

# Potenziale und Herausforderungen durch sekundäre Nutzung von Gesundheitsdaten

## Potentials and challenges through secondary use of health data

Authors Names: Dominik Schrahe\*, Velko Iltchev†, Thomas Städter\*

\*FOM Hochschule für Oekonomie und Management, Institut für Gesundheit und Soziales, Standort München, dominik.schrahe@fom-net.de/ thomas.staedter@fom.de

†Technische Universität Sofia, Fakultät für Deutsche Ingenieur- und Betriebswirtschaftsusbildung, velko.iltchev@fdiba.tu-sofia.bg

**Zusammenfassung** — Zur Behandlung von Patienten werden von behandelnden Ärzten seit jeher Daten zu Lebensumständen, Erkrankungen und Medikation erhoben. Neben der akuten Behandlung eines Einzelpatienten, können derartige Daten jedoch auch in der Versorgungsforschung, zur Medikamentenentwicklung oder Optimierung von Therapien relevant sein (sekundäre Nutzung) [1][2]. Vor dem Hintergrund aktueller politischer Initiativen wie dem deutschen Gesundheitsdatennutzungsgesetz sowie dem Electronic Health Data Space (EHDS) der EU-Kommission wird in diesem Beitrag der potenzielle Nutzen von Big Data Analysen und Machine Learning Verfahren im Gesundheitswesen aufgezeigt. Diese sind in der Lage Diagnosen, Medikation sowie Therapien prinzipiell und spezifisch auf den einzelnen Patienten abgestimmt zu optimieren und damit Kosten im Gesundheitswesen zu senken.

**Abstract** — For the treatment of patients, data on living conditions, diseases and medication have always been collected by physicians. In addition to the acute therapy of an individual patient, however, such data can also be relevant in healthcare research, for drug development or optimization of therapies in general (secondary use) [1][2]. Considering current political initiatives such as the German Health Data Usage Act and the Electronic Health Data Space (EHDS) of the EU Commission, this paper highlights the benefits of Big Data analytics and machine learning methods in healthcare. These can optimize diagnoses, medication, and therapies, both in general and specifically adapted to the individual patient, and thus reducing costs in the healthcare system.

### I. EINLEITUNG

Obwohl bei jedem Behandlungsprozess Behandlungsdaten von verschiedenen Akteuren des Gesundheitswesens erzeugt werden, ist der uneingeschränkte Austausch dieser Daten zwischen Ärzten, Kliniken, Pharmaunternehmen und weiteren Beteiligten aufgrund der Sensibilität der Daten sowie technischer Hürden komplex [3]. Während der vergangenen COVID-19-Pandemie war das Konfliktfeld Datenschutz vs. Datennutzungsinteresse besonders sichtbar, da Daten über die variantenabhängige Verbreitung des Virus sowie über Impfeffekte in Deutschland fehlten [4][5]. Auch außerhalb von Pandemien klagen Forschende in Deutschland regelmäßig über uneinheitliche Auslegungen von Datenschutzgesetzen bei der Datenbeschaffung sowie über Unterschiede zwischen Bundesländern aufgrund von 13 verschiedenen Krankenhausgesetzen, was Deutschland für internationale Forschungsprojekte unattraktiv macht [6]. Gleichzeitig führen mutmaßliche Risiken durch Deanonymisierung in vermeintlich anonymen Patientendatensätzen zu teils emotionalen Debatten [7]. Der Deutsche Ethikrat kritisiert dabei nicht den Datenschutz an sich, sondern dessen Umsetzung, die für betroffene Patienten vielfach zu kompliziert ist und die in der Folge Einwilligungserklärungen ohne Prüfung unterschreiben [8]. Ein Opt-Out auf die Verwendung der eigenen Gesundheitsdaten zu Forschungszwecken solle zwar niederschwellig möglich sein, allerdings ist der erwartete Nutzen für die Allgemeinheit gegen die Risiken der datenspendenden Patienten abzuwägen [9]. Als Schlagwort fällt in diesem Zusammenhang regelmäßig der Begriff der Datensolidarität als Mittelweg zwischen Schadens- und Nutzenpotenzialen [5].

Neben Politik und Wissenschaft wirbt auch die Pharmaindustrie für eine bessere Verfügbarkeit von Gesundheitsdaten zu Forschungszwecken [10][2]. Der deutsche Bundesgesundheitsminister Lauterbach betont, dass eine Weiterentwicklung des Gesundheitswesens nur durch die Datennutzung und den Austausch von Daten in Europa gelingen wird [9]. Wissenschaftler beklagen, dass durch die aktuellen regulatorischen Hürden die Potenziale der Datennutzung nicht ausgeschöpft werden können [5].

In anderen europäischen Ländern werden Gesundheitsdaten aus nahezu flächendeckend verbreiteten elektronischen Patientenakten (ePA) zum Teil schon seit Jahren zu Forschungszwecken weiterverwendet [11]. So werden in Dänemark Gesundheitsdaten aus verschiedenen Institutionen mittels einer ‚central person registration number‘ zu Forschungszwecken weitergegeben, sodass eine Verknüpfung zwischen verschiedenen Registern möglich ist [12].

Diese Arbeit soll sich mit den Potenzialen der Nutzung von Gesundheitsdaten in der Forschung sowie aktuellen Herausforderungen in diesem Umfeld beschäftigen. Insbesondere soll die Verfügbarkeit neuer Datenkategorien in einem bisher unbekanntem Ausmaß in Hinblick auf technische Möglichkeiten wie Machine Learning sowie Big Data im Gesundheitswesen beleuchtet werden. Folgende Forschungsfragen ergeben sich aus den vorausgegangenen Erläuterungen

- Welche Herausforderungen bestehen aktuell in der Nutzung von Gesundheitsdaten?
- Welche Potenziale ergeben sich für das Gesundheitswesen durch die sekundäre Nutzung von Gesundheitsdaten?

- Für welche Akteure des Gesundheitswesens können Technologien wie Big Data und Machine Learning einen Mehrwert bringen?

## II. SEKUNDÄRNUTZUNG VON GESUNDHEITSDATEN IM ELECTRONIC HEALTH DATA SPACE DER EU

Auch ein aktueller Entwurf der EU-Kommission für den Electronic Health Data Space (EHDS) trägt zur gesellschaftlichen und politischen Diskussion über die Sekundärnutzung von Gesundheitsdaten zur Forschung bei. Die Pläne der EU-Kommission sehen neben einem zukünftigen europaweiten Austausch von Gesundheitsdaten zu Behandlungszwecken (Primärnutzung), was einer europaweiten ePA nahekommt, explizit auch die Nutzung zu Forschungszwecken vor. Dabei werden umfangreiche Anforderungen an die IT-Sicherheit und den Datenschutz gestellt, wonach unter anderem Anonymisierungs- oder Pseudonymisierungstechniken zum Einsatz kommen und je nach Risiko für die Privatsphäre auch Opt-Out Verfahren für Patienten angeboten werden müssen [13]. Datenzugriffe für Forschende erfolgen im Falle pseudonymisierter Daten innerhalb einer sicheren Verarbeitungsumgebung, die allerdings neben der Analyse lediglich den Export von Ergebnissen erlaubt. Nur anonymisierte Datensätze dürfen durch Forschende heruntergeladen und lokal analysiert werden [8][9]. Nationale Ämter sind für Zugriffsanfragen seitens der Forschenden vorgesehen, auch wenn die Daten nicht vollständig bei diesen Ämtern, sondern bei verschiedenen Einrichtungen gespeichert sind. Die Bereitstellung erfolgt im Anschluss ebenfalls über nationale Forschungsdatenzentren, die die Forschungsdaten von dezentralen Einrichtungen zusammenfassen und für die Nutzung vorbereiten [14]. Somit erfolgt eine Vereinfachung, sowohl für die Datenhaltenden- als auch für Datennutzenden Instanzen, da beide lediglich mit einer Stelle kommunizieren müssen [9]. Für Forschende werden damit Zugänge zu Forschungsdaten vereinfacht und kostengünstiger gestaltet. Durch die verbesserte Datenverfügbarkeit verspricht sich die EU-Kommission positive Auswirkungen auf Innovationen im Gesundheitswesen sowie verbesserte gesundheitspolitische Entscheidungen. Gemäß Berechnungen sollen durch die verbesserte Nutzung von Gesundheitsdaten für Forschung, Innovation und Politikgestaltung im Laufe von 10 Jahren 5,4 Mrd. Euro eingespart werden [15].

## III. PRAKTISCHE HERAUSFORDERUNGEN IN DER SEKUNDÄRNUTZUNG VON GESUNDHEITSDATEN

Da insbesondere im deutschen Gesundheitswesen die Daten weitgehend dezentralisiert sind, regelmäßig nicht digitalisiert werden und ebenso wenig in standardisierten Formaten vorliegen, ist es derzeit schwierig, eine flächendeckende Nutzung großer Datenbestände aus dem Gesundheitswesen zu ermöglichen. Die weitere Verbreitung der ePA in Deutschland, etwa durch den angestrebten Wechsel zum Opt-Out-Modell [16] sowie die Einführung des EHDS, werden in den kommenden Jahren die Rahmenbedingungen für die Sekundärnutzung von Gesundheitsdaten verbessern. Auch das im Entwurf befindliche Gesundheitsdatennutzungsgesetz wird einen Beitrag dazu leisten [5].

Derzeit wird die Sekundärnutzung von Gesundheitsdaten in Deutschland durch siloartige Datenhaltungen in unterschiedlichen Institutionen sowie mangelnde Standardisierung bei Datenformaten grundlegend erschwert [8]. Institutionen

oder Hersteller verwenden regelmäßig eigene Formate, was institutionsübergreifende Datenanalysen erschwert [17]. International werden Nomenklaturen wie SNOMED oder ICD verwendet, die auch in Deutschland institutionsübergreifende Forschungsaktivitäten erleichtern würden [1][6]. Sowohl der EHDS als auch das deutsche Gesundheitsdatennutzungsgesetz strebt daher eine verbesserte technische Standardisierung an. Trotz technischer Interoperabilität sind jedoch Verzerrungseffekte zu befürchten, nachdem bisherige Bemühungen lediglich ePA- und Abrechnungsdaten von gesetzlich versicherten Personen umfassen. Bestimmte Berufsgruppen sowie Beamte in Deutschland werden in Forschungsdatensätzen zumindest stark unterrepräsentiert sein [5].

Eine weitere praktische Herausforderung ergibt sich aus dem Zielkonflikt von Datenschutz- und Datennutzungsinteressen, da sich etwa die Anwendung von Anonymisierungstechniken negativ auf den Erkenntnisgewinn bei Big Data Analysen auswirkt [3]. Auch fortschrittlichere Techniken wie Differential-Privacy oder K-Anonymisierung führen durch Verrauschung der Daten zu einer statistischen Unsicherheit, die negative Auswirkungen auf die Analyseergebnisse haben kann [5]. Diesbezüglich ist neben weiteren Forschungen zu Sekundärnutzungsfreundlichen Pseudonymisierungs- und Anonymisierungsverfahren auch eine breite gesellschaftliche Debatte erforderlich, welche Privatsphäre-Risiken tolerabel sind, um die im Folgenden dargestellten Nutzeneffekte realisieren zu können.

## IV. POTENZIALE VON BIG DATA UND MACHINE LEARNING IM GESUNDHEITSWESEN

Die verbesserte Verfügbarkeit von Gesundheitsdaten wirkt sich besonders positiv aus, wenn die beiden technischen Innovationen Machine Learning und Big Data genutzt werden. Bei letzterem handelt es sich um Techniken zur Analyse von großen, häufig semi- oder unstrukturierten Datenmengen, die mit herkömmlichen Technologien kaum auswertbar wären. Dabei wird darauf abgezielt, Muster wie etwa Zusammenhänge aus Medikationen und Folgeerkrankungen zu erkennen, um bisher unbekannte Nebenwirkungen aufzudecken [18].

Beide Technologieansätze wurden in Deutschland bisher nur in vereinzelten Forschungsprojekten mit begrenzten Datenbeständen angewandt. Angesichts der Stärken beider Technologien wäre jedoch die Verarbeitung größerer Datenmengen, z.B. aus den ePAs aller Patientinnen und Patienten von Vorteil. Die Fähigkeit und auch die Notwendigkeit, große Datenmengen zu Analyse- bzw. Trainingszwecken zu verarbeiten, unterscheidet Big Data und Machine Learning-Ansätze deutlich von bisherigen statistischen Analyseverfahren, die auf die Verarbeitung kleinerer Datenmengen ausgelegt waren [10]. So ist es bei der Analyse von Behandlungsdaten einer Arztpraxis oder eines Krankenhauses kaum möglich, Optimierungspotenziale in der Therapie zu entdecken. Wenn jedoch die Behandlungsdaten aller Praxen sowie aller Kliniken eines Landes analysiert werden könnten, ist es wahrscheinlich möglich, Risiken bestimmter Behandlungsmethoden oder Vorteile alternativer Therapiemethoden zu identifizieren [1]. Auch kann eine solche umfangreiche Datenbasis erstmals für die Entwicklung von Medikamenten und Therapien für seltene Erkrankungen verwendet werden, für die bisher mangels Datenverfügbarkeit kaum Untersuchungen möglich waren [19]. Zudem werden Trendanalysen, etwa bei periodisch auftretenden Grippewellen aber auch Pandemien durch den Zugriff auf anonymisierte Daten aus allen ePAs eines Landes deutlich erleichtert [1].

Aktuelle Studiendesigns wie z.B. klinische Studien arbeiten in der Regel mit eng abgegrenzten Datensätzen, die nur zu diesem Zweck erhoben oder explizit auf den erforderlichen Umfang gefiltert wurden. Echte Versorgungsdaten werden von Experten als wertvoller angesehen, weil sie die Versorgungspraxis besser abbilden als Studiendaten [8]. Während beispielsweise klinische Studien in der Regel hypothesengesteuert sind, das heißt analytische Verfahren werden eingesetzt, um zuvor aufgestellte Hypothesen zu falsifizieren, sind Big Data-Vorgehensmodelle in der Regel hypothesengenerierend. Demnach werden große Datenmengen nach bisher unbekanntem Mustern durchsucht. So umfasst der Analyseumfang die gesamten ePA-Daten eines Landes anstelle von Behandlungsdaten einer spezifischen Erkrankung. Dabei ist es denkbar, dass auch Daten in die Analysen einfließen, die bisher nicht mit medizinischen Datensätzen verknüpft werden konnten. Die deutsche ePA und auch der Entwurf des EHDS sehen für alle ePAs in Europa den Datenimport aus Fitnesstrackern oder Smartwatches vor [6]. Aufgrund der Vielfalt der auf diesem Wege verfügbaren Daten ergeben sich neue Einblicke in die Lebensumstände erkrankter Menschen und somit möglicher Risikofaktoren, die Erkrankungen hervorrufen oder fördern [20]. Hypothesengenerierende Forschungen können dabei als Ergänzung, jedoch nicht als Ersatz von hypothesengetriebenen Ansätzen betrachtet werden [1], insbesondere weil dadurch zwar Korrelationen aufgedeckt, jedoch keine Kausalität nachgewiesen werden kann [21].

Neben gezielten Analysen können Daten aus ePAs auch auf bisher unbekanntem Muster untersucht werden. Diesem sogenannten EHR-Mining (Electronic Health Record Mining in Analogie zum Data Mining) werden Potenziale bei der Erkennung von Medikations- und Therapienebenwirkungen zugeschrieben, indem Abrechnungs- und Behandlungsdaten in der ePA betrachtet werden [17]. Zwar handelt es sich auch bei derartigen Erkenntnissen zunächst um Hypothesen, diese können jedoch im Nachgang durch klassische Untersuchungsdesigns weiter geprüft werden. Zudem ist es möglich, Zusammenhänge zwischen bisher unbekanntem Mustern in medizinischem Bildmaterial und Krankheitsverläufen oder bisher nicht zuordenbaren Symptomen zu erkennen [1]. In der Folge können Behandlungsmethoden qualitativ optimiert, aber auch Kosten für unnötige Therapien gesenkt werden [20].

Anstelle von Big Data Analysen können Gesundheitsdaten zunächst auch zum Training von Machine Learning Modellen verwendet werden. Klassischerweise arbeitet die IT mit deterministischen Algorithmen, die Inputdaten entsprechend der Anweisungen des Programmcodes verarbeiten. Machine Learning versucht dagegen anhand vorhandener Erkenntnisse bzw. Trainingsdaten Modelle zu entwickeln, die in der Lage sind, komplexe Fragestellungen eigenständig zu beantworten. So ist es beispielsweise denkbar, ein Modell mit medizinischem Bildmaterial so zu trainieren, dass es in der Lage ist, weiteres Bildmaterial zu analysieren und zu beurteilen, ob Hinweise auf Erkrankungen vorliegen [22]. Eine besondere Bedeutung kommt hierbei den Trainingsdaten zu, wobei für qualitativ hochwertige Modelle sehr große Datenmengen ohne Verzerrungen benötigt werden [3]. Hochwertige Modelle sind dann in der Lage, beispielsweise anhand von Laborergebnissen oder medizinischem Bildmaterial, Diagnosen zu treffen oder genetische Proben nach Auffälligkeiten zu untersuchen [23]. Die Genauigkeit von Machine Learning Modellen übertrifft manchmal die durch menschliche Experten gestellten Diagnosen [3]. In der Behandlungspraxis sind Machine Learning Modelle in der Lage, Vorab-Diagnosen zu stellen oder relevante Informa-

tionen für behandelnde Ärzte aus umfassenden ePAs zu extrahieren [1]. Letzteres wird mit wachsenden Datenbeständen immer wichtiger, weil es Ärzten im Praxis- bzw. Klinikalltag kaum möglich ist, die Menge der via ePA zur Verfügung stehenden Gesundheitsdaten manuell zu sichten und zu bewerten [25]. Hier kann zukünftig eine spezifische Ausprägung maschineller Lernverfahren helfen: Sogenannte ‚Large Language Models‘ könnten bei Bedarf große Datenbestände explizit auf bestimmte Zusammenhänge untersuchen oder behandelnde Ärzte durch inhaltliche Zusammenfassungen unterstützen [9].

Ein weiterer Nutznießer der Verfügbarkeit von Gesundheitsdaten zum Training von Machine Learning Modellen ist die Pharmaindustrie. Relevant ist dies insbesondere für die Bereiche Arzneimittelentwicklung sowie der Erforschung weiterer Anwendungsfelder von Medikamenten. Bereits heute werden Big Data Methoden in der Vorauswahl und Priorisierung vielversprechender Wirkstoffkandidaten oder zur Abschätzung der Toxizität dieser verwendet. Im Rahmen des sogenannten ‚virtuellen Screenings‘ kann darüber hinaus die Wirkungsweise bestehender Medikamente zur Behandlung anderer Krankheiten simuliert werden [21]. Weitere Potenziale bestehen insbesondere durch die verbesserte Verfügbarkeit von genetischen Daten, um Machine Learning Methoden für die Wirkstoffidentifikation zu nutzen. Aufgrund der sehr hohen Kosten in der Medikamentenentwicklung sind Effizienzsteigerungen in diesem Bereich hoch rentabel [2].

Einen Schritt weiter als die allgemeine Medikamentenentwicklung geht die sogenannte ‚Precision Medicine‘, wobei hier kein datengetriebener Erkenntnisgewinn für die Allgemeinheit im Fokus steht, sondern die Anpassung von Therapien und Medikation auf individuelle Patienten oder spezifische Patientengruppen und ihre Bedürfnisse [21]. So soll die richtige Behandlung zur richtigen Zeit für den individuellen Patienten zur Verfügung stehen [24][25]. In diesem Zusammenhang werden beispielsweise eine gezieltere Auswahl von Medikamenten sowie eine verbesserte Dosierung oder eine personalisierte Therapie auf Grundlage von Erfahrungen aus bestehenden Behandlungsdaten angestrebt [20][26]. Auch ermöglicht die Precision Medicine Patienten mit bestimmten Krankheitsrisiken gezielter vorzuwarnen und zu individuellen Präventionsmaßnahmen zu ermutigen [20].

## V. FAZIT UND AUSBLICK

Zusammenfassend lässt sich sagen, dass die Sekundärnutzung von Gesundheitsdaten vielfältige Potenziale bietet, die von der Unterstützung bei der Diagnose über Optimierung von Medikationen und Therapien bis hin zur individuellen Früherkennung von Krankheitsrisiken reicht. Zur Nutzung dieser mangelt es aktuell sowohl an durchgängiger digitaler Datenerfassung im Gesundheitswesen als auch der Interoperabilität und Standardisierung der Datenobjekte. Auch wird zumindest die deutsche ePA bisher patientenseitig nur wenig genutzt, sodass noch keine für die Forschung verwendbare Datenbasis entstehen konnte. Aktuelle Vorhaben wie die Standardisierung von Medizinischen Informationsobjekten in der deutschen ePA sowie die Bestrebungen des EHDS setzen hier an, um die Rahmenbedingungen für eine verbesserte Datenverfügbarkeit zu schaffen. Unter Berücksichtigung der Komplexität dieses Vorhabens allein auf nationaler Ebene ist davon auszugehen, dass eine Standardisierung von Datenformaten auf europäischer Ebene noch Jahre in Anspruch nehmen wird. Erschwerend kommen deutlich unterschiedliche Entwicklungsstände in der

Digitalisierung des Gesundheitswesens hinzu, die parallel angegangen werden müssen. Daraus lassen sich drei Handlungsfelder identifizieren, die durch weitere politische Maßnahmen zu fördern und durch die weiteren Akteure des Gesundheitswesens umzusetzen sind:

1. Konsequente Weiterverfolgung des Ansatzes des EHDS zur Einführung europäischer Standards und der Etablierung einer Art europaweiten ePA
2. Flächendeckende Ermöglichung von Datenbezug zu Forschungszwecken aus dieser ePA für Forschungseinrichtungen und Wirtschaft sowohl durch gesetzliche Rahmenbedingungen als auch technische Standardisierung und
3. Ausgestaltung der Datenweitergabe unter angemessenen IT-Sicherheits- und Datenschutzbedingungen, um die Akzeptanz der Sekundärnutzung von Gesundheitsdaten in der Bevölkerung zu wahren.

Vorteile ergeben sich aus einer deutlich verbesserten Sekundärnutzung von Gesundheitsdaten für traditionelle Versorger wie Ärzte und Kliniken aber auch für Wissenschaftler und die Pharmaindustrie. Technologische Innovationen wie Big Data und Machine Learning bilden die Grundlage für die beschriebenen Potenziale. Die derzeit prognostizierten Kosteneinsparungen sowie die weiteren Nutzeneffekte können nur den aktuellen Stand widerspiegeln. Es ist zu erwarten, dass die Weiterentwicklung der technischen Möglichkeiten in Zukunft neue Nutzenpotenziale aufzeigen wird.

## VI. QUELLEN

[1] Hulsen, T., Jamuar, S. S., Moody, A. R., Karnes, J. H., Varga, O. et al., "From big data to precision medicine," *Frontiers in medicine*, vol. 6, 34, March 2019.

[2] Qian, T., Zhu, S., and Hoshida, Y., "Use of big data in drug development for precision medicine: an update," *Expert review of precision medicine and drug development*, vol. 4, pp. 189-200, 2019.

[3] Guse, R., Thiebes, S., Hennel, P., Rosenkranz, C., and Sunyaev, A., "Datenmarktplätze für Künstliche Intelligenz im Gesundheitswesen: Potenziale, Herausforderungen und Strategien zur Bewältigung," *HMD Praxis der Wirtschaftsinformatik*, vol. 59, pp. 1527-1544, 2022.

[4] ExpertInnenrat der Bundesregierung zu COVID-19, „4. Stellungnahme des ExpertInnenrates der Bundesregierung zu COVID-19 - Dringende Maßnahmen für eine verbesserte Datenerhebung und Digitalisierung,“ <https://www.bundesregierung.de/resource/blob/974430/2000794/f189a6b7b0f581965f746e957db90af7/2022-01-22-nr-4expertenrat-data.pdf>, 2022, Zugegriffen: 26. Januar 2022.

[5] Schmitt, J., Bierbaum, T., Geraedts, M., Gothe, H., Härter, M. et al., „Das Gesundheitsdatennutzungsgesetz–Potenzial für eine bessere Forschung und Gesundheitsversorgung,“ *Das Gesundheitswesen*, vol. 85 pp. 215-222, 2023.

[6] Welzel, C., Cotte, F., Brückner, S., Lauber-Rönsberg, A., Muck, J. et al., „Gesundheitsdaten: Paradigmenwechsel steht auch Deutschland bevor,“ *Deutsches Ärzteblatt*, vol. 120, A-1162, 2023.

[7] Steiner, F., „Verfahren gegen zentrale Massenspeicherung von Gesundheitsdaten ruht,“ <https://heise.de/-7516830>, 2023, Zugegriffen: 20. März 2023.

[8] Lau, T. and Richter-Kuhlmann, E., „Gesundheitsdaten: Ein Balanceakt zwischen Datennutzung und Datenschutz,“ *Deutsches Ärzteblatt*, vol. 120, A-598, 2023.

[9] Lau, T., „Gesundheitsdateninfrastruktur: Ein neues Datenzeitalter,“ *Deutsches Ärzteblatt*, vol. 120, A-1154, 2023.

[10] Nature Communications Editorial, "Data sharing and the future of science," *Nature Communications*, vol. 9, 2817, 2018.

[11] Thiel, R., Deimel, L., Schmidtmann, D. et al., „#Smart Health Systems – Digitalisierungsstrategien im internationalen Vergleich,“ Gütersloh: Bertelsmann Stiftung, 2018, pp. 290-316

[12] Holm, S. and Ploug, T., "Big Data and health research – The governance challenges in a mixed data economy," *Journal of bioethical inquiry* vol. 14 pp. 515–525, 2017.

[13] Europäische Kommission, „Verordnung des Europäischen Parlaments und des Rates über den europäischen Raum für Gesundheitsdaten,“ [https://eur-lex.europa.eu/resource.html?uri=cellar:dbfd8974-cb79-11ec-b6f4-01aa75ed71a1.0003.02/DOC\\_1&format=PDF](https://eur-lex.europa.eu/resource.html?uri=cellar:dbfd8974-cb79-11ec-b6f4-01aa75ed71a1.0003.02/DOC_1&format=PDF), 2022, Zugegriffen: 15. März 2023.

[14] Towards the European Health Data Space, "TEHDAS' proposals for the implementation of EHDS technical infrastructure," <https://tehdas.eu/results/tehdas-proposals-for-the-implementation-of-ehds-technical-infrastructure/>, 2023, Zugegriffen: 22. August 2023.

[15] Europäische Kommission, „Europäischer Raum für Gesundheitsdaten # EUDigitalHealth,“ <https://ec.europa.eu/commission/presscorner/api/files/attachment/872467/EHDS-factsheet-DE.pdf.pdf>, 2022, Zugegriffen: 22. August 2023.

[16] Gematik GmbH, „Pressemitteilung | Gesellschafter beauftragen gematik mit Prüfauftrag für ‚Opt-out-ePA‘,“ <https://www.gematik.de/newsroom/news-detail/pressemitteilung-gesellschafter-beschliessen-opt-out-epa>, 2022, Zugegriffen: 15. März 2023.

[17] Pigeot, I., Fröhlich, H., Intemann, T., Prause, G. and Wright, M. N., „KI und die Nationale Forschungsdateninfrastruktur für personenbezogene Gesundheitsdaten (NFDI4HEALTH),“ in Dössel, O., Schäffter, T., Rutert, B. (ed.), „Künstliche Intelligenz in der Medizin,“ Berlin: Berlin-Brandenburgische Akademie der Wissenschaften, 2023, pp. 62-74.

[18] Berros, N., El Mendili, F., Filaly, Y., and El Bouzekri El Idrissi, Y., "Enhancing digital health services with big data analytics," *Big data and cognitive computing*, vol. 7, 64, 2023.

[19] Sardana, D., Zhu, C., Zhang, M., Gudivada, R. C., Yang, L. et al., "Drug repositioning for orphan diseases," *Briefings in bioinformatics*, vol. 12, pp. 346-356, 2011.

[20] Pastorino, R., De Vito, C., Migliara, G., Glocker, K., Binenbaum, I. et al., "Benefits and challenges of Big Data in healthcare: an overview of the European initiatives," *European journal of public health*, vol. 29, pp. 23-27, 2019.

[21] König, H., Heyen, N. B., Baumann, M., Schmoch, U., Bachlechner, D. et al., „Künstliche Intelligenz in der genomischen Medizin-Potenziale und Handlungsbedarf,“ Karlsruhe: Karlsruher Institut für Technologie, 2021.

[22] Sharifani, K. and Amini, M., "Machine Learning and Deep Learning: A Review of Methods and Applications," *World Information Technology and Engineering Journal*, vol. 10, pp. 3897-3904, 2023.

[23] Mathew, C. J., David, A. M. and Mathew, C. M. J., "Artificial intelligence and its future potential in lung cancer screening," *EXCLI Journal*, vol. 19, pp. 1552-1562, 2020.

[24] Ashley, E. A., "Towards precision medicine," *Nature Reviews Genetics*, vol. 17, pp. 507-522, 2016.

[25] Cammarota, G., Ianiro, G., Ahern, A., Carbone, C., Temko, A., et al., "Gut microbiome, big data and machine learning to promote precision medicine for cancer," *Nature reviews gastroenterology & hepatology*, vol. 17, pp. 635-648, 2020.

[26] You, J., McLeod, R. D. and Hu, P., "Predicting drug-target interaction network using deep learning model," *Computational biology and chemistry*, vol. 80, pp. 90-101, 20.