104DSK-Petersberger-Erklaerung.html

4. Die Bedeutung von Gesundheitsdaten für die klinische Forschung und Versorgung aus Sicht der DGIM

Die Nutzung von Daten als Grundlage einer rationalen Entscheidungsfindung und zur Fortentwicklung von evidenz-basierten Handlungsempfehlungen spielt eine zentrale Rolle in der Inneren Medizin. Die Verfügbarmachung und Heranziehung von Daten, die der Patientenversorgung entstammen, stellen dabei neben solchen aus klinischen Studien eine besonders wertvolle Quelle der Information für die internistische Forschung dar. Diese Daten spiegeln unmittelbare Besonderheiten und Entwicklungen innerhalb eines betroffenen Kollektivs wider. Auch fallen diese Daten regulär im Rahmen der Versorgung an und existieren daher unabhängig von Forschungsprojekten repräsentativ für einen Bevölkerungsquerschnitt.

Die Möglichkeit, große Datenmengen mit patienten-relevanten Endpunkten aus vielfältigen Quellen wie der elektronischen Patientenakte (ePA), bildgebenden Verfahren und genetischen Analysen zu integrieren und zu analysieren, bildet eine Grundlage der modernen Präzisionsmedizin und der personalisierten Behandlung und der hierzu erforderlichen medizinischen Forschung. Aus den Daten lassen sich komplexe Muster erkennen, um die Prädiktion von Krankheitsverläufen zu verbessern und maßgeschneiderte Therapieansätze zu entwickeln, welche auf die individuellen Bedürfnisse von Patienten zugeschnitten sind. Für Betroffene resultiert eine umfassende und transparente Information über ihre Erkrankung, die zu einem Mehr an Gesundheitskompetenz führt.

Die Implementierung moderner, datengestützter Analyseverfahren und Technologien in die klinische Praxis und Forschung verspricht darüber hinaus eine Verbesserung von Diagnostik, Therapie und Nachsorge.

In diesem Kontext ist ein verantwortungsvoller Umgang mit Gesundheitsdaten von entscheidender Bedeutung. Der Datenschutz und die Sicherheit patientenbezogener Daten müssen gewährleistet sein, um das Vertrauen der Patienten in die medizinische Forschung zu stärken und ethische Standards zu wahren.

Forschende in der Inneren Medizin sehen sich angesichts der Vielfalt datenschutzrechtlicher Normen für den Forschungsbereich und deren Auslegungsspielräumen mit Unklarheiten konfrontiert. Dies hat eine bremsende und verhindernde Wirkung zur Folge, da Rechtsunsicherheit bei den Forschenden besteht und der Prozess der Datenerhebung und -analyse erschwert wird. Dies kann auch zu Lasten der Patienten gehen, denen so Innovationen in Diagnostik und Therapie vorenthalten werden.

Die DGIM sieht daher in der Entwicklung eines Leitfadens zum Umgang mit dem Datenschutz in der Forschung einen wichtigen Beitrag, um unter Wahrung der Patientenrechte das Potential von Gesundheitsdaten für die klinische Forschung und Versorgung zu erschließen.

5. Beratungsgrundlagen des Hessischen Beauftragten für Datenschutz und Informationsfreiheit (HBDI)

Im Bereich der medizinischen Forschung nimmt der HBDI gerne Beratungsanfragen von Forschenden an. Aber auch Eingaben von Betroffenen oder Berichterstattungen in der Presse können eine Befassung auslösen. In die Beratung fließen dabei auch die aktuellen Entwicklungen auf Bundesebene sowie auf europäischer Ebene ein, die der HBDI nach seinen Möglichkeiten mitgestaltet.

Auf politischer Ebene gibt es eine Vielzahl von Bestrebungen, die eine Förderung der Datennutzung und des Datenverkehrs im Gesundheitsbereich zum Ziel haben. Hervorzuheben ist die europäische Datenstrategie, die einen umfassenden Ansatz verfolgt und einen einheitlichen europäischen Datenraum für eine attraktive, sichere und dynamische Datenwirtschaft anstrebt. In diesem sollen die Potentiale der Datennutzung in unterschiedlichsten Bereichen erschlossen und eine Datenökonomie gefördert werden. Hierzu gehört auch der Europäische Raum für Gesundheitsdaten (EHDS), in dem sowohl die Primärnutzung als auch die Sekundärnutzung von Gesundheitsdaten neu geregelt werden.[4]

Die Anonymisierung personenbezogener Daten spielt in der medizinischen Forschung eine große Rolle. Für die damit verbundenen technischen, rechtlichen und ökonomischen Herausforderungen helfen die dogmatischen Ansätze eines absoluten oder relativen Begriffs personenbezogener oder anonymisierter Daten nicht weiter. Notwendig ist vielmehr, die Anforderungen an eine ausreichende Anonymisierung möglichst präzise zu bestimmen. Für diese Aufgabe wird es angesichts der enormen Breite der Anwendungen anonymisierter Daten nicht eine einzige Generallösung geben können. Vielmehr werden diese Anforderungen nach Anwendungsbereichen differenziert zu bestimmen sein. Für die Bestimmung der Risiken für die Verwendung anonymisierter Daten und der Anforderungen an einen ausreichenden Schutz sollten typische Szenarien der Erstellung und Verwendung anonymisierter Daten unterschieden und jeweils eine risikogerechte Anpassung an Maßnahmen zum Schutz der Grundrechte gesucht werden.

Da die DS-GVO hierfür keine einzelfallbezogenen Lösungen bietet, kommt den Datenschutzaufsichtsbehörden bei der Beratung eine besondere Rolle zu. Sie sollten eine führende und konstruktive Position einnehmen und dabei unterstützen, Wege hin zu einer datenschutzrechtskonformen Datennutzung zu beschreiten. Dies sollte in jedem Fall auch die Sensibilisierung für die Risiken der Datennutzung und -weitergabe sowie des Missbrauchs insbesondere für Rechte und Freiheiten der betroffenen Personen einschließen. Auf etwaigen Regelungsbedarf und -lücken sollte hingewiesen sowie an Vorschlägen zu deren Behebung mitgearbeitet werden.

[4] EU-Verordnung über den europäischen Raum für Gesundheitsdaten, 2025/327.

6. Datenschutzrechtliche Grundlagen in der Verarbeitung von Gesundheitsdaten

Bei der Verarbeitung personenbezogener und personenbeziehbarer Daten muss der jeweils Verantwortliche datenschutzrechtliche Anforderungen erfüllen. Keine Anwendung finden die datenschutzrechtlichen Regelungen demgegenüber für Daten, die sich nicht auf eine identifizierte oder identifizierbare Person beziehen lassen, also auf anonymisierte Daten.

Die Anonymisierung selbst ist eine Datenverarbeitung und bedarf daher einer Rechtsgrundlage.[5] In der Regel lässt sich bei der Anonymisierung zu Forschungszwecken eine Rechtsgrundlage finden.[6]

A. Forschung ohne Einwilligung

Es ist wichtig, dass sich Forschende im Vorfeld eines Forschungsvorhabens die Frage stellen, ob die Forschung mit anonymisierten – und in diesem Sinne beschränkten Daten – zur Erreichung des Forschungszieles geeignet ist. Wenn dem nicht so ist, sollten Alternativen angedacht werden. Hierzu gehört neben der Forschung mit Einwilligung die Forschung *ohne* Einwilligung. Diese ist für private Stellen in § 27 BDSG sowie für öffentliche Stellen in § 24 HDSIG geregelt. Der hier vorzunehmende Abwägungsprozess zwischen den Interessen der Betroffenen und dem Forschungsinteresse sollte bei Investigator initiierten Studien (IIT) in vielen Fällen zugunsten der Forschung ausfallen. Mit dem Gesundheitsdatennutzungsgesetz (GDNG) ist außerdem eine bundeseinheitliche Rechtgrundlage zur Forschung mit dem im Versorgungskontext erhobenen Behandlungsdaten hinzugekommen (siehe unten Fallbeispiel 2 B. II. 3 a (4)).

B. Anonymisierte Daten

Anonymisierung beschreibt den Prozess, potenziell identifizierende Faktoren aus einem Datensatz zu entfernen, so dass ein Personenbezug für die resultierenden Daten faktisch nicht mehr möglich ist, die Personen nicht re-identifizierbar sind.[7] Werden personenbezogene Daten in solcher Art in anonymisierte Daten überführt und dann weiterverarbeitet, finden datenschutzrechtliche Anforderungen keine Anwendung.[8] Dies bedeutet, dass diese Daten für Forschungszwecke aus datenschutzrechtlicher Sicht uneingeschränkt zur Verfügung stehen.

Ein absoluter Ausschluss einer Personenbeziehbarkeit von Gesundheitsdaten, die für Forschung und Versorgung verarbeitet werden sollen, ist weder gefordert, noch dürfte ein solcher stets möglich sein. Also kann es nur darum gehen, ein ausreichendes Maß an Risikoreduktion zu begründen.

Für die Feststellung, ob die zu verarbeitenden Daten personenbeziehbar sind, werden alle Mittel berücksichtigt, die eine Personenbezug herstellen können.[9] Andererseits soll

---

[5] nach Art. 4 Nr. 2 DS-GVO

[6] In Betracht kommen Art. 5 Abs. 1 lit. b i.V.m. Art. 6 Abs. 4 DS-GVO und Art. 6 Abs. 1 lit. c, e, Art. 9 Abs. 2 lit. j, Art. 89 DS-GVO i.V.m. nationalen Gesetzen (z.B. § 6 Abs. 3 S. 3 GDNG, § 27 BDSG oder § 24 HDSIG).

[7] ErwG 26 DS-GVO

[8] Konkretere Regelungen zur Anonymisierung personenbezogener Daten finden sich in der DS-GVO nicht, insbesondere auch nicht zu Anonymisierungsverfahren, zu Pflichten im Falle der Feststellungen der Personenbeziehbarkeit oder zur Verhinderung einer De-Anonymisierung.

[9] ErwG 26 DS-GVO; Die in der DS-GVO enthaltenen Erwägungsgründe haben keinen Rechtsnormcharakter, sie sollen allerdings bei der Auslegung und Interpretation der bestehenden

der Aufwand Berücksichtigung finden, mit dem eine Personenbeziehbarkeit herzustellen wäre. Die zum Zeitpunkt der Verarbeitung verfügbare Technologie und technologische Entwicklungen sollen ebenfalls berücksichtigt werden.

Aus einer Anonymisierung von Daten zu Forschungszwecken folgt regelmäßig kein hohes Risiko, sodass regelmäßig keine Datenschutzfolgenabschätzung (DSFA) notwendig ist.[10] Bei der im Rahmen der Anonymisierung zu erstellenden Risikoanalyse kann es aber sinnvoll sein, Bausteine einer DSFA, insbesondere zur Bewertung von Risiken, zu verwenden.[11]

Eine Anonymisierung kann durch die Minimierung der Risiken für Rechte und Freiheiten der betroffenen Personen (Patienten) die Nutzung von Patientendaten zu Forschungszwecken vereinfachen oder überhaupt erst ermöglichen.

Eine Wiederherstellung der Personenbeziehbarkeit kann für anonymisierte Daten in der Regel nicht vollständig ausgeschlossen werden. Prinzipiell besteht ein Risiko einer Re-Identifizierbarkeit. Risikoverstärkende Faktoren für eine Re-Identifizierbarkeit können beispielsweise die folgenden Punkte sein:

- der Einsatz eines nicht ausreichend wirksamen Anonymisierungsverfahrens,
- unvollständige Annahmen hinsichtlich des zum Datensatz erwerbbaren Zusatzwissens,
- Fehleinschätzungen im Zusammenhang mit den technischen Möglichkeiten zur Re-Identifizierung.

Für die Feststellung, dass sich durch die vorgenommene Anonymisierung das Risiko eines Personenbezugs ausreichend reduziert hat, sind auch risikomindernde Faktoren zu berücksichtigen:

- Zugriffskontrollen und Zugriffsbegrenzungen (z.B. geschützter Datenraum mit Use- & Access-Mechanismen, Verwendung von Treuhandmodellen),
- Datenverarbeitung in geschützter Umgebung (z.B. Federated Learning),
- Verbot gezielter Re-Identifizierung (strafbewehrt),
- Zeitnahe Löschung (ggf. können Daten auch bei einem Treuhänder verwahrt werden, sofern das Forschungsprojekt oder die Regeln guter wissenschaftlicher Praxis eine längere Aufbewahrung erfordert).

Für personenbezogene Gesundheitsdaten wird, im Sinne eines risikoverstärkenden Faktors, grundsätzlich ein hohes Missbrauchs- und Schadenspotenzial angenommen.[12] Umgekehrt kann sich eine Verarbeitung dieser Daten zu eingegrenzten Zwecken (z.B. KI-Training eines spezifischen KI-Systems und anschließende Löschung, Untersuchung ei-

Rechtsnormen helfen. Da andere Orientierungen fehlen, stellen sie aber meist einen wesentlichen und relevanten Faktor dar.

[10] Eine Datenschutz-Folgenabschätzung muss nach Art. 35 Abs. 1 DS-GVO immer dann vorgenommen werden, wenn die Datenverarbeitung voraussichtlich ein hohes Risiko für die Rechte und Freiheiten natürlicher Personen zur Folge hat.

[11] Art. 35 Abs. 7 DS-GVO enthält gesetzliche Anforderungen an den Inhalt einer Datenschutz-Folgenabschätzung.

[12] Art. 9 DS-GVO

ner spezifischen Fragestellung zu einem Krankheitsbild etc.) im Sinne eines risikomindernden Faktors begünstigend in der datenschutzrechtlichen Beurteilung des Forschungsvorhabens auswirken.[13]

Ob in Anbetracht risikoverstärkender oder -mindernder Faktoren eine ausreichende Risikoreduktion erfolgt ist, muss dann jeweils im Einzelfall von der verantwortlichen Stelle selbst bewertet werden. Verantwortlich ist die verantwortliche Stelle, die die Entscheidungshoheit über die Daten hat. Der Datenschutzbeauftragte der verantwortlichen Stelle ist im Regelfall unterstützend einzubinden.

Hierbei ist eine Dokumentation zur Erfüllung der Rechenschaftspflicht aus Art. 5 Abs. 2 DS-GVO notwendig. Im Hinblick auf die Dokumentation gibt es keine festen Formvorgaben, insbesondere was den Umfang oder die Genauigkeit betrifft. Die Ausführungen müssen in sich schlüssig und widerspruchsfrei sein. Mehr ist nicht gefordert. Die Forschung darf hierdurch nicht unnötig erschwert werden.

Ein hundertprozentiger Nachweis der Anonymität muss durch den Anwender nicht erbracht werden. Ein solcher Ansatz würde ansonsten sowohl dem breiten Spektrum möglicher Einsatzgebiete und deren Spezifika als auch den Herausforderungen der Anonymisierung nicht gerecht werden. Hier bestünde nicht zuletzt das Risiko, mögliche Einsatzgebiete der Anonymisierung durch die Formulierung von zu restriktiven Anforderungen von vornherein auszuschließen. Dies würde der europäischen Datenstrategie widersprechen und die Erschließung von Potentialen in unterschiedlichen Bereichen verhindern.

Die vorangegangenen Ausführungen machen deutlich, dass die Feststellung, ob Daten hinreichend anonym sind, die Bewertung von bestehenden Risiken für die Betroffenen unter Berücksichtigung der Gesamtumstände im Einzelfall die Art der Daten und Datenverarbeitungsprozesse bestimmt. Hierbei ist auch eine differenzierte Betrachtung der unterschiedlichen Einsatzgebiete der Anonymisierung erforderlich.

C. Anwendungsbereich medizinische Forschung

In der medizinischen Forschung können anonymisierte Daten in unterschiedlichen Disziplinen und Teilgebieten zum Einsatz kommen.

Hierzu müssen die Forschenden je nach Forschungsgegenstand auf unterschiedliche Patientendaten zurückgreifen. Bei solchen Patientendaten handelt es sich um Gesundheitsdaten und somit um besondere Kategorien personenbezogener Daten.[14]

Gerade bei umfangreichen Daten, etwa zu Patienten- und Medikationshistorien, werden hohe Anforderungen an die eingesetzten Anonymisierungsverfahren gestellt. Gleichzeitig muss bei ihrer Anwendung sichergestellt werden, dass die resultierenden anonymisierten Daten für den Forschungszweck weiterhin geeignet sind.

D. Sonderfall: Pseudonymisierung

Falls ein Forschungsvorhaben mit anonymisierten Daten nicht erreicht werden kann, so sollten grundsätzlich pseudonymisierte Daten verwendet werden, bei denen in Kenntnis des Pseudonyms die Personen re-identifizierbar sind.[15]

---

[13] vgl. Roßnagel, Anonymisierung personenbezogener Daten und Nutzung anonymer Daten, DuD 2024, 513 (519).
[14] gemäß Art. 9 DS-GVO
[15] Siehe Art. 89 Abs. 1 S. 3 DS-GVO; Voraussetzung für die Nutzung pseudonymisierter Daten ist, dass eine datenschutzrechtliche Rechtsgrundlage anwendbar ist.

Sollte ein Forschungsvorhaben erfolgreich sein und beispielsweise ein neues Heilverfahren für eine Erkrankung entwickelt werden, so könnte dieses auch Patienten zugutekommen, deren Daten in anonymisierter Form Eingang in die Forschung gefunden haben. Dies würde zur Kontaktaufnahme jedoch zumindest eine Identifikation der relevanten Patienten und somit eine Re-Identifizierung erfordern. Hierfür würden sich pseudonymisierte Daten eignen, deren Zuordnungsinformationen bei einem Treuhänder hinterlegt sind und die – eventuell nur nach Beteiligung mehrerer Stellen – re-identifiziert werden können. Der Treuhänder könnte rechtlich in besonderer Weise verpflichtet sein.

Der Europäische Datenschutzausschuss (EDSA) hat in seinen Leitlinien zur Pseudonymisierung Informationen und Hilfestellungen zum Begriff der Pseudonymisierung veröffentlicht.[16]

7. Fallbeispiel 1: Training eines Modells künstlicher Intelligenz (KI) zur Unterstützung eines kontinuierlichen Monitorings von Qualitätsparametern und zur Erstellung von Befunden in der Darmkrebsvorsorge

A. Sachverhalt und zu klärende Fragestellung

Medizinisch-wissenschaftlicher Hintergrund/Notwendigkeit

Mit Hilfe von auf KI basierenden Algorithmen lässt sich im Bereich der Darmkrebsvorsorge in der Gastroenterologie die Qualitätssicherung erheblich beschleunigen und objektivieren: Anstelle von Überprüfungen zu festen Audit-Zeitpunkten wird ein kontinuierliches Monitoring von Qualitätsparametern ermöglicht, so dass z.B. rechtzeitig vor negativen Veränderungen der Qualitätsparameter gewarnt wird.

Ebenfalls können solche Algorithmen anhand der während der Untersuchungen erhobenen Bilddaten Befunde automatisch erstellen und somit den Dokumentationsprozess erheblich beschleunigen.

Der Einsatz der KI-basierten Algorithmen trägt also zur Erhöhung der Versorgungsqualität und Schonung knapper ärztlicher Ressourcen bei.

Nutzung von / Umgang mit medizinischen Daten

Um die o.g. KI zu trainieren und die Diversität der Befunde mit den dazugehörigen Bilddaten im Training zu erlernen, wird eine große Menge an frei verfügbaren anonymisierten Datensätzen, die Endoskopiebefunde mit histopathologischen Befunden, Endoskopiebildern und -videos verbinden, aus verschiedenen Einrichtungen benötigt. Solch öffentlich zugängliche Datensätze verschiedener Zentren können KI-Lösungen deutlich fördern bzw. oft erst ermöglichen.

Das Ziel des vorliegenden Projektbeispiels ist die Zusammenführung und frei zugängliche Veröffentlichung anonymisierter Datensätze aus verschiedenen klinischen Zentren (pro Zentrum und Jahr ca. 1000 bis 4000). Hierfür sollen Daten anonymisiert und dann projektbezogen genutzt werden, die sowohl im Rahmen der klinischen Routine als auch im Rahmen von Investigator Initiated Trials (ITT) erhoben wurden. Umfasst sind dabei

[16] https://www.edpb.europa.eu/our-work-tools/documents/public-consultations/2025/guidelines-012025-pseudonymisation_de

Befundberichte, verlinkt zu feingeweblichen (histopathologischen) Analysen, Endoskopiebildern und -videos. Die anonymisierten Datensätze aus mehreren Zentren sollen zusammengeführt werden, um Diversität abbilden zu können.

Datenschutz-rechtliche Probleme / Unsicherheiten

Derzeit existieren große, frei verfügbare Datensätze zu diesem Thema in Deutschland nicht, obwohl solche Daten verfügbar sind und zur aktiven Forschung daran benötigt werden. Um solche Datensätze frei verfügbar zu machen, müssen die folgenden Fragen beantwortet werden:

- Ist ein Endoskopiebild/ -video (d.h. Bilddaten aus dem Inneren des Körpers) als anonym zu bewerten?
- Wie muss ein Befundbericht gestaltet sein, um als anonym eingestuft werden zu können?
- Dürfen retrospektive Daten aus der klinischen Routine ohne Einwilligung der betreffenden Patienten für die angeführte Fragestellung verwendet werden?
- Welche Besonderheiten gelten für die Nutzung der Daten bereits verstorbener Patienten für das Projekt?
- Welche Folgen hat die denkbare Re-Identifizierung veröffentlichter anonymisierter Daten durch den Patienten oder einen behandelnden Arzt, welche jeweils prinzipiell über das hierfür erforderliche Zusatzwissen verfügen?

Angewendete Maßnahmen zur Anonymisierung

Für den vorliegenden Fall wurden folgende Bearbeitungsschritte zur Anonymisierung der Bilder durchgeführt:

- sensible Stellen (s. unten) wurden immer geschwärzt oder abgeschnitten und absichtlich nicht nur verpixelt,
- für IDs wurden durchgehend Zufallszahlen verwendet,
- der Bilddateiname wurde durch eine zufällig generierte Untersuchungs-ID aus dem Befundbericht ersetzt und mit einer laufenden Nummer versehen,
- Metadaten des Bildes wurden entfernt,
- Der Name des Patienten wurde aus dem Bild entfernt,
- Das Datum der Untersuchung wurde mit Ausnahme des Untersuchungsmonats und -jahrs entfernt (analog zum Untersuchungsbericht),
- Es wurde sichergestellt, dass nur Inhalte aus dem Inneren des Körpers auf dem Bild zu sehen sind,
- Die Namen der Untersuchenden wurden entfernt,
- Der Name des Zentrums wurde entfernt.

Folgende Bearbeitungsschritte wurden zur Anonymisierung der Befundberichte durchgeführt:

- Vergabe zufällig generierter IDs für Zentrum, Patienten, Untersucher und Untersuchung,
- Das Datum der Untersuchung wurde auf Monat und Jahr beschränkt,
- Das Alter des Patienten wurde auf ganze Jahreszahlen begrenzt,
- im Befundbericht wurden kein Name oder Initialen des Patienten verwendet.

Die Abhängigkeiten der neuen zufällig generierten IDs untereinander blieben, wie im Originaldatensatz erhalten. Sprich eine Patienten-ID kann mehrere Untersuchungs-IDs haben. Diese wiederum hat Bilder, die aus einer Untersuchungs-ID und einer fortlaufenden Nummer bestehen.

Exemplarischer Befundbericht

Zentrum: 123
Patienten-ID 8923023A479
Untersucher-ID: 39213821KD12
Datum der Untersuchung: August, 2022
Geschlecht: männlich, Alter 62
Untersuchung: Koloskopie
Gerät: Olympus CF-HQ190L
Anamnese: Vorsorgedarmspiegelung
Indikation: Kontrollkoloskopie - Polypen?
Prämedikation: 200 mg Propofol i. v. (fraktioniert). Pulsoxymetrische- +
Kreislaufüberwachung. 3 l Sauerstoff über Nasensonde.
Maximale Einsicht: Terminales Ileum.
Anus: Es zeigen sich mehrere reizlose Marisken.
Colon: Im Coecum Kaltschlingenabtragung von einem sessilen Polypen < 5mm.
Histologie: PE 1: Polyp Coecum
Komplikation: nein
Beurteilung vorläufig: Reizlose Marisken. (K64.4)
Empfehlung: Kontrolle in 5 Jahren
Histologie: Fraktion 1: Hyperplastischer Polyp der Dickdarmschleimhaut (Polyp
Coecum, laut klinischen Angaben) mit angrenzend regelhafter Dickdarmschleimhaut.
Keine schweren Epitheldysplasien. Kein invasives Karzinomwachstum.

B. Datenschutzrechtliche Diskussion des Fallbeispiels

Ist ein Bild aus dem Inneren des Körpers (z.B. ein Endoskopiebild/-video) als anonym zu bewerten?

Für die Beantwortung der Frage, ob ein Bild aus dem Inneren des Körpers als anonym zu bewerten ist, ist es entscheidend, wie die jeweilige Fachdisziplin selbst das Risiko der Re-Identifizierbarkeit einordnet. Im Falle eines Lichtbildes aus der Radiologie, das einen Lungenflügel zeigt, ist das maßgebliche Kriterium der Re-Identifizierbarkeit u. U. anders zu bewerten als im Falle eines Lichtbildes aus dem Inneren des Körpers.

Fragen, die hier innerhalb der jeweiligen Fachdisziplin zu stellen und durch den Forscher zu beantworten und darzulegen sind, wären etwa die Folgenden:

- Kann man anhand des Bildes aus dem Inneren des Körpers den Patienten ohne die Zuhilfenahme von durch jedermann erwerbbares Zusatzwissen unter mehreren herausfiltern?
- Gibt es eine Vielzahl an identischen Bildern aus dem Inneren des Körpers, bei denen der Unterschied ohne die Zuhilfenahme von durch jedermann erwerbbares Zusatzwissen erkennbar ist?

- Wie stark wächst das Risiko der Re-Identifizierbarkeit durch die weiteren Merkmale, die verarbeitet werden, in der Gesamtschau an?
- Können Datensätze in größeren Gruppen zusammengefasst werden oder müssen individuelle Einzeldatensätze bestehen bleiben?

Sofern sich hier jeweils ein für die Re-Identifizierung risikoerhöhender Faktor zeigt, kann dies durch ein Mehr an Schutz der Betroffenenrechte, wie eingangs dargestellt, ausgeglichen werden. Die Einschränkung des Nutzerkreises von einer Veröffentlichung an die Allgemeinheit zu einem beschränkten Zugang für ausgewählte Forschende kann z.B. zur Reduzierung der Risiken eingesetzt werden.

Der Nachweis über eine absolute Anonymisierung, die den Wert 100 % erreicht, kann dieser Ansicht nach nicht gefordert werden. Anderenfalls wäre für jedes medizinische Forschungsvorhaben, das anonyme Daten zum Gegenstand haben möchte, ein wissenschaftliches Gutachten erforderlich. Da die DS-GVO hier keine definitiven Kriterien an die Hand gibt, muss diese Wertung zugunsten des Forschungsprivilegs ausfallen.

Eine Re-Identifizierung eines Patienten durch den behandelnden Arzt anhand von ihm erhobener Befunde ist grundsätzlich anzunehmen. Aus Sicht der DGIM gilt die Anonymisierung hierbei dennoch als vollzogen. Der behandelnde Arzt wurde vom Patienten berechtigt, den Befund zu erheben und unterliegt der ärztlichen Schweigepflicht.

Einschätzung der DGIM: Möglichkeit der Re-Identifizierung anhand von Bildern aus dem Inneren des Körpers

Nach Einschätzung der DGIM besteht für Bilder aus dem Inneren des Körpers* (wie hier im Beispiel Endoskopiebilder / -videos aus dem Dickdarm) die Möglichkeit der Re-Identifizierung eines Patienten nur für diejenigen Personen, die als Zusatzwissen über das exakt selbe Vergleichsbild (im Beispiel also das Endoskopiebild/ -video aus dem Dickdarm) des Patienten sowie in Kombination damit über dessen Identität verfügen. Bei Personen mit diesem Zusatzwissen handelt es sich jedoch immer um Personen, denen die Identität in Kombination mit dem medizinischen Befund sowieso bekannt sind: Behandelnde Ärztinnen und Ärzte (und deren Assistenzpersonal) sowie der Patient / die Patientin selbst. Dies trifft auch dann zu, wenn sich ein behandelnder Arzt/Ärztin in Einzelfällen anhand besonders ausgefallener Befunde (Bilder) an die Identität eines Patienten erinnert, der ihm aber auch in diesem Fall ja schon bekannt ist.

Zusammenfassend erscheint also aus Sicht der DGIM das Risiko der Re-Identifizierung eines Patienten anhand von Bildern aus dem Inneren des Körpers durch Personen ohne Zusatzwissen als vernachlässigbar klein. Das Risiko besteht nur für Personenkreise, denen die Information zulässigerweise sowieso schon bekannt ist.

Im Unterschied dazu kommt solchen Bilddaten eine besondere Bedeutung zu, auf denen das Gesicht bzw. Teile des Gesichts des Patienten enthalten sind. Für solche Bilddaten muss aufgrund der vielfach öffentlich vorhandenen und technisch vergleichsweise einfach zugänglichen identifizierenden Bilder und Begleitinformationen über das Internet im Sinne eines durch jedermann erwerbbaren Zusatzwissens ein deutlich erhöhtes Risiko einer Re-Identifizierung durch jedermann angenommen werden. Dies ist prinzipiell auch denkbar für Bilder bildgebender Verfahren (z.B. Schnittbilder oder auch konventionelles Röntgen), die eine Rekonstruktion des Kopfes/Gesichts erlauben. Bei bildgebenden Verfahren, die andere Körperregionen abbilden und u.U. eine Rekonstruktion von Körper-

konturen erlauben, erscheint ein gewisses Risiko der Re-Identifizierung auch durch Personen ohne spezifisches Zusatzwissen denkbar. Hier ist eine Prüfung im Einzelfall erforderlich.

*Hierunter fallen im Sinne dieser Ausführungen endoskopische und histopathologische Bildaufnahmen.

Durch den Forscher wurde dargelegt, dass eine Bewertung des Risikos der Re-Identifizierung erfolgte und dies durch spezifische Gegenmaßnahmen ausgeglichen wurde. Für den HBDI reicht hier zunächst die Feststellung aus, dass von der verantwortlichen Stelle eine sorgfältige und nachvollziehbare Risikobewertung getroffen wurden. Die Bewertung, ob die Risiken einer Re-Identifizierung ausreichend gering sind, obliegt der verantwortlichen Stelle, in der Regel also den Forschenden. Diese Prüfung ist zu dokumentieren, um die Nachweispflichten aus Art. 5 Abs. 2 DS-GVO zu erfüllen und der Aufsichtsbehörde diese Dokumentation im Falle einer Prüfung vorlegen zu können.

Wie muss ein Befundbericht gestaltet sein, um als anonym eingestuft werden zu können?

Befundberichte können dergestalt bearbeitet werden, dass sie als anonym gelten können. Hierfür ist es wichtig, eindeutige Identifier aus dem Bericht zu entfernen, die nach dem Schlüssel-Schloss Prinzip nur zu einer Person passen können. Hierunter fallen in der Regel Patienten-IDs, die auch an anderer Stelle hinterlegt sind und den Zugriff zu weiteren Daten ermöglichen.

Zudem empfiehlt sich eine Betrachtung, ob Daten in der Gesamtschau den Personenkreis zu stark einengen, so etwa das Erfassen des genauen Aufnahme- und Entlassungsdatums. Hier wird es in der Regel nur wenige Personen geben, bei denen diese Daten übereinstimmen.

Auf die Angabe des kompletten Geburtsdatums ist generell zu verzichten (TT.MM.JJ). Die Altersangabe ist insoweit immer vorzuziehen (z. B. 62 Jahre), wenngleich auch hier die Möglichkeit einer Vergröberung geprüft werden sollte (60–65 Jahre, 50–55 Jahre). Hiervon sollte u. U. auch Gebrauch gemacht werden, wenn die Diagnosen und die empfohlenen Behandlungsmaßnahmen in der Exaktheit nur auf wenige Patienten zutreffen. Hier könnte das exakte Alter sonst der entscheidende Identifier sein. Auch das Datum der Untersuchung könnte heruntergebrochen werden (statt „August 2025" z.B. „Juli bis August 2025“).

Dürfen retrospektive Daten aus der klinischen Routine ohne Einwilligung des Patienten für die angeführte Fragestellung verwendet werden?

Personenbezogene Daten aus der klinischen Routine dürfen retrospektiv auch ohne Einwilligung für Forschungszwecke verwendet werden, wenn hierfür eine tragfähige gesetzliche Rechtsgrundlage vorliegt. Hierfür kommen insbesondere die Normen § 27 BDSG sowie § 24 HDSIG in Betracht, in welchen ein Forschungsprivileg zum Ausdruck kommt, das sich über die DS-GVO hinaus in vielen Bereichen wiederfindet. Diese Normen ermöglichen die Forschung ohne Einwilligung, wenn zuvor eine Abwägung der Forschungsinteressen mit den Interessen der Betroffenen erfolgt ist, und diese zugunsten der Forschungsinteressen ausfällt.

Außerdem kommt auch § 6 Abs. 1 Nr. 2 GDNG als Rechtsgrundlage für die retrospektive Auswertung von Daten aus der klinischen Routine in Betracht (siehe Abschnitt 10.B).

Bei allen Forschungsvorhaben sollte generell geprüft werden, ob die Forschung mit Einwilligung der Patienten umsetzbar und im Sinne des Forschungsvorhabens möglicherweise zielführender sein könnte.

Eignet sich für das Forschungsvorhaben dahingegen das Einholen einer Einwilligung aus diversen Gründen nicht, sollte das Vorhaben mit einem Datenschutzkonzept abgesichert werden. Auch damit kann sichergestellt werden, dass die Rechte der Betroffenen trotz fehlender, ausdrücklicher Zustimmung ausreichend Berücksichtigung finden.[17] Ein Datenschutzkonzept muss bei Forschungsvorhaben auf Basis von 24 HDSIG erstellt werden.[18] Besondere gesetzliche Anforderungen an Form und Umfang eines Datenschutzkonzeptes sind nicht gesetzlich vorgesehen. Ein Datenschutzkonzept sollte die wesentlichen datenschutzrechtlichen Fragestellungen beantworten (Datenflüsse, Verantwortlichkeiten, Rechtsgrundlagen, Betroffenenrechte, technische und organisatorische Maßnahmen etc.).

Welche Besonderheiten gelten für die Forschung mit Daten bereits verstorbener Patienten?

Für Daten verstorbener Personen gilt die DS-GVO nicht mehr. Die ärztliche Schweigepflicht und das postmortale Persönlichkeitsrecht sind aber zu beachten.

Da hier nicht mehr mit Einwilligungen gearbeitet werden kann, sind Daten verstorbener Patienten im Grunde besonders für eine Verarbeitung in anonymisierter Form geeignet, sofern diese Daten beforscht werde sollen und für ein bestimmtes Forschungsprojekt relevant sind.

Welche Folgen hat die prinzipiell denkbare Re-Identifizierung veröffentlichter anonymisierter Daten durch den Patienten oder einen behandelnden Arzt, welche jeweils über das hierfür erforderliche Zusatzwissen verfügen?

In einem Szenario, in dem Bilddaten ausschließlich von der betroffenen Person re-identifiziert werden können, sind diese Daten zunächst für Dritte ohne diese Zusatzinformationen anonym – jedenfalls, solange der Patient/die Patienten die Zusatzinformation nicht weitergibt oder veröffentlicht. Das Risiko der Re-Identifizierung ist daher als gering zu betrachten, da der Patient hier selbst Herr darüber ist, inwieweit er sein spezielles Zusatzwissen preisgibt. Besondere Folgen ergeben sich daher nicht aus dieser Konstellation.

Für den Europäischen Gerichtshof (EuGH) ist eine Identifizierung insbesondere dann nicht nach allgemeinem Ermessen wahrscheinlich, wenn die „Herausgabe“ des Zusatzwissens an Dritte gesetzlich verboten ist bzw. keine rechtlichen Mittel zur Erlangung des Zusatzwissens bestehen.[19]

Dies ist hier für den weiteren potenziellen Inhaber von Zusatzwissen der Fall, also für den Arzt oder die Ärztin. Denn diese sind nach dem Berufs- und Strafrecht zur Verschwiegenheit verpflichtet und dürfen Dritten nicht das identifizierende Zusatzwissen zukommen lassen. Rechtliche Mittel Dritter, um an die von der Schweigepflicht geschützten Daten zu kommen, bestehen nicht. Im Gegenteil, die Daten sind sogar vor Zugriffen

[17] Ein entsprechendes Muster zur Orientierung findet sich auf der Homepage des HBDI (siehe https://datenschutz.hessen.de/datenschutz/statistik-und-wissenschaft/datenschutzkonzepte-fuer-akademische-abschlussarbeiten-oder-promotionsvorhaben).
[18] § 24 Abs. 1 S. 3 HDSIG
[19] EuGH, Urt. v. 19.10.2016, Rs. C-582/14 (Breyer), ECLI:EU:C:2016:779, Rn. 46

durch staatliche Stellen besonders geschützt.[20] Auch hier wird also im Ergebnis davon auszugehen sein, dass das Risiko der Re-Identifizierung gering ist. Im Ergebnis bleibt also auch diese Konstellation ohne Folgen für die zuvor getroffenen Bewertungen.

8. Fallbeispiel 2: Künstliche Intelligenz (KI) Anwendungen in der Pathologie

Medizinisch-wissenschaftlicher Hintergrund / Notwendigkeit

In mehreren Drittmittel-geförderten Forschungsprojekten wurden KI-Verfahren an histopathologischen Schnittpräparaten angewandt, um klinisch relevante Parameter vorherzusagen und dadurch die Prognoseabschätzung oder die Vorhersage des Therapieerfolgs zu verbessern.

Nutzung von / Umgang mit medizinischen Daten

Bei den zum Training der KI-Modelle verwendeten Daten handelt es sich um digitale Bilddaten (Whole-Slide-Images), welche von analogen histopathologischen Schnittpräparaten generiert werden.* Als Zielvariablen werden je nach Fragestellung verschiedene klinisch relevante Parameter definiert (z. B. Tumordiagnose, Überlebenszeit, Therapieansprechen oder molekulare Marker). Das Training der Modelle erfolgt lokal („On-Prem“) und nach Anonymisierung oder, wenn zulässig, nach mehrfacher Pseudonymisierung der Daten. Neuere Ansätze basieren darauf, die gelernten Modellparameter mehrerer Institutionen während des Trainings dynamisch auszutauschen. Beim so genannten föderierten Lernen bzw. Schwarmlernen werden z. B. nur die trainierten Gewichte ausgetauscht. Dies kann verschlüsselt und unter Verwendung der Blockchaintechnologie erfolgen, um eine zusätzliche Sicherheit zu schaffen[21] Am Ende profitieren alle Schwarmteilnehmer von dem gemeinsamen Lernfortschritt, ohne dass die Trainingsdaten weitergegeben werden müssen. In einem aktuellen Projekt wird das Verfahren außerdem verwendet, um synthetische Daten zu generieren, welche dann für die Forschung nutzbar gemacht werden können.[22]

*Obgleich das Gewebe natürlich von einem individuellen Patienten stammt, ist es im Gegensatz zu anderen medizinischen bildgebenden Verfahren (z. B. Schädel-CT oder dermatologischen Aufnahmen) nach dem heutigen Stand der Technik nicht möglich, allein vom Aussehen der einzelnen Zellen unter dem Mikroskop auf die jeweilige Person zurückzuschließen.

Datenschutzrechtliche Probleme und Unsicherheiten

[20] Durch das Zeugnisverweigerungsrecht (§ 53, 53a StPO) und das Beschlagnahmeverbot (§ 97 StPO).
[21] s. hierzu: Saldanha et al. (2023). Direct prediction of genetic aberrations from pathology images in gastric cancer with swarm learning. Gastric Cancer, 26(2), 264-274.; Saldanha et al (2022). Swarm learning for decentralized artificial intelligence in cancer histopathology. Nat Med, 28(6), 1232-1239.; Sieling et al. (2025). Urheber- und datenschutzrechtliche Aspekte zur Nutzung von Bildern aus der Pathologie in Social Media. Die Pathologie, 46(4), 207-212.; Truhn et al. (2024). Encrypted federated learning for secure decen-tralized collaboration in cancer image analysis.
[22] Schulz et al. 2025 (in preparation)

Datenschutz spielt in der Durchführung der skizzierten Projekte eine zentrale Rolle. Gleichzeitig bestehen bei vielen Wissenschaftlern erhebliche Unsicherheiten, die zunehmend als Belastung empfunden werden. In der Praxis kann dies dazu führen, dass wertvolle Datensätze ungenutzt bleiben oder vielversprechende Forschungsansätze nicht weiterverfolgt werden.

Insbesondere forschende Ärzte befinden sich hier in einem Dilemma: Einerseits könnten KI-gestützte Methoden zu erheblichen Fortschritten in der medizinischen Versorgung – etwa bei der Behandlung von Krebspatienten – beitragen. Es wäre ethisch kaum vertretbar, diese Möglichkeiten ungenutzt zu lassen, da dies potenziell zu vermeidbarem Leid oder einer Verlängerung von Krankheitsverläufen führen könnte. Andererseits wird von ihnen erwartet, einen „allumfassenden“ / „100%igen“ Datenschutz sicherzustellen – ein Anspruch, der kaum vollständig erfüllbar ist.[23] Dies gilt umso mehr in einem dynamischen Umfeld, in welchem sich sowohl die technologischen Rahmenbedingungen als auch der Rechtsrahmen laufend verändern.

Verfahren wie das föderierte Lernen bieten zwar potenzielle Lösungen, sind jedoch mit erheblichem technischem und organisatorischem Aufwand verbunden – eine Hürde, die nicht jede Forschungsgruppe problemlos überwinden kann. Zudem stoßen auch diese Ansätze an ihre Grenzen, etwa bei der Entwicklung großer KI-Modelle (z. B. Large Language Models oder Foundation Models) oder bei multimodalen KI-Anwendungen.

Um datenschutzrechtlichen Anforderungen gerecht zu werden, setzt die betreffende Arbeitsgruppe derzeit konsequent auf lokal betriebene („On-prem“) Infrastruktur und arbeiten mit anonymisierten oder mehrfach pseudonymisierten Daten. Für die verwendeten Kohorten und Verfahren liegen positive Ethikvoten der zuständigen Kommission vor. Bei Unklarheiten – beispielsweise im Rahmen eines Healthcare Hackathons – erfolgt eine Abstimmung mit dem lokalen Datenschutzbeauftragten; bei Bedarf werden zusätzliche Rechtsgutachten eingeholt. Dieser Ansatz hat sich bislang als tragfähig erwiesen.

B. Datenschutzrechtliche Diskussion des Fallbeispiels

Unter welchen Voraussetzungen sind histopathologische Bilddaten als anonym einzustufen?

Für die histopathologische Bilddaten sind die im Abschnitt 7.B beschriebenen Anforderungen an die Anonymisierung zu untersuchen.

Dabei ist zu prüfen, ob Dritte über Zusatzwissen verfügen, mit dessen Hilfe eine Re-Identifizierung einzelner Patientinnen oder Patienten durchgeführt werden kann. Bei histopathologisches Bilddaten ist positiv zu berücksichtigen, dass keine dauerhaften anatomischen Besonderheiten (wie z.B. bei Thorax-CT) erkennbar sein dürften und so eine Verbindung zu einer konkreten Person deutlich schwieriger ist. Ein wichtiger Aspekt bei der Bewertung der Re-Identifizierungsrisiken ist auch, ob die anonymisierten Daten veröffentlicht werden oder nur innerhalb einer Forschungskooperation genutzt werden.

Aus Sicht der DGIM erlaubt die reine Bildinformation eines histopathologischen Schnittbilds keine Rückschlüsse auf den betreffenden Patienten.

[23] Sieling & von Petersdorff-Campen, 2025

Welche Anforderungen bestehen an Austausch und Zusammenführung von Daten im Rahmen von Forschungskooperationen? Welche Anforderungen ergeben sich aus der Art der Kooperationsteilnehmer?

Auch bei Forschungskooperation muss für jede Übermittlung von personenbezogenen Daten an einen anderen Verantwortlichen eine datenschutzrechtliche Rechtsgrundlage vorliegen. Dies gilt auch dann, wenn es sich um eine gemeinsame Verantwortlichkeit der Beteiligten nach Art. 26 DS-GVO handelt.

Bei einer solchen gemeinsamen Verantwortlichkeit muss ein Vertrag nach Art. 26 Abs. 2 DS-GVO geschlossen werden und es ist darauf zu achten, dass die Verarbeitungsvorgänge unter gemeinsamer Verantwortlichkeit (z.B. Speicherung in gemeinsamer Datenbank) von den Bereichen der getrennten Verantwortlichkeit (z.B. Erhebung von Daten aus Routineversorgung) sauber getrennt werden.

Je nach Art der Kooperationsteilnehmer und der konkreten Datenverarbeitungen ist zu untersuchen, ob es sich eine gemeinsame Verantwortlichkeit, eine getrennte Verantwortlichkeit oder eine Auftragsverarbeitung (Art. 28 DS-GVO) handelt. Bei der Einbindung eines technischen Dienstleisters, der eine in der Cloud betriebene Software bereitstellt, kann es sich beispielsweise um eine Auftragsverarbeitung handeln.

Hinsichtlich des KI-Modells muss datenschutzrechtlich zwischen der Entwicklungsphase und der Einsatzphase getrennt werden.[24] Für die Entwicklung eines KI-Modells (Erhebung Trainingsdaten und Training) können andere datenschutzrechtliche Rechtsgrundlagen gelten (insb. auch zu wissenschaftlichen Forschungszwecken) als für den späteren Einsatz in der medizinischen Versorgung. Es ist auch zu prüfen, ob das KI-Modell selbst personenbezogene Daten enthält.

Das Training mit anonymen Daten ist datenschutzrechtlich vorzugswürdig.

Für welche Gruppen von Forschungsgegenständen empfiehlt sich aus Sicht der Forschung der Einsatz von besonderen Verfahren wie Schwarmlernen, federated learning oder Blockchain-Technologie?

Datenschutzfreundliche Technologien (engl. Privacy Enhancing Technologies, PET) wie Schwarmlernen und föderiertes Lernen (engl. federated learning, FL) sollten in der medizinischen Forschung weiter erprobt werden, um einen idealen Ausgleich zwischen Forschungsinteressen und Datenschutz erzielen zu können. Häufig ist eine dezentrale Datenverarbeitung datenschutzrechtlich gegenüber einer Verarbeitung in einer zentralen Instanz vorzugswürdig. Vor jedem Forschungsvorhaben sollte geprüft werden, wie das Vorhaben möglichst datensparsam und datenschützend umgesetzt werden kann. Die oben genannten Verfahren eignen sich nicht für alle Vorhaben, gerade bei besonders sensiblen Datensätzen und größeren Forschungsvorhaben sollten sie aber in Betracht gezogen werden.

Föderiertes Lernen

Die Genauigkeit von KI-Modellen hängt maßgeblich von der Quantität (und Qualität) der verfügbaren Trainingsdaten ab.[25] Da Gesundheitsdaten jedoch besonders schützenswert sind, können sie aus rechtlichen, ethischen und organisatorischen Gründen nicht

[24] EDSA Stellungnahme 28/2024 zu gewissen Datenschutzaspekten der Verarbeitung personenbezogener Daten im Zusammenhang mit KI-Modellen; https://www.edpb.europa.eu/our-work-tools/our-documents/opinion-board-art-64/opinion-282024-certain-data-protection-aspects_en

[25] Bahri, Y., Dyer, E., Kaplan, J., Lee, J., and Sharma, U. (2024). Explaining neural scaling laws. Proc Natl Acad Sci U S A 121, e2311878121. 10.1073/pnas.2311878121

ohne weiteres zwischen Institutionen geteilt oder in zentralen Repositorien (z. B. über Cloud-Services) zusammengeführt werden. Hier setzt das Konzept des föderierten Lernens an. Anstatt die Daten selbst übertragen zu müssen, werden die Modelle lokal trainiert und es werden lediglich die Modellparameter übermittelt, aggregiert und ausgetauscht, sodass ein globales Modell entsteht.

Ziel ist es eine multizentrische Kooperation bei gleichzeitiger Wahrung der Datensouveränität und -sicherheit zu ermöglichen. So können selbst kleine Kohorten (z. B. bei seltenen Erkrankungen) an mehreren Standorten verwendet werden, um eine größtmögliche Modelgenauigkeit zu erzielen. In Studien konnte gezeigt werden, dass an medizinischen Daten trainierte FL-Modelle genauso leistungsfähig sind wie zentral trainierte Modelle (Non-Inferiority).[26]

Verschiedene Verfahren wie z. B. besondere Verschlüsselungstechniken können die Sicherheit zusätzlich erhöhen.[27]

Ein zentrales Hindernis für den Einsatz von Föderiertem Lernen sind deutlich komplexere IT- und Netzwerkanforderungen. Jede beteiligte Einrichtung benötigt eigene Rechenressourcen für das lokale Training sowie umfangreiche Software- und Sicherheitsumgebungen. Hinzu kommt ein hoher Kommunikationsaufwand: Modellparameter müssen in einzelnen Runden zwischen den Standorten ausgetauscht werden, was Bandbreite und stabile Netzwerke erfordert. Die Orchestrierung wird durch lokale Ausfälle, Verzögerungen und unterschiedliche Hardware erschwert. Zusätzlich sind erweiterte Sicherheitsmaßnahmen wie verschlüsselte Übertragungskanäle und/oder VPNs nötig. Daher wird Föderiertes Lernen bislang meist nur in spezialisierten Konsortien durchgeführt.

[26] Saldanha, O.L., Quirke, P., West, N.P., James, J.A., Loughrey, M.B., Grabsch, H.I., Salto-Tellez, M., Alwers, E., Cifci, D., Ghaffari Laleh, N., et al. (2022). Swarm learning for decentralized artificial intelligence in cancer histopathology. Nat Med 28, 1232-1239. 10.1038/s41591-022-01768-5; Warnat-Herresthal, S., Schultze, H., Shastry, K.L., Manamohan, S., Mukherjee, S., Garg, V., Sarveswara, R., Handler, K., Pickkers, P., Aziz, N.A., et al. (2021). Swarm Learning for decentralized and confidential clinical machine learning. Nature 594, 265-270. 10.1038/s41586-021-03583-3; Dayan, I., Roth, H.R., Zhong, A., Harouni, A., Gentili, A., Abidin, A.Z., Liu, A., Costa, A.B., Wood, B.J., Tsai, C.S., et al. (2021). Federated learning for predicting clinical outcomes in patients with COVID-19. Nat Med 27, 1735-1743. 10.1038/s41591-021-01506-3

[27] Truhn, D., Tayebi Arasteh, S., Saldanha, O.L., Muller-Franzes, G., Khader, F., Quirke, P., West, N.P., Gray, R., Hutchins, G.G.A., James, J.A., et al. (2024). Encrypted federated learning for secure decentralized collaboration in cancer image analysis. Med Image Anal 92, 103059. 10.1016/j.media.2023.103059

9. Fallbeispiel 3: Entwicklung einer Künstlichen Intelligenz (KI) zur Vorhersage der Mortalität bei Intensivpatienten

A. Sachverhalt und zu klärende Fragestellung

Im folgenden Fallbeispiel wird eine KI zur Vorhersage der Mortalität bei Intensivpatienten in Kooperation mit einem wirtschaftlichen Anbieter entwickelt. Hierzu werden Daten verschiedener Kliniken zusammengeführt und verarbeitet. Die Daten sollen so anonymisiert werden, dass für die datenverarbeitende Stelle, Kooperationspartner und Dritte kein Rückschluss auf Identitäten mehr möglich ist.

Vorgehensweise

Alle Identifikatoren werden aus dem Datenset entfernt und alle Quasi-Identifikatoren werden durch eine K-Anonymisierung mit Tupeln der Größe k = 5 anonymisiert. Medikamenten-Namen werden darüber hinaus pseudonymisiert und verhindern, dass die datenverarbeitende Stelle Rückschlüsse auf Therapien ziehen kann. Die Pseudonymisierung dieser Daten läuft in zwei Schritten ab:

- **Encrypt**: Verschlüsselung der Daten. Die Schlüssel werden bei diesem Vorgang erstellt und bleiben innerhalb des Zentrums.
- **Decrypt**: Nach Abschluss der Studie hat das Personal des Zentrums die Möglichkeit die Ergebnisse der Studie mittels der Schlüssel auf die Originalwerte der Medikamente umzuwandeln. Die Umwandlung ist für das Personal der datenverarbeitenden Stelle nicht möglich, da dieses nicht im Besitz der Schlüssel ist.

Folgende Identifikatoren befinden sich im Datenset und werden entfernt:

- Fallnummer / Patientennummer
- Freitextfelder (können direkte Identifikatoren enthalten)

Folgende Quasi-Identifikatoren befinden sich im Datenset und werden entfernt:

- Größe
- Gewicht

Folgende Quasi-Identifikatoren befinden sich im Datenset und werden durch Tupelbildung der Größe k = 5 anonymisiert. Bei der Tupelbildung werden alle Quasi-Identifikatoren berücksichtigt:

- Alter
- Geschlecht
- Body-Mass-Index (BMI)
- Medikamente
- Blutgruppe
- Rhesusfaktor
- Aufenthaltsdauer im Krankenhaus
- Behandlungsstatus (verstorben / entlassen)

Im Folgenden werden die einzelnen Elemente des Datensets gelistet und deren Verarbeitungsmethode erläutert:

Parameter	**Methode**
Name	Entfernt
Fallnummer und Patientennummer	Entfernt
Gewicht und Größe	Entfernt
Jegliche Freitextfelder	Entfernt
Quasi-Identifikatoren (Alter, Geschlecht, BMI, Medikamente, Blutgruppe, Rhesusfaktor, Aufenthaltsdauer, Behandlungsstatus)	Zu Tupeln der Größe k = 5 zusammengefasst (k-Anonymität)
Medikamente	Ersetzt durch Zufallszeichen (Pseudonymisierung) Zum Beispiel: Original: Paracetamol Nach Encrypt: JSKSLDLE
Zeitstempel in Messwerte	Ersetzt durch Zahlenfolge mit patientenbezogenem Startpunkt (Pseudonymisierung). Das heißt jeder Patient beginnt bei Zeitpunkt 0, ein Rückschluss auf Datumsangaben ist so ausgeschlossen. Eine Rückumwandlung in die originalen Werte ist nicht mehr möglich. Zum Beispiel: Original: 2020-04-08 04:42:00.000 Nach Encrypt: 12312412
Labor- / Messwerte	Labor- und Messwerte, z.B. Bilirubin, bleiben unverändert

Zufallszahlen und Schlüssel werden bei jedem Encrypt-Vorgang neu generiert.

Zusätzlich zu den Anonymisierungs- und Pseudonymisierungsmethoden, werden die Daten nur auf verschlüsselten Servern gehostet, auf welche nur autorisierte Personen Zugriff haben.

Erläuterung der Maßnahmen

Namen, Adressen, Freitextfelder

Da Namen, Adressen und Freitextfelder entfernt werden liegt hier kein Risiko vor.

Fallnummer

Die Fallnummern werden unumkehrbar entfernt. Die Verknüpfung der Daten wird durch eine zufällig generierte Zahl gewährleistet.

Quasi – Identifikatoren

Als Quasi-Identifikatoren wurden folgende Werte identifiziert:

- Gewicht
- Größe
- Alter
- Geschlecht
- Body-Mass-Index (BMI)
- Medikamente
- Blutgruppe
- Rhesusfaktor
- Aufenthaltsdauer im Krankenhaus
- Behandlungsstatus (verstorben / entlassen)

Gewicht und Größe werden aus dem Datensatz entfernt. Alle anderen Quasi-Indentifikatoren werden durch k-Anonymisierung zu Tupeln der Größe k=5 zusammengefasst. Dies bedeutet, dass in jeder Gruppe mit derselben Altersklasse, BMI-Klasse, Medikation, Blutgruppe, Rhesusfaktor, Aufenthaltsdauer im Krankenhaus und Behandlungsstatus sich immer mindestens 5 Leute befinden müssen. Lassen sich für manche Kombinationen keine Gruppen von mindestens 5 Personen finden, werden diese aus dem Datensatz entfernt. Die Tupelbildung wird mithilfe des Mondrian-Algorithums durchgeführt. Die Tupel werden mit Hilfe der folgenden Features gebildet:

Feature	**Repräsentation nach K-Anonymisierung**
Behandlungsstatus (gestorben/entlassen)	Bleibt erhalten - Gruppen von weniger als 5 Personen werden komplett aus dem Datensatz entfernt
Blutgruppe	Bleibt erhalten, Gruppen von weniger als 5 Personen werden komplett aus dem Datensatz entfernt
Rhesusfaktor	Bleibt erhalten, Gruppen von weniger als 5 Personen werden komplett aus dem Datensatz entfernt

Medikation	Medikamente, welche weniger als 5 Personen bekommen, werden aus dem Datenset entfernt, anschließend werden alle Medikamente als eigene Anonymisierungsfeatures behandelt. Können für bestimmte Medikamente keine homogenen Gruppen von 5 Personen mit derselben Medikation gebildet werden, werden diese Medikamente ebenfalls aus dem Datenset entfernt. S
Geschlecht	Bleibt erhalten, falls Gruppen von weniger als 5 Personen vorhanden sind (z.B. Geschlecht: Divers), werden diese komplett aus dem Datensatz entfernt
Alter	Bildung von Altersgruppen, z.B. 15 – 20 Jahre
BMI	Bildung von BMI – Gruppen: z.B. 15 – 20
Aufenthaltsdauer	Bildung von Gruppen, z.B. 15 - 20 Tage

Die Tupel werden über folgende Matrix gebildet, die aus mehreren Tabellen des originalen Datensatzes zusammengesetzt wird:

Behandlungs-status	**Blut-gruppe**	**Rhesus-faktor**	**Medik.-1**	**Medik.-2**	**…**	**Medik.-n**	**Ge-schlecht**	**Alter**	**BMI**	**Aufenthalts-dauer**
z.B. „entlassen“	z.B. „A“	z.B. „positiv“	z.B. „Ibuprofen“				z.B. „m“	z.B. 15-25"	z.B. 15-25	z.B. 15-25

Medikamente

Medikamenten-Namen können Aufschluss über die Therapie geben und sollen der datenverarbeitenden Stelle nicht bekannt werden. Deswegen werden diese durch zufällig generierte Zahlenfolgen ersetzt. Der Schlüssel für die Umwandlung verbleibt beim Zentrum, sodass die Ergebnisse für das Zentrum einsehbar sind.

Zeitstempel

Es soll nicht bekannt werden, zu welchen Zeiten sich Personen im Krankenhaus befanden, deswegen werden die Zeitstempel durch relative Zeitangaben auf folgende Art und Weise ersetzt:

- Die kleinste Datumsangabe für jeden einzelnen Patienten wird gefunden und als Zeitpunkt 0 gespeichert
- Weitere Messwerte werden als Distanz zum ersten Zeitpunkt abgespeichert
- Die originalen Zeitangaben werden gelöscht

Auf diese Art und Weise gehen die originalen Zeitangaben verloren und es ist pro Patienten nur noch der zeitliche Verlauf der Werte nachzuvollziehen.

Datenaustausch

Der Datenaustausch läuft in folgenden Schritten ab:

1. Anonymisierung der Daten auf einem Rechner im Zentrum ohne Anschluss ans Internet
2. Austausch der anonymisierten Daten über einen Austauschdatenträger auf einen PC mit Internet
3. Upload der Daten auf eine gesicherte Plattform
4. Löschen der Daten auf dem lokalen Computer und Formatieren des Austauschdatenträgers
5. Download der anonymisierten, ausgewerteten Daten aus der Plattform
6. Austausch der Daten über eine Austauschdatenträger auf einen PC im Zentrum ohne Anschluss ans Internet
7. Formatieren des Datenträgers

Datenverarbeitung und Entwicklungssystem

Über die Datenaustausch-Plattform werden die Daten in eine Entwicklungsplattform (hier: Amazon Web Services, AWS[28]) eingespielt. Die Entwicklung und Verarbeitung der Daten erfolgten lediglich auf Servern in Frankfurt, Deutschland. Das verwendete System erfüllt folgende Zertifizierungen und Standards:

- HIPAA
- DS-GVO
- ISO9001, ISO27001, ISO27017, ISO27018,
- SOC1, SOC2, SOC3, C5

Gesamtbewertung

Durch Tupelbildung über Quasi-Identifikatoren werden die Daten so anonymisiert, dass es für die datenverarbeitende Stelle und andere Dritte nicht möglich ist Aufschluss über die Einzelpersonen zu gewinnen. Durch Pseudonymisierung von wichtigen Parametern (Medikamente) wird darüber hinaus sichergestellt, dass die datenverarbeitende Stelle oder andere Dritte keine Erkenntnisse über diese gewinnen können. Die Anonymisierung wird innerhalb des Zentrums auf Computern ohne Internetzugriff durchgeführt und nur anonymisierte Daten verlassen das Zentrum. Die Schlüssel für die De-Pseudonymisierung der wichtigen Parameter werden nur im Zentrum auf Computern ohne Internetzugriff oder Datenträgern gespeichert.

Die anonymisierten Daten werden darüber hinaus von der datenverarbeitenden Stelle nur auf zertifizierten Systemen verarbeitet, wodurch ein unbefugter Zugriff auf diese Daten vermieden wird.

[28] Diese ist der de-facto Standard für die Verarbeitung vertraulicher Daten und der Entwicklung technischer Anwendungen im Medizinbereich. Weitere Informationen sind unter folgender Adresse einsehbar: https://aws.amazon.com/de/compliance/hipaa-compliance/?nc1=h_ls

Auf Basis der Ausführungen hat die datenverarbeitende Stelle das datenschutzrechtliche Risiko als gering eingestuft.

Risiko	Schwere	Eintritt	Maßnahmen
Die Schlüssel für die Entschlüsselung der wichtigen Parameter (Medikamente) werden entschlüsselt.	1	1	Es werden nur Parameter verschlüsselt, die für die Anonymisierung nicht relevant sind. Für jeden Parameter wird eine einzelne Verschlüsselung durchgeführt, wobei der originale Parameter durch eine zufällig erzeugte Zeichenfolge ersetzt wird. Eine direkte Umrechnung des verschlüsselten Parameters in den originalen Wert ist somit nicht möglich.
Der Datenträger mit den anonymisierten Daten wird nicht formatiert und kommt abhanden	2	2	Die Datenübertragung von einem Computer findet im selben Raum statt und die Formatierung muss von 2 Personen überprüft werden. Daten sind auf dem Datenträger schon anonymisiert.
Daten werden nicht vom PC mit Internetanschluss gelöscht	2	3	Daten sind bereits anonymisiert. Auf dem PC wird aktuelle Virensoftware installiert.
Fremde Personen verschaffen sich Zugriff auf die Plattform, in der die anonymisierten Daten gespeichert sind	2	1	Die Daten sind verschlüsselt und anonymisiert. Das Cloud-System garantiert höchste Sicherheit.
Eine unautorisierte Person erhält die Zugriffsschlüssel für die Daten in der Cloud	3	2	Die Daten sind bereits anonymisiert und können nur innerhalb des Cloud-Systems weiterverarbeitet werden (Virtual Private Cloud Netzwerk). Zugriff zu diesem Netzwerk nur über Zwei-Faktor-Authentifizierung möglich

B. Datenschutzrechtliche Diskussion des Fallbeispiels

Das vorliegende Fallbeispiel zeigt anschaulich, wie mit geringen Modifikationen an den Datensätzen der Personenbezug entfernt und Daten nutzbar gemacht werden können.

Auch in der Gesamtschau besteht ein zu vernachlässigendes Risiko einer Reidentifizierung, so dass auch größere Datensätze ausgewertet und genutzt werden können.

Besonderheiten durch die Beteiligung eines kommerziellen Anbieters an der Datenverarbeitung ergeben sich beim vorliegenden Fall aus datenschutzrechtlicher Sicht nicht.

Sofern die weitere Analyse ergibt, dass ein höherer Bedarf an spezifischen u. a. auch personenbeziehbaren Daten besteht, sind punktuelle Anpassungen von einzelnen Datensätzen möglich. Es sollte aber dann auch geprüft werden, ob dies noch in technisch-organisatorischer Hinsicht aufgefangen werden kann, oder über eine alternative Erhebungsmethode zur Anwendung kommen sollte, um das Forschungsziel nicht zu gefährden (Forschung mit Einwilligung, Datenerhebung auf gesetzlicher Grundlage o. ä.).

Auch im vorliegenden Fall sollte der KI-Einsatz und die Risiken für die Betroffenen über eine DSFA abgesichert werden.

10. Fallbeispiel 4: Abgrenzung von Qualitätssicherung und Forschung

Das folgende Gutachten gliedert sich in eine Darstellung des Sachverhalts bzw. der an den HDBI herangetragenen Fragestellung und eine materiell-rechtliche Bewertung.

A. Sachverhalt und zu klärende Fragestellung

Medizinisch- wissenschaftlicher Hintergrund / Notwendigkeit

Bei einer Analyse von medizinischen Daten der eigenen Einrichtung zum Zweck der Qualitätssicherung können allgemein relevante Erkenntnisse gewonnen werden, die auch eine Veröffentlichung, z.B. in wissenschaftlichen Journals, auch eine Weiterverfolgung der Fragestellung sinnvoll machen.

Eine Auswertung der Daten zu Forschungszwecken benötigt jedoch eine vorherige Genehmigung durch die zuständige Ethikkommission und muss dann auch streng retrospektiv angelegt sein, wenn keine Aufklärung der Patientinnen und Patienten vorliegt.

Nutzung von / Umgang mit medizinischen Daten

Weitergehende Sekundärnutzung von Behandlungsdaten über die interne Qualitätssicherung hinaus, wenn forschungsrelevante Fragestellungen aus dem Prozess der Qualitätssicherung resultieren.

Datenschutz-rechtliche Probleme / Unsicherheiten

Es besteht Unsicherheit, an welcher Stelle die Grenze zwischen Qualitätssicherung und Forschung verläuft und wie eine retrospektive Analyse zu definieren ist.

B. Datenschutzrechtliche Diskussion des Fallbeispiels

Zur Bewertung des Sachverhalts der DGIM sollen zunächst die grundlegenden Begriffe der Zweckbestimmungen „Qualitätssicherung“ und „Forschung“ erarbeitet und voneinander abgegrenzt werden. Sodann soll der Frage nachgegangen werden, inwieweit eine Verarbeitung von Gesundheitsdaten zum Zweck der Qualitätssicherung und zu Forschungszwecken datenschutzrechtlich zulässig ist.

Begriffsbestimmungen und Abgrenzungsfragen

Zunächst soll also erarbeitet werden, was unter Qualitätssicherung als Zweckbestimmung einer Datenverarbeitung i.S.d. Datenschutzrechts zu verstehen ist. Sodann wird der Forschungsbegriff der DS-GVO erarbeitet, bevor zur Abgrenzung knapp die Unterschiede herausgestellt werden.

Versuch einer einheitlichen Begriffsdefinition der Qualitätssicherung i.S.d. Datenschutzrechts

Das allgemeine Datenschutzrecht gibt keine einheitliche gesetzliche Definition dafür vor, was unter dem Zweck der Qualitätssicherung zu verstehen ist. Im Sozialrecht[29] ist hingegen sehr genau vorgesehen, wie qualitätssichernde Maßnahmen durchzuführen sind, wobei hier vor allem der Gemeinsame Bundesausschuss (G-BA) viele Prüfpflichten vorgibt.[30] Hier dienen demnach alle Datenverarbeitungsschritte der Qualitätssicherung, die zur Umsetzung der sozialrechtlichen Vorgaben[31] nötig sind.

[29] §§ 135a ff. SGB V
[30] §§ 136 ff. SGB V
[31] §§ 135a ff. SGB V

Nach der Gesetzesbegründung zum GDNG muss das Gesundheitssystem den Angehörigen der Heilberufe ermöglichen, „*ihr eigenes Handeln kritisch zu reflektieren, die Wirksamkeit und Effizienz ihrer Maßnahmen und der Strukturen, in denen sie wirken, zu analysieren und auf der Basis eigener Erfahrung forschend die Möglichkeiten der Therapie, Prävention, Diagnostik und Beurteilung zu erweitern*“. [32] Die Qualitätssicherung sei hiernach Teil eines lernendes Gesundheitssystems.

Eine klare Abgrenzung zwischen Qualitätssicherung und wissenschaftliche Forschung nimmt die Gesetzesbegründung zum GDNG nicht vor.

Das Bundesgesundheitsministerium fasst auf seiner Homepage treffend zusammen, dass unter Qualitätssicherung im Krankenhaus im Sinne dieser Regelungen des SGB V „die Abbildung, Sicherung und Verbesserung der Qualität insbesondere der ärztlichen und pflegerischen Tätigkeiten“ zu verstehen sei.[33] Daran lehnt sich auch das hessische Landeskrankenhausrecht an.[34] Im Sinne der Einheit der Rechtsordnung sollte auch der datenschutzrechtliche Begriff des Zwecks der Qualitätssicherung – auch im Bundesdatenschutzgesetz (BDSG), wo es keinen expliziten Verweis auf die Qualitätssicherung i.S.d. Sozialrechts gibt[35] – hiermit grundlegend korrelieren.

Hiernach handelt es sich also dann um den Zweck der Qualitätssicherung, wenn eine Gesundheitseinrichtung personenbezogene Daten bzw. Gesundheitsdaten verarbeitet, um die Versorgungsqualität abzubilden, zu sichern und/oder zu verbessern.

Definition von wissenschaftlichen Forschungszwecken i.S.d. Datenschutzrechts

Wann eine Datenverarbeitung dem Zweck der wissenschaftlichen Forschung[36] dient bzw. was genau unter einem wissenschaftlichen Forschungszweck i.S.d. Datenschutzrechts zu verstehen ist, wird ebenfalls nicht ausdrücklich vom Normtext der DS-GVO vorgegeben. Die Erwägungsgründe der DS-GVO geben jedoch Hinweise hierzu, woran auch der Europäische Datenschutzausschuss (EDSA) und die Konferenz der unabhängigen Datenschutzbehörden des Bundes und der Länder (DSK) anknüpfen.

a. Erwägungsgrund 159 der DS-GVO und EDSA

Wie der EDSA in einer Stellungnahme betont,[37] gibt Erwägungsgrund 159 der DS-GVO eine Auslegungshilfe zur Definition des datenschutzrechtlichen Forschungsbegriffs. Hiernach sollte die „Verarbeitung personenbezogener Daten zu wissenschaftlichen Forschungszwecken [...] weit ausgelegt werden und die Verarbeitung für beispielsweise die technologische Entwicklung und die Demonstration, die Grundlagenforschung, die angewandte Forschung und die privat finanzierte Forschung einschließen. Darüber hinaus sollte sie dem festgeschriebenen Ziel, einen europäischen Raum der Forschung zu schaffen, Rechnung tragen.[38] Die wissenschaftlichen Forschungszwecke sollten auch

[32] Kabinettvorlage „Gesetz zur verbesserten Nutzung von Gesundheitsdaten“, S. 61, https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/3_Downloads/Gesetze_und_Verordnungen/GuV/G/GDNG_Kabinett.pdf
[33] Vgl. BMG, Qualitätssicherung im Krankenhausbereich, https://www.bundesgesundheitsministerium.de/qualitaet-krankenhausversorgung.html (Abruf am 27.9.2023).
[34] § 8 Abs. 1 S. 2 HKHG
[35] § 22 BDSG
[36] Zum Begriff "wissenschaftliche Forschung " siehe auch 10.B.a. (Erwägungsgrund 159 der DS-GVO und EDSA)
[37] *EDSA*, Leitlinien 3/2020 für die Verarbeitung von Gesundheitsdaten für wissenschaftliche Forschungszwecke im Zusammenhang mit dem COVID-19-Ausbruch, Angenommen am 21. April 2020, Rn. 9.
[38] Artikel 179 Absatz 1 AEUV

Studien umfassen, die im öffentlichen Interesse im Bereich der öffentlichen Gesundheit durchgeführt werden.

Der EDSA zitierte zudem seine Vorgängerinstitution, die frühere Artikel-29-Datenschutzgruppe, die bereits darauf hinwies, „dass der Begriff ‚wissenschaftliche Forschung' nicht über seine allgemeine Bedeutung hinaus ausgeweitet werden sollte und in diesem Kontext als ein Forschungsprojekt verstanden wird, das in Übereinstimmung mit den maßgeblichen, für den Sektor relevanten methodischen und ethischen Standards und in Übereinstimmung mit bewährten Verfahren entwickelt wird".[39] Diese Aussagen bekräftigte der EDSA sodann in seiner Neuauflage der Leitlinie zu Einwilligungen.[40]

b. DSK

Die Konferenz der unabhängigen Datenschutzbehörden des Bundes und der Länder (DSK) stellt in ihrem Positionspapier zum Begriff „wissenschaftliche Forschungszwecke" eine Definition bereit. Danach müssen die folgenden fünf Kriterien erfüllt sein, damit es sich um wissenschaftliche Forschungszwecke handelt:

I. Methodisches und systematische Vorgehen
II. Erkenntnisgewinn
III. Nachprüfbarkeit
IV. Unabhängigkeit und Selbstständigkeit
V. Gemeinwohlinteresse

Eine Veröffentlichung (als Publikationen, Vorträge, etc.) der Forschungsergebnisse ist zwar keine zwingende Voraussetzung wissenschaftlicher Forschung. Eine Absicht oder zumindest grundsätzliche Bereitschaft zur Veröffentlichung der Forschungsergebnisse ist aber im Rahmen des Kriteriums der Nachprüfbarkeit erforderlich.

Den Ansatz des Gemeinwohlinteresses findet man auch nicht selten in der Literatur, wo eine „Wissensgenerierung für die Allgemeinheit" oder jedenfalls die geplante Erzielung eines gesellschaftlichen Nutzens zur Bedingung für das Vorliegen von Forschungszwecken gemacht wird.[41] Dies stützt etwa Erwägungsgrund 113 der DS-GVO, der bei Forschungszwecken von „legitimen gesellschaftlichen Erwartungen in Bezug auf einen Wissenszuwachs" spricht.[42]

Bei der medizinischen Forschung ist grundsätzlich von einem Gemeinwohlinteresse auszugehen. Das Kriterium V (Gemeinwohlinteresse) schließt nicht aus, dass auch Vorhaben mit wirtschaftlicher Motivation und finanziellen Interessen wissenschaftliche Forschungszwecke im Sinne der DS-GVO verfolgen können, wenn die Ergebnisse der Allgemeinheit zugutekommen (sollen).

Interessant kann auch der vergleichende Blick auf die Rechtsprechung des Bundesverfassungsgerichts sein, die Forschung i.S.d. deutschen Grundrechts der Forschungsfrei-

[39] *EDSA*, Leitlinien 3/2020 für die Verarbeitung von Gesundheitsdaten für wissenschaftliche Forschungszwecke im Zusammenhang mit dem COVID-19-Ausbruch, Angenommen am 21. April 2020, Rn. 10 mit Verweis auf die Leitlinien der früheren Artikel-29-Datenschutzgruppe in Bezug auf die Einwilligung gemäß der Verordnung 2016/679 vom 10.4.2018, WP259 rev.01, 17DE, S. 33 (vom EDSA gebilligt).
[40] *EDSA*, Leitlinien 05/2020 zur Einwilligung gemäß Verordnung 2016/679, Version 1.1, angenommen am 4. Mai 2020, Rn. 159.
[41] Siehe etwa *Buchner/Schnebbe* in Sturma/Lanzerath (Hrsg.), Big Data in der Medizin, 2020, S. 49, 71 f.; *Buchner/Tinnefeld* in Kühling/Buchner (Hrsg.), DS-GVO/BDSG, 3. Aufl. 2020, Art. 89 Rn. 13; ähnlich auch *Roßnagel*, ZD 2019, 157, 158, wonach das Erzielen und Publizieren von Erkenntnisse Wesensmerkmal der Forschung sei.
[42] Vgl. *Schildbach* Big-Data-Anwendungen als Herausforderung für das Gesundheitsdatenschutzrecht, 2023, S. 226.

heit als jede geistige Tätigkeit mit dem Ziel, in methodischer, systematischer sowie nachprüfbarer Art und Weise neue Erkenntnisse zu gewinnen sieht.[43] Dies ist jedoch für den in letzter Instanz zur Auslegung des Forschungsbegriffs der DS-GVO berufenen EuGH nicht verbindlich.

Abgrenzung zum Zweck der Qualitätssicherung

In Abgrenzung zur Qualitätssicherung als Zwecksetzung einer Datenverarbeitung fällt bei dem Begriff der wissenschaftlichen Forschungszwecke auf, dass zunächst beide Datenverarbeitungszwecke der Wissensgenerierung dienen.

Bei der Qualitätssicherung zielt dieser Erkenntnisgewinn jedoch darauf ab herauszufinden, ob intern in einer Gesundheitseinrichtung oder in einem Bereich des Gesundheitswesens die Qualität der Versorgung bestimmten Vorgaben entspricht.

Bei der Forschung sollen für die Allgemeinheit nützliche Erkenntnisse gewonnen und dieser jedenfalls in irgendeiner Weise zur Verfügung gestellt werden.

Solange also nur zur „internen Optimierung“ Daten verarbeitet und Erkenntnisse generiert werden sollen, liegt noch kein Forschungszweck i.S.d. Datenschutzrechts vor. Sollen jedoch gerade für die (Fach-)Öffentlichkeit Erkenntnisse generiert werden, sind also insbesondere Publikationen oder ein Wissensaustausch im Rahmen von Fachkongressen o.ä. über die Ergebnisse der Datenverarbeitung geplant, so handelt es sich um Forschungszwecke.

Grenzbereich Qualitätssicherung-Forschung: Beispiele aus der klinischen Praxis

Die folgenden realen Beispiele zeigen, dass es in der Praxis oftmals zu Überschneidungen und Grenzfällen zwischen Qualitätssicherung und Forschung kommt.

Beispiel 1: Testung auf Mikrosatelliteninstabilität (MSI) beim kolorektalen Karzinom

Es wurde überprüft, ob eine Testung auf MSI[44] bei kolorektalem Karzinom im Patientengut einer großen universitär-medizinischen Fachabteilung gemäß den Empfehlungen der Deutschen Krankenhausgesellschaft (DKG) durchgeführt wurde. Aufgrund einer seit einigen Jahren bestehenden neuen therapeutischen Option für Betroffene, sollte bei allen Patienten im fortgeschrittenen Tumorstadium IV überprüft werden, ob eine Testung auf MSI erfolgt war. Die Ergebnisse der Überprüfung führten dazu, interne Prozesse umzustellen und sollten publiziert werden, da anzunehmen war, dass möglicherweise auch in Fachabteilungen und Instituten anderer Kliniken mit Blick auf eine MSI-Testung beim kolorektalen Karzinom Versorgungsdefizite vorliegen.

Beispiel 2: Standardisierte Nachsorge beim Ösophaguskarzinom

Eine regelhafte klinische Nachsorge beim Ösophaguskarzinom ist nicht evidenzbasiert, allerdings in vielen Kliniken gängige Praxis. Um zu evaluieren, ob eine standardisierte Nachsorgesprechstunde auch wirklich alle betroffenen Patienten einer großen universitär-medizinischen Fachabteilung erreichte, wurde überprüft, ob und wenn ja wie die Nachsorge fallbezogen durchgeführt wurde und ob dies einen Einfluss auf das Überle-

[43] BVerfG, Urteil vom 29. 5. 1973 - 1 BvR 424/71 u. 325/72 = NJW 1973.

[44] Die sogenannte Mikrosatelliteninstabilität (MSI) ist ein charakteristisches Merkmal zur Unterscheidung von verschiedenen Krebsarten des Magen-Darm-Trakts und bestimmt, ob Patienten mit diesen Erkrankungen besonders gut auf eine Immuntherapie mit Checkpoint-Inhibitoren ansprechen. Siehe: https://dktk.dkfz.de/ueber-uns/news/magen-und-darmkrebs-geeignete-patienten-fuer-eine-immuntherapie-mit-kuenstlicher-intelligenz-fruehzeitig-identifizieren

ben der Patienten hatte. Die Ergebnisse sollten dann publiziert werden, da den Ergebnissen der internen Untersuchung wichtige Erkenntnisse mit Relevanz für das praktische Vorgehen anderer Fachkliniken zu entnehmen waren.

Beispiel 3: Adjuvante Therapie beim Ösophaguskarzinom

Nach Einführung einer adjuvanten Therapiestrategie beim operierten Ösophaguskarzinom hielt diese Empfehlung für betroffene Patienten Einzug in das interdisziplinäre Tumorboard einer Universitätsklinik. Es sollte konsekutiv überprüft werden, ob die Empfehlung zur adjuvanten Therapie in der Praxis in dem weiten Einzugsgebiet der betreffenden Universitätsklinik auch umgesetzt wird. Die Ergebnisse der Untersuchung führten zu einer Umstellung interner Prozesse, da die Empfehlung nicht in einem ausreichenden Prozentsatz durchgeführt wurde. Diese Ergebnisse sollen publiziert werden, da die Aufarbeitung der möglichen Ursachen voraussichtlich auch für andere versorgende Einrichtungen von Relevanz ist.

Wenn schwerpunktmäßig die internen Abläufe anhand bestehender (Qualitäts-)Kriterien überprüft und verbessert werden sollen, spricht dies für den Zweck der Qualitätssicherung.

Demgegenüber handelt es sich dann eher um Forschung, wenn neue Erkenntnisse über den medizinischen Nutzen von Vorgaben bzw. Leitlinien erzielt werden sollen (z.B. Einfluss von Nachsorgemaßnahmen aufs Überleben).

Die Nachprüfbarkeit ist zwar ein Kriterium für wissenschaftliche Forschungszwecke. Die Absicht zur Veröffentlichung anonymisierter Ergebnisse aus der Qualitätssicherung schließt aber dann den Zweck der Qualitätssicherung nicht per se aus, wenn die Daten zunächst schwerpunktmäßig zur Qualitätssicherung verarbeitet wurden.

Zulässigkeit der Datenverarbeitung

Die DGIM schilderte zudem, dass bei der Qualitätssicherung immer wieder Erkenntnisse erzielt werden, die auch eine wissenschaftliche Verwertung der Daten interessant macht. Hier stellt sich die Frage, ob dies datenschutzrechtlich zulässig ist.

Die von der DGIM aufgeworfene Problematik stellt sich jedoch schon nicht, soweit die zu Qualitäts- und/oder Forschungszwecken zu verarbeitenden Daten nicht in den Anwendungsbereich des Datenschutzrechts fallen. Ist der Anwendungsbereich hingegen eröffnet, so gibt es verschiedene Zulässigkeitsregelungen, die eine Datenverarbeitung zum Zwecke der Qualitätssicherung erlauben. Sodann stellt sich die Frage, ob diese Daten auch zu Forschungszwecken verarbeitet werden dürfen.

Anwendbarkeit des Datenschutzrechts

Anonyme Daten fallen nicht in den Anwendungsbereich der DS-GVO. Soweit möglich sollte die Qualitätssicherung mit anonymen Daten durchgeführt werden. Eine Weiternutzung anonymer Daten zu Forschungszwecken kann ohne weitere datenschutzrechtliche Anforderungen erfolgen. Insbesondere soweit nur aggregierte Daten aus der Qualitätssicherung für Forschungszwecke benötigt werden, scheint dies denkbar. Gleichsam wird in vielen Fällen jedoch ein Agieren außerhalb des Anwendungsbereichs der DS-GVO nicht möglich sein oder jedenfalls mit Unsicherheiten behaftet sein, so dass in diesen Fällen anschließend die datenschutzrechtliche Zulässigkeit der Datenverarbeitung zu prüfen sein wird.

a. *Personenbezogene oder anonyme Daten*[45]

Die (sachliche) Anwendbarkeit des Datenschutzrechts richtet sich nach Art. 2 DS-GVO. Entscheidend dürfte dabei in den allermeisten Fällen sein, ob Daten *personenbezogen* sind[46], so dass das Datenschutzrecht Anwendung findet. Sind Daten hingegen *anonym*, können sie ohne weitere datenschutzrechtliche Vorgaben verarbeitet werden.[47]

Eine vorherige Anonymisierung ist jedoch eine Verarbeitung und bedarf daher einer datenschutzrechtlichen Rechtsgrundlage.[48] In der Regel lässt sich bei der Anonymisierung zu Forschungszwecken aber eine Rechtsgrundlage finden.[49]

Personenbezogene Daten sind „alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person […] beziehen".[50] Wichtigstes Merkmal ist dabei die Frage der Identifizierbarkeit. Denn wenn einzelne Personen oder Patienten nicht mehr aus den Daten zu identifizieren sind, so kann mit den Daten – jedenfalls aus datenschutzrechtlicher Sicht – nach Belieben verfahren werden.

Die Frage der Identifizierbarkeit ist ebenfalls gesetzlich definiert: „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".[51] Der nicht unmittelbar geltende, aber für die Auslegung des Begriffs der Identifizierbarkeit bedeutsame Erwägungsgrund 26 der DS-GVO gibt weiterführende Hinweise, indem dieser zusammengefasst die Wahrscheinlichkeit einer Identifizierung nach allgemeinem Ermessen zum Kriterium hierfür erklärt.

Damit kommt es letztlich auf eine Risiko- bzw. Wahrscheinlichkeitsanalyse der Identifizierung natürlicher Personen an.[52] Ohne gesetzliche Regelung lässt sich dies nur anhand einer Risikobetrachtung im Einzelfall beantworten.[53] Besteht bei einer konkret geplanten Datenverarbeitung kein oder nur ein vernachlässigbar geringes Risiko der Identifizierung von Personen, so ist das Datenschutzrecht nicht anwendbar. Dass auch Szenarien der Datenverarbeitung existieren müssen, in denen die Identifizierungswahrscheinlichkeit zwar nicht „gleich null" ist, aber dennoch wegen der vernachlässigbar geringen Wahrscheinlichkeit der Identifizierung anonyme Daten vorliegen, hat auch der EuGH anerkannt, als er in seiner Rechtsprechung wörtlich von Risiken spricht, die „de facto vernachlässigbar" erscheinen.[54]

---

[45] s. dazu auch die Ausführungen unter **Error! Reference source not found.** und 6.B, S.6 ff.
[46] i.S.v. Art. 2 Abs. 1 i.V.m. Art. 4 Nr. 1 DS-GVO
[47] Dies ergibt der Umkehrschluss aus Art. 2 Abs. 1 DS-GVO insbesondere unter Berücksichtigung des Erwägungsgrundes 26 der DS-GVO.
[48] Art. 4 Nr. 2 DS-GVO
[49] In Betracht kommen Art. 5 Abs. 1 lit. b i.V.m. Art. 6 Abs. 4 DS-GVO und Art. 6 Abs. 1 lit. c, e, Art. 9 Abs. 2 lit. j, Art. 89 DS-GVO i.V.m. nationalen Gesetzen (z.B. § 6 Abs. 3 S. 3 GDNG, § 27 BDSG oder § 24 HDSIG).
[50] Art. 4 Nr. 1 DS-GVO
[51] Art. 4 Nr. 1 DS-GVO
[52] Siehe hierzu etwa *Klar/Kühling* in Kühling/Buchner (Hrsg.), DS-GVO/BDSG, 3. Aufl. 2020, Art. 4 Nr. 1 Rn. 22; *Schildbach* Big-Data-Anwendungen als Herausforderung für das Gesundheitsdatenschutzrecht, 2023, S. 151 f. mwN.
[53] Die DSK hat mit Entschließung vom 23.11.2023 (S. 3) eine solche gesetzliche Regelung für besondere Forschungsgegenstände (wie z.B. radiologische Bilddaten) gefordert; https://www.datenschutzkonferenz-online.de/media/en/2023-11-23_DSK-Entschliessung_DS.pdf.
[54] EuGH, Urt. v. 19.10.2016, Rs. C-582/14 (Breyer), ECLI:EU:C:2016:779, Rn. 46.

b. Beispiel aggregierte Daten

Für die vorliegende Fragestellung kann dies etwa Bedeutung haben, sofern für die Forschung gar kein Zugriff auf personenbezogene Daten (mehr) nötig ist. Das könnte etwa der Fall sein, wenn die Weiterverarbeitung von *aggregierten* Daten aus dem Kontext der Qualitätssicherung genügt – also keine auf einzelne Personen bezogenen Daten mehr vorhanden sind, sondern nur „die Ergebnisse". Solche aggregierten Daten sind meist anonym,[55] sofern nicht ausnahmsweise doch Rückschlüsse auf Einzelpersonen möglich sind.[56]

Doch in aller Regel sind aggregierte Daten nicht personenbezogen und können deshalb auch zu Forschungszwecken verwendet werden, ohne dass das – sodann nicht anwendbare – Datenschutzrecht dem entgegensteht.

Ebenso dürfen aus der Qualitätssicherung gewonnene Daten nach einer ausreichenden Anonymisierung mittels Aggregation veröffentlicht werden, wenn eine Rechtsgrundlage für die Anonymisierung vorliegt.[57]

c. Verbleibende Restrisiken in anderen Konstellationen

In vielen anderen Fällen wird eine zuverlässige Anonymisierung hingegen schwerer umsetzbar sein. Insbesondere, soweit Forschende mit auf einzelne Personen bezogenen Daten arbeiten wollen, sind die Hürden zu einer erfolgreichen Anonymisierung höher und es besteht – je nach Fallkonstellation – oft noch Unsicherheit, ob eine Identifizierung ausgeschlossen ist.[58]

Für die externe Qualitätssicherung und die Forschung durch Dritte sind Ansätze denkbar, die auf eine Art von Datentreuhänder oder ähnliches zurückgreifen, so dass die nutzenden Dritten selbst keinen Zugriff auf personenbezogene Daten haben.[59] Auch ähnliche Ansätze, bei denen nur eine (berechtigte) Stelle tatsächlich Zugriff auf den Datensatz hat und andere (forschenden) Stellen lediglich ein Analyseskript übergeben und sodann die (aggregierte) Ergebnisdaten erhalten ist denkbar, damit Forschende nicht mit personenbezogenen Daten arbeiten müssen.

Gelingt eine Anonymisierung nicht oder bestehen jedenfalls zu große Unsicherheiten, so kommt es auf die im Folgenden beleuchteten Zulässigkeitsregelungen des Datenschutzrechts an.

Zulässigkeit der Datenverarbeitung zum Zwecke der Qualitätssicherung

Die Verarbeitung von personenbezogenen Daten bzw. Gesundheitsdaten zum Zwecke der Qualitätssicherung ist in vielen Fällen erlaubt. Das EU-Recht erlaubt mitgliedstaatliche Zulässigkeitstatbestände zu Qualitätssicherungszwecken. In Deutschland finden sich daran anknüpfenden verschiedene Regelungen, die Datenverarbeitungen zu Qualitätssicherungszwecken erlauben.

a. Art. 9 Abs. 2 DS-GVO als Öffnungsklausel

Die auf der Ebene der EU angesiedelte DS-GVO als wichtigste datenschutzrechtliche Regelung steht einer Verarbeitung von Gesundheitsdaten zum Zwecke der Qualitätssicherung offen gegenüber. Unter zwei Regelungen in Art. 9 Abs. 2 DS-GVO kann der

[55] Vgl. anstatt vieler *Klar/Kühling* in Kühling/Buchner (Hrsg.), DS-GVO/BDSG, 3. Aufl. 2020, Art. 4 Nr. 1 Rn. 15.
[56] Vgl. zu Re-Identifikationsszenarien *Kühling/Schildbach* NZS 2020, 43 f. mwN.
[57] insb. § 6 Abs. 3 S. 3 GDNG oder § 22 Abs. 1 Nr. 1 lit. c BDSG
[58] Weiterführend etwa *Gierschmann* ZD 2021, 482, 483 ff.
[59] Vgl. zum Anonymisierungspotenzial von Datentreuhändern *Buchner/Haber* et al., DuD 2021, 806, 810; *Kühling*, DuD 2021, 783, 784; *Selzer/Timm*, DuD 2021, 816, 816.

Verarbeitungszweck „Qualitätssicherung“ gefasst werden, nämlich die Buchstaben h und j.

Hier ist unter anderem die Datenverarbeitung zum Zweck der „Verwaltung von Systemen und Diensten im Gesundheits- oder Sozialbereich“ erfasst, wozu oft auch die Qualitätssicherung in diesem Bereich gezählt wird.[60, 61] Auch wird sodann die „Gewährleistung hoher Qualitäts- und Sicherheitsstandards bei der Gesundheitsversorgung“ im Bereich der öffentlichen Gesundheit – also, sofern staatliche Stellen zur Qualitätssicherung tätig werden, adressiert.[62, 63] Beide Normen erlauben es EU-Mitgliedstaaten hierzu Regelungen zu treffen.

b. § 299 SGB V als bereichsspezifischer Zulässigkeitstatbestand

Von dieser Möglichkeit haben deutsche Gesetzgeber Gebrauch gemacht. Für das Gesundheitswesen existiert eine Regel für die Datenverarbeitung zur Qualitätssicherung.[64] Soweit diese Regelung anwendbar ist – also nur im Bereich der gesetzlichen Krankenversicherung – erlaubt sie den an der vertragsärztlichen Versorgung teilnehmenden Ärzten, den zugelassenen Krankenhäuser und weiteren Leistungserbringern (Absatz 1) sowie den Kassen (Absatz 1a), Sozialdaten zum Zweck der Qualitätssicherung[65] zu verarbeiten. Einer Einwilligung der Versicherten bedarf es hierfür somit gerade nicht.[66] Dabei regeln die genannten Normen kleinteilig, wie mit den Daten zu verfahren ist.

c. § 6 GDNG

Seit März 2024 gilt das neue Gesundheitsdatennutzungsgesetz (GDNG). Nun dürfen Gesundheitseinrichtungen bei ihnen im Rahmen der Versorgung erhobene Daten zu Zwecken der Qualitätssicherung weiterverarbeiten, soweit dies für diesen Zweck erforderlich ist.[67] Die weiteren Anforderungen des § 6 GDNG sind dabei zu beachten (insb. Pseudonymisierung, frühestmögliche Anonymisierung, Rechte und Rollenkonzept, Protokollierung, öffentliche Informationen etc.).

§ 6 Abs. 1 Nr. 1 GDNG erlaubt als bundeseinheitliche Regelung eine umfangreiche Verarbeitung von Behandlungsdaten zu Zwecken der internen Qualitätssicherung.

Da die Regelung sehr neu ist, hat sich noch keine Auslegungspraxis etabliert. Die Verantwortlichen sollten daher die weitere Diskussion zum GDNG im Blick behalten.

d. Weitere Tatbestände des Landeskrankenhausrechts

Die Zulässigkeit der Datenverarbeitung zur Qualitätssicherung im Krankenhaus sieht zudem – für Hessen – das Hessische Krankenhausgesetz (HKHG) vor.[68] Dies gilt im Anwendungsbereich des Gesetzes, sofern die Datenverarbeitung erforderlich ist zur Qualitätssicherung in der stationären Versorgung, der Empfänger eine Ärztin oder ein Arzt oder eine ärztlich geleitete Stelle ist, der genannte Zweck nicht mit anonymisierten oder pseudonymisierten Daten erreicht werden kann und nicht überwiegende schutzwürdige

[60] Art. 9 Abs. 2 lit. h DS-GVO
[61] Siehe etwa *Weichert* in Kühling/Buchner (Hrsg.), DS-GVO/BDSG, 3. Aufl. 2020, Art. 9 Rn. 107.
[62] Art. 9 Abs. 2 lit. i DS-GVO
[63] Vgl. *Weichert* in Kühling/Buchner (Hrsg.), DS-GVO/BDSG, 3. Aufl. 2020, Art. 9 Rn. 116.
[64] etwa in § 299 SGB V
[65] §§ 135a ff. SGB V
[66] Vgl. etwa *Michels* in Becker/Kingreen (Hrsg.), SGB V, 8. Aufl. 2022, § 299 Rn. 1.
[67] § 6 Abs. 1 Nr. 1 GDNG
[68] § 12 Abs. 2 Nr. 7 HKHG

Interessen der Betroffenen entgegenstehen. Für die Frage, was unter Qualitätssicherung im Sinne des hessischen Krankenhausrechts zu verstehen ist, verweist das HKHG auf die Regelungen des SGB V.[69]

e. § 22 Abs. 1 Nr. 1 lit. c BDSG als allgemeiner Zulässigkeitstatbestand

Schließlich steht auch die Regelung des BDSG[70] einer Datenverarbeitung im Krankenhaus zu Qualitätssicherungszwecken offen gegenüber. Die allgemeine Regel gilt, soweit die oben genannten speziellen Regeln[71] ausnahmsweise nicht einschlägig sein sollten. Denn das BDSG[72] spiegelt im Wesentlichen die Regelungen der DS-GVO[73], die für Qualitätssicherung als Verarbeitungszweck offen sind.

Das BDSG[74] greift dabei die DS-GVO[75] und dort den Begriff der „Verwaltung von Systemen und Diensten im Gesundheits- und Sozialbereich" auf. Hierunter können Datenverarbeitungsvorgänge zum Zwecke der Qualitätssicherung nicht zuletzt durch private Akteure zählen.[76] § 22 Abs. 1 Nr. 1 lit. c BDSG ist hingegen die passende Regelung zu Art. 9 Abs. 2 lit. i DS-GVO und erlaubt als Auffangregelung Datenverarbeitung „zur Gewährleistung hoher Qualitäts- und Sicherheitsstandards bei der Gesundheitsversorgung".[77]

Zulässigkeit der Datenverarbeitung zu Forschungszwecken

Für die Verarbeitung von personenbezogenen Daten zu Forschungszwecken bestehen eine Vielzahl von möglichen Zulässigkeitstatbeständen. Sofern personenbezogene Daten, die eigentlich zum Zwecke der Qualitätssicherung verarbeitet werden, nun zu Forschungszwecken verarbeitet werden sollen, müssen unter dem Stichwort der Sekundärnutzung zudem die Anforderungen an eine legale Zweckänderung vorliegen.

a. Mögliche Zulässigkeitstatbestände für Forschende

(1) Einwilligung

Zunächst können Forschende bei der Verarbeitung von Gesundheitsdaten grundsätzlich immer auf eine Einwilligung zurückgreifen.[78]

(2) Art. 9 Abs. 2 lit. j DS-GVO als Öffnungsklausel

Daneben sieht die DS-GVO jedoch auch vor, dass EU-Mitgliedstaaten Zulässigkeitstatbestände für Datenverarbeitungen zu „wissenschaftlichen […] Forschungszwecken" erlassen können, wovon der Gesetzgeber in Deutschland umfangreich Gebrauch gemacht hat.[79]

(3) Bereichsspezifische Forschungsklauseln

Zunächst findet sich eine Vielzahl an speziellen bzw. bereichsspezifischen Forschungsklauseln im deutschen Recht. So erlaubt etwa das Sozialrecht, sofern es anwendbar ist,

---

[69] in § 8 Abs. 1 S. 2
[70] § 22 Abs. 1
[71] § 299 SGB V, § 12 Abs. 2 Nr. 7 HKHG
[72] § 22 Abs. 1 Nr. 1 lit. b und c
[73] Art. 9 Abs. 2 lit. h und i
[74] § 22 Abs. 1 Nr. 1 lit. b
[75] Art. 9 Abs. 2 lit. h
[76] Vgl. etwa *Rose* in Taeger/Gabel (Hrsg.), DS-GVO/BDSG/TTDSG, 4. Aufl. 2022, BDSG § 22 Rn. 26.
[77] Vgl. etwa *Rose* in Taeger/Gabel (Hrsg.), DS-GVO/BDSG/TTDSG, 4. Aufl. 2022, BDSG § 22 Rn. 33.
[78] nach den Art. 9 Abs. 2 lit. a, Art. 7, Art. 4 Nr. 11 DS-GVO
[79] Vgl. Roßnagel, Datenschutz in der Forschung, ZD 2019, 157.

unter den jeweiligen Voraussetzungen die Datenverarbeitung zum Zwecke der Forschung.[80]

Auch das Landeskrankenhausrecht sieht in manchen Bundesländern eine Forschungsklausel vor.[81] In Hessen existiert keine solche Regelung im Krankenhausgesetz, sondern ein Verweis auf die allgemeine datenschutzrechtliche Forschungsklausel. Weitere bereichsspezifische Forschungsklauseln sind im AMG und im MPG vorgesehen.[82]

(4) § 6 GDNG

Nach dem GDNG dürfen Gesundheitseinrichtungen bei ihnen bei ihnen im Rahmen der Versorgung erhobene Daten auch zu medizinischen, rehabilitativen und pflegerischen Forschungszwecken weiterverarbeiten, soweit dies hierfür erforderlich ist.[83] Die weiteren Anforderungen des § 6 GDNG sind dabei zu beachten (insb. Pseudonymisierung, frühestmögliche Anonymisierung, Rechte und Rollenkonzept, Protokollierung, öffentliche Informationen etc.).

In § 6 GDNG ist weder eine Einwilligung noch eine Widerspruchsrecht der betroffenen Personen vorgesehen. Auch eine Interessenabwägung im Einzelfall ist nicht vorgesehen. Die Gesundheitseinrichtungen sind aber dazu verpflichtet, die Öffentlichkeit über laufende Forschungsvorhaben zu informieren[84], Forschungsvorhaben in einem Register zu registrieren und Forschungsergebnisse zu veröffentlichen.[85]

Eine Weitergabe der Daten an Dritte ist untersagt[86]. Unter bestimmten Voraussetzungen ist allerdings die gemeinsame Nutzung der Daten durch öffentlich geförderte Zusammenschlüsse von datenverarbeitenden Gesundheitseinrichtungen (insb. Verbundforschungsvorhaben und Forschungspraxen-Netzwerke) zulässig.[87]

(5) Allgemeine datenschutzrechtliche Forschungsklauseln

Sofern keine Spezialnorm einschlägig ist, erlaubt auch das allgemeine deutsche Datenschutzrecht in vielen Fällen eine Datenverarbeitung zu Forschungszwecken. § 27 BDSG, soweit dieser einschlägig ist, erlaubt die Datenverarbeitung zu Forschungszwecken, wenn die Verarbeitung hierfür erforderlich ist und die Interessen des Verantwortlichen an der Verarbeitung die Interessen der betroffenen Person an einem Ausschluss der Verarbeitung erheblich überwiegen. Ist auf einen Verantwortlichen hingegen das Landesdatenschutzrecht anwendbar, so sieht für Hessen etwa § 24 HDSIG eine ähnliche Regelung vor, die allerdings kein „erhebliches“ Überwiegen der Forschungsinteressen fordert, sondern lediglich ein „einfaches“ Überwiegen.

---

[80] § 75 SGB X, § 67c Abs. 5 SGB X, §§ 287 und 287a SGB V, §§ 303a ff. SGB V sowie § 363 SGB V

[81] Siehe eine Liste aller forschungsrelevanten Landesnormen der Krankenhausgesetze bei *Dierks/Kircher et al.*, Lösungsvorschläge für ein neues Gesundheitsforschungsdatenschutzrecht in Bund und Ländern, 15.9.2019, S. 38, abrufbar unter: https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/5_Publikationen/Gesundheit/Berichte/RECHTSGUTACHTEN_Gesundheitsforschungsdatenschutzrecht_BMG.pdf.

[82] Siehe eine Auflistung weiterer forschungsrelevanter Spezialgesetze bei *Dierks/Kircher et al.*, Lösungsvorschläge für ein neues Gesundheitsforschungsdatenschutzrecht in Bund und Ländern, 15.9.2019, S. 69 ff.

[83] § 6 Abs. 1 Nr. 2 GDNG

[84] § 6 Abs. 4 S. 2 GDNG

[85] § 8 S. 3 GDNG

[86] § 6 Abs. 3 S. 1 GDNG

[87] § 6 Abs. 3 S. 4

Auf der Basis von § 27 BDSG und § 24 HDSIG kann grundsätzlich auch eine Übermittlung von Daten zu Forschungszwecken an Dritte zulässig sein, wobei die damit verbundenen Erhöhung der Risiken für die betroffenen Personen im Rahmen der Interessenabwägung zu berücksichtigen ist.

Ergebnis

Soweit nicht mit zuverlässig anonymisierten Daten gearbeitet werden kann, bedarf es eines Zulässigkeitstatbestandes für die Verarbeitung von Gesundheitsdaten. Sowohl für die Qualitätssicherung als auch für die Forschung gibt es in der EU und in Deutschland eine Vielzahl solcher Zulässigkeitstatbestände.

Unter *Qualitätssicherung* als Zwecksetzung wird dabei nach dem hier entwickelten Verständnis jede Datenverarbeitung zu fassen sein, die der Abbildung, Sicherung und Verbesserung der Qualität insbesondere der ärztlichen und pflegerischen Tätigkeiten dienen soll.

Wann eine Datenverarbeitung *Forschungszwecken* dient, ist weit zu verstehen und umfasst auch die technologische Entwicklung, die Demonstration, die Grundlagenforschung, die angewandte Forschung sowie privat finanzierte Forschung. Letztlich wird es dabei stets um einen weitgehend unabhängig erzielten und in einem möglichst transparenten Prozess gewonnenen Erkenntnisgewinn für die Allgemeinheit gehen, so dass insbesondere eine grundsätzliche Publikationsabsicht für das Vorliegen wissenschaftlicher Zwecke spricht.

Gerade wenn bei der Datenverarbeitung zur Qualitätssicherung verallgemeinerungsfähige Erkenntnisse gewonnen werden, die sodann wissenschaftlich ergründet werden sollen, *so können die Zwecksetzungen Qualitätssicherung und Forschung durchaus sukzessive aufeinanderfolgen*. Das Datenschutzrecht erlaubt dies, soweit ein *Zulässigkeitstatbestand* vorliegt.

Sollen die bereits im Rahmen der Versorgung und Behandlung erhobenen Daten durch dieselbe Gesundheitseinrichtung genutzt werden, ist § 6 GDNG hierfür als Zulässigkeitstatbestand anwendbar.

Die erforderliche *Zweckänderung* ist wegen der Privilegierung der Wissenschaft grundsätzlich erlaubt.[88] Selbstverständlich müssen Forschende dabei alle weiteren Vorgaben des Datenschutzrechts beachten, also etwa geeignete technische und/oder organisatorische Maßnahmen zum Schutz der betroffenen Personen ergreifen.

Ob hieraus hervorgehende Studien *retrospektiv* (=Datenerhebung vor Hypothesenerstellung/Studienbeginn) angelegt sind oder nicht, ist dabei nach dem hier vorliegenden Verständnis keine primär datenschutzrechtliche Frage. Denn nach der Sachverhaltsbeschreibung geht es stets um die Verarbeitung von bestehenden Daten aus der Versorgung oder Qualitätssicherung. Sollen für eine prospektive Studie (d.h. Datenerfassung nach Hypothesenerstellung/Studienbeginn) neue Daten erhoben werden, so kann hierfür etwa im Wege der (medizin-ethisch oft nötigen und datenschutzrechtlich sinnvollen) Einwilligung das Einverständnis der Studienteilnehmer erfragt werden.

[88] gem. Art. 5 Abs.1 lit. b Hs. 2 DS-GVO

Deutsche Gesellschaft
für Innere Medizin e.V.
Irenenstrasse 1
65189 Wiesbaden

www.dgim.de
info@dgim.de

Tel: +49 611 205 80 40 0
Fax: +49 611 205 80 40 46

DGIM Deutsche Gesellschaft für Innere Medizin