

Bundesgesundheitsbl  
<https://doi.org/10.1007/s00103-024-03880-y>  
 Eingegangen: 13. Dezember 2023  
 Angenommen: 4. April 2024

© The Author(s) 2024



Toralf Kirsten<sup>1,2</sup> · Philip Kleinert<sup>3</sup> · Marie Gebhardt<sup>3</sup> · Johannes Drepper<sup>3</sup> ·  
 Anne-Katrin Andreeff<sup>4</sup> · Fabian Prasser<sup>5</sup> · Oliver Kohlbacher<sup>6,7,8</sup>

<sup>1</sup> Institut für Medizinische Informatik, Statistik und Epidemiologie, Universität Leipzig, Leipzig, Deutschland; <sup>2</sup> Medizininformatikzentrum, Dept. Medical Data Science, Universitätsklinikum Leipzig, Leipzig, Deutschland; <sup>3</sup> TMF – Technologie- und Methodenplattform für die vernetzte medizinische Forschung e. V., Berlin, Berlin, Deutschland; <sup>4</sup> Institut für Medizinische Informatik und Biometrie, Medizinische Fakultät Carl Gustav Carus der Technischen Universität Dresden, Dresden, Deutschland; <sup>5</sup> Center of Health Data Science, Medizininformatik, Berliner Institut für Gesundheitsforschung in der Charité – Universitätsmedizin Berlin, Berlin, Deutschland; <sup>6</sup> Institut für Biomedizinische Informatik, Universität Tübingen, Tübingen, Deutschland; <sup>7</sup> Fachbereich Informatik, Universität Tübingen, Tübingen, Deutschland; <sup>8</sup> Institut für Translationale Bioinformatik, Universitätsklinikum Tübingen, Tübingen, Deutschland

# Grundlagen für die wissenschaftliche Nutzung umfangreicher Versorgungsdaten in Deutschland – Ergebnisse der AG Data Sharing der Medizininformatik-Initiative

## Einleitung

Klinische Daten zu gesundheitlichen Aspekten bilden die Grundlage für die angewandte medizinische Forschung. Sie werden in all ihrer Vielfalt in verschiedenen Institutionen und für verschiedene Zwecke aufgenommen und eingesetzt.

Klinische Studien fokussieren hauptsächlich auf klinische Fragestellungen, z. B. die Wirksamkeit eines neuen Medikaments, einer neuen Therapie oder eines angepassten Therapie-Settings. Dagegen zielen epidemiologische Studien oftmals auf die Erhebung der Prävalenzen von ausgewählten Erkrankungen in der Bevölkerung ab. Gezielte medizinische Studien haben oftmals sehr eingeschränkte Zielstellungen, sodass die dafür erhobenen Daten in ihrer Aussagekraft limitiert sind. Die Routinedaten, die während der Aufenthalte von Patientinnen und Patienten in Kliniken erhoben werden, dienen primär der Dokumentati-

on des Arzt-Patienten-Kontakts in Hinsicht auf die ausgeführte Diagnostik und die durchgeführten Behandlungen (Therapien). Gleichwohl stellen sie eine fundierte Datenquelle für die medizinische Forschung dar, auch wenn ihr primärer Zweck sich von dem der o. g. Studien unterscheidet. Routinedaten aus der gelebten klinischen Praxis sind umfangreicher, heterogener und deutlich komplexer als Studiendaten.

Routinedaten aus den Dokumentationssystemen der Krankenhäuser sind meist umfangreicher und aussagekräftiger als Krankenkassendaten, die zudem aufgrund ihres Zwecks der Abrechnung erbrachter Leistungen nicht repräsentativ sind. Letztlich erfordert die medizinische Forschung den Vergleich und die Zusammenschau unterschiedlicher Datenquellen.

Die Forschung mit klinischen Routinedaten kann der beständigen Verbesserung der Patientensicherheit, der schnelleren Diagnostik und Behandlung und damit einem erstrebenswerten Gemeinwohl dienen. In vielen Fällen sind mul-

tizentrische Auswertungen in der Lage, die nötigen Fallzahlen für eine statistisch robuste Evidenzgenerierung zu ermöglichen. Die Verwendung klinischer Routinedaten für medizinische Forschungszwecke über Klinikbereiche, Standorte und Bundesländergrenzen hinweg bedarf jedoch einer Harmonisierung. Da es sich um personenbezogene Gesundheitsdaten handelt, die besonders schützenswert sind, gelten besondere Bedingungen für die Nutzung, so gibt es verschiedene europäische, landes- und bundeslandspezifische Regelungen zum Datenschutz, was Auswirkungen auf die Nutzung der Daten hat. Erklärtes Ziel der im Jahr 2016 gegründeten Medizininformatik-Initiative (MII) ist es, den Konflikt zwischen vollständigem Datenschutz und bestmöglicher Nutzung von Daten zur medizinischen Forschung zu lösen und bundeslandübergreifend in den deutschen Universitätskliniken umzusetzen. Datenintegrationszentren (DIZ) werden aufgebaut, um Forschenden den Zugriff auf die klinischen Routinedaten zu ermöglichen. Die Arbeitsgruppe *Data Sharing*,

Die Autoren T. Kirsten und P. Kleinert teilen sich die Erstautorenschaft.

als Teil einer übergreifenden Koordinierungsstruktur innerhalb der MII, hat zum Ziel, organisatorische Grundlagen für das deutschlandweite Teilen und Nutzen der Daten zu schaffen.

Das Teilen von Daten ist oft mit hohen technischen, rechtlichen, aber auch kulturellen Hürden verbunden. Patientinnen und Patienten unterstützen die Nutzung ihrer Daten für die Forschung überwiegend, sofern entsprechende Maßnahmen zum Schutz ihrer Daten gegeben sind [2]. Die AG Data Sharing hat Prozesse und Methoden mitgestaltet und etabliert, die multizentrisches Datenteilen sicher und effizient ermöglichen und gleichzeitig die Bedenken gegen die Datennutzung berücksichtigen.

Die FAIR-Prinzipien [1] bilden die Grundlage für das heutige Data Sharing. Die 4 Buchstaben von FAIR stehen für die Anforderungen Auffindbarkeit („findable“), Zugriff („accessible“), Interoperabilität („interoperable“) und Wiederverwendbarkeit („reusable“). Während die Interoperabilität ein Grundanliegen der Arbeitsgruppe (AG) Interoperabilität in der MII-Koordinierungsstruktur ist, obliegen die Auffindbarkeit, der Zugriff und die Regelungen zur Wiederverwendung der AG Data Sharing, wobei die Wiederverwendung zum Teil auch von der AG Consent bearbeitet wird, die sich der informierten, breiten Einwilligung der Patientinnen und Patienten widmet.

Mit diesem Artikel geben wir einen Überblick über die Struktur und Arbeitsweise der AG Data Sharing und fassen die wichtigsten Ergebnisse der AG im Rahmen der MII zusammen. Insbesondere steht die klinikübergreifende Datennutzung im Fokus, also die integrierte Nutzung von Daten mehrerer Kliniken.

## Struktur und Arbeitsweise der AG Data Sharing

Die AG Data Sharing wurde mit Beginn der Vorbereitungs- und Konzeptionsphase der MII im Jahr 2017 vom Nationalen Steuerungsgremium (NSG) einberufen. Sie hat die Aufgabe, übergreifende Aspekte des Teilens medizinischer Daten aus Universitätskliniken zu diskutieren und abzustimmen. Die AG ist in den Verbund verschiedener AGs

eingebettet, zu denen die AG Consent, die AG Interoperabilität, die AG Kommunikation und die AG externe Daten gehören. Diese AGs berichten ihre Ergebnisse dem NSG, welches diese beschließen kann und sie damit für die in der MII vertretenen Universitätskliniken verbindlich macht.<sup>1</sup>

Teil der AG Data Sharing sind sogenannte Task-Forces (TFs), die einzelne Themen im Detail diskutieren und bearbeiten. Dies sind derzeit die TFs „Use & Access“, „Widerruf“, „Prozessmodelle“, „Audit“ und „Verteilte Analysen“, die im Folgenden vorgestellt werden.

Die *TF Use & Access* widmet sich der Aufgabe, organisatorische Regelungen für einen standortübergreifenden Datenzugang zu konzipieren und zu etablieren und in einer Nutzungsordnung zu fassen. Dazu trägt langjährige Expertise aus Zentren bei, an denen große epidemiologische Studien durchgeführt werden, darunter insbesondere die Universität Greifswald mit den SHIP-Studien [3] sowie die Universität Leipzig mit dem LIFE – Leipziger Forschungszentrum für Zivilisationserkrankungen und den dort durchgeführten Studien (z. B. LIFE-Adult [4]). Als konzeptionelle Grundlage diente das „Open Archive Information System“ (OAIS [5]), das die 3 am Data-Sharing-Prozess beteiligten Strukturen und deren Aufgaben benennt: den Datenproduzenten („data producer“), den Datenkonsumenten („data consumer“) sowie die datenverwaltende Stelle.

Im Fokus der *TF Prozessmodelle* stehen die Prozesse des übergreifenden Data Sharing, die sich von der Vorbereitung und Planung bis hin zum abschließenden Datenaustausch erstrecken. Auf Basis der Prozessmodelle begann die Arbeit an deren erweiterten Versionen, in denen insbesondere Teilprozesse identifiziert und harmonisiert wurden, sodass diese standardisiert ablaufen können. Insbesondere wurden die Prozessmodelle mit der von der TF Use & Access entworfenen Nutzungsordnung be-

ständig abgeglichen. Die Prozessmodelle wurden in der „Business Process Modeling Notation“ (BPMN<sup>2</sup>) in Abstimmung mit lokalen Expertinnen und Experten aufgezeichnet.

Um den Fortgang der Aufbauarbeiten und damit den Zielerreichungsgrad an den Datenintegrationszentren zeitnah und regelmäßig überprüfen zu können, wurden von der *TF Audit* geeignete Qualitätsindikatoren geschaffen. Dazu wurde die Expertise aus dem Netzwerk der Koordinierungszentren für Klinische Studien (KKS) herangezogen. Diese Zentren unterliegen regelmäßigen Audits, d. h. Prüfungen, die von einem bestellten unabhängigen Prüfinstitut durchgeführt werden. Die mit der TF Audit geschaffenen Ergebnisse waren Grundlage für „Friendly Audits“ (Vorausaudits zur Feststellung der Zertifizierungsfähigkeit) an beinahe allen Standorten sowie für die Vorbereitung des Auditierungsprozesses, der 2021 vor Ende der ersten Förderperiode an den Datenintegrationszentren aller Konsortien von einem extern bestellten Prüfer durchgeführt wurde.

Die *TF Verteilte Analysen* arbeitet an Methoden und Strukturen, die eine standortübergreifende, dezentrale Analyse ermöglichen. Dazu wurden Infrastrukturen wie DataSHIELD [6], das Personal-Health-Train-Konzept [7, 8] und verschiedene Lösungen unter Nutzung von Secure Multi Party Computation (SMPC; [9–11]) untersucht. Eine Datenherausgabe ausschließlich in materieller Form würde für die Forschenden zu kurz greifen. Wissenschaftliches Personal kann bereits im Sinne der Eigenforschung (abhängig von den gesetzlichen Grundlagen je Bundesland) lokal am eigenen Klinikum auf diese Daten (ohne Einwilligung der Patientinnen und Patienten) zugreifen. Eine zentrale Datenverwaltung und -analyse ist jedoch nur mit Daten von Patientinnen und Patienten möglich, die ihre Einwilligung dazu erteilt haben.

Die *TF Widerruf* befasst sich mit dem Verfahren zur Berücksichtigung von Widerruf der Einwilligung von Patienten

<sup>1</sup> <https://www.medizininformatik-initiative.de/de/geschaeftsordnung-des-nationalen-steuerungsgremiums-der-deutschen-medizin-informatik-initiative>. Zugegriffen am 20.12.2023.

<sup>2</sup> <https://www.omg.org/spec/BPMN>. Zugegriffen am 20.12.2023.

oder Patientinnen in die Nutzung ihrer Daten für die Forschung. Der Prozess berücksichtigt die Rechte der Patientinnen und Patienten bei lokalen wie auch übergreifenden Forschungsprojekten.

## Ergebnisse der AG Data Sharing

In der AG Data Sharing und den etablierten Task-Forces wurde eine Vielzahl an Ergebnissen erzielt, von denen an dieser Stelle nur eine Auswahl wiedergegeben und kurz charakterisiert werden kann.

## Prozesse zur Datennutzung

Die medizinischen Behandlungsdaten der Universitätskliniken sind personalisiert. Forschende können deshalb nur unter Wahrung des Datenschutzes, also unter Beachtung bestimmter Restriktionen und Bedingungen, auf die Daten zugreifen. Die im Rahmen der MII geschaffenen Datenintegrationszentren (DIZ) der Standorte bereiten die Daten interoperabel (basierend auf Spezifikationen der AG Interoperabilität) auf und stellen erste Datenschutzmaßnahmen sicher (z. B. eine Pseudonymisierung der Daten). Außerdem verwalten sie den Zugriff auf die Daten gemäß dem OAIS-Konzept als Transferstelle. Zur Verpflichtung auf die von der AG Data Sharing entworfene *Nutzungsordnung*<sup>3</sup> und zur Prüfung der Einhaltung der Restriktionen sind alle Daten vor der Datennutzung zu beantragen. Der *Datennutzungsantrag*<sup>4</sup> wurde von der AG spezifiziert und abgestimmt.

Der *Datennutzungsvertrag*<sup>5</sup> wurde von der AG zur schnellen Bearbeitung standardisiert und abgestimmt. **Abb. 1** gibt eine Übersicht über den vereinheitlichten Antragsprozess, der für die Forschenden grafisch aufbereitet wurde.

<sup>3</sup> <https://www.medizinformatik-initiative.de/de/nutzungsordnung>. Zugegriffen am 20.12.2023.

<sup>4</sup> [https://www.medizinformatik-initiative.de/sites/default/files/2020-12/DatNutzAntr\\_v1.0.pdf](https://www.medizinformatik-initiative.de/sites/default/files/2020-12/DatNutzAntr_v1.0.pdf). Zugegriffen am 20.12.2023.

<sup>5</sup> <https://www.medizinformatik-initiative.de/de/nutzungsvertrag>. Zugegriffen am 20.12.2023.

Bundesgesundheitsbl <https://doi.org/10.1007/s00103-024-03880-y>  
© The Author(s) 2024

T. Kirsten · P. Kleinert · M. Gebhardt · J. Drepper · A.-K. Andreeff · F. Prasser · O. Kohlbacher

## Grundlagen für die wissenschaftliche Nutzung umfangreicher Versorgungsdaten in Deutschland – Ergebnisse der AG Data Sharing der Medizinformatik-Initiative

### Zusammenfassung

Versorgungsdaten stellen eine wichtige Ressource in der angewandten medizinischen Forschung dar. Sie liegen multizentrisch vor. Es bleibt jedoch eine Herausforderung, standardisierte Datenaustauschprozesse zwischen Bundesländern und ihren individuellen Gesetzen und Vorschriften zu ermöglichen. Die Medizinformatik-Initiative (MII) wurde im Jahr 2016 gegründet, um Prozesse zu implementieren, die einen klinikübergreifenden Zugriff auf Versorgungsdaten in Deutschland ermöglichen. Mehrere eingerichtete Arbeitsgruppen konzipieren standardisierte Datenstrukturen (AG Interoperabilität), Patienteninformationen und Einwilligungserklärungen (AG Consent) sowie Regelungen zum Datenaustausch (AG Data Sharing). Hier stellen wir die wichtigsten Ergebnisse der Arbeitsgruppe Data Sharing vor, die unter anderem abgestimmte

Nutzungsbedingungen, rechtliche Regelung und Datenzugriffsprozesse enthalten. Diese werden bereits von den etablierten „Datenintegrationszentren“ (DIZ) und „Use and Access Committees“ (UACs) umgesetzt. Wir beschreiben die Dienste, die notwendig sind, um Forschenden einen standardisierten Datenzugang zu ermöglichen. Sie werden u. a. mit dem Forschungsdatenportal für Gesundheit (FDPG) umgesetzt. Auf dieser Basis wurden seit der Pilotphase die Prozesse von 406 aktiven Forschenden verwendet, was zum Stand April 2024 zu 19 registrierten Projekten und 31 eingereichten Forschungsanträgen geführt hat.

### Schlüsselwörter

Medizinformatik · Datenaustausch · Real-World-Daten · Universitätsmedizin · Multizentrische klinische Forschung

## Foundations for the scientific use of extensive health care data in Germany—results of the Data Sharing working group of the Medical Informatics Initiative

### Abstract

Healthcare data are an important resource in applied medical research. They are available multicentrically. However, it remains a challenge to enable standardized data exchange processes between federal states and their individual laws and regulations. The Medical Informatics Initiative (MII) was founded in 2016 to implement processes that enable cross-clinic access to healthcare data in Germany. Several working groups (WGs) have been set up to coordinate standardized data structures (WG Interoperability), patient information and declarations of consent (WG Consent), and regulations on data exchange (WG Data Sharing). Here we present the most important results of the Data Sharing working group, which include agreed terms of use, legal regulations, and data access processes.

They are already being implemented by the established Data Integration Centers (DIZ) and Use and Access Committees (UACs). We describe the services that are necessary to provide researchers with standardized data access. They are implemented with the Research Data Portal for Health, among others. Since the pilot phase, the processes of 385 active researchers have been used on this basis, which, as of April 2024, has resulted in 19 registered projects and 31 submitted research applications.

### Keywords

Medical informatics · Data sharing · Real-world data · University medicine · Multicentric clinical research

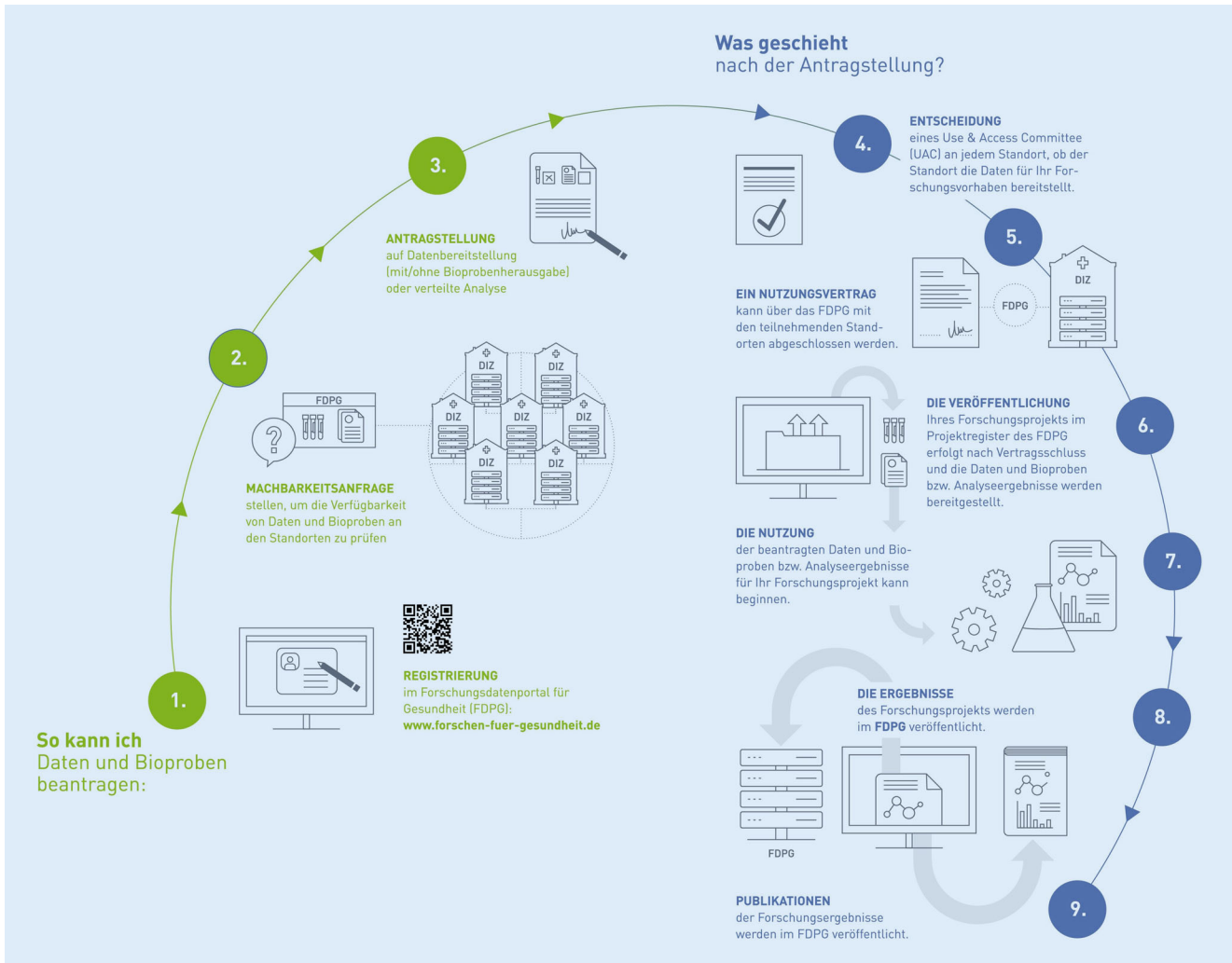


Abb. 1 ▲ Prozess der Datennutzung von der Registrierung und Beantragung bis zur Publikation. (Quelle: TMFe. V.)

### Informationsaufbereitung und Darstellung der verfügbaren Daten für Forschende

Die Forschenden sollten vor der Datenbeantragung genau recherchieren, welche Daten verfügbar sind. Dazu wurde von der AG Data Sharingspezifiziert, welche Einsichtnahme in verfügbare Daten hierfür benötigt wird. Die AG unterscheidet 4 Stufen des Data Sharing, von denen sich die ersten beiden auf Informationen zum Datenmodell und zu den verfügbaren Daten beziehen.

Die erste Stufe beinhaltet ausschließlich Informationen zum Datenmodell. Innerhalb der MII hat man sich darauf geeinigt, alle strukturierten medizinischen Daten nach dem FHIR-Standard (Fast

Healthcare Interoperable Resources)<sup>6</sup> aufzubereiten. FHIR und seine Adaption für die Datenintegrationszentren sind frei zugänglich.<sup>7</sup>

Die zweite Stufe zielt auf die aggregierte Deskription von Datenbeständen. Darunter werden in erster Linie die Häufigkeiten ausgewählter Datenitems verstanden, z. B. die Anzahl von Patientinnen und Patienten oder die Anzahl von Diagnosen, Prozeduren und Laborwerten. Diese deskriptiven Daten werden in einem „Schaufenster“ öffentlich ohne Einschränkung des Zugriffs zur Verfü-

gung gestellt und regelmäßig aktualisiert (Abb. 2).

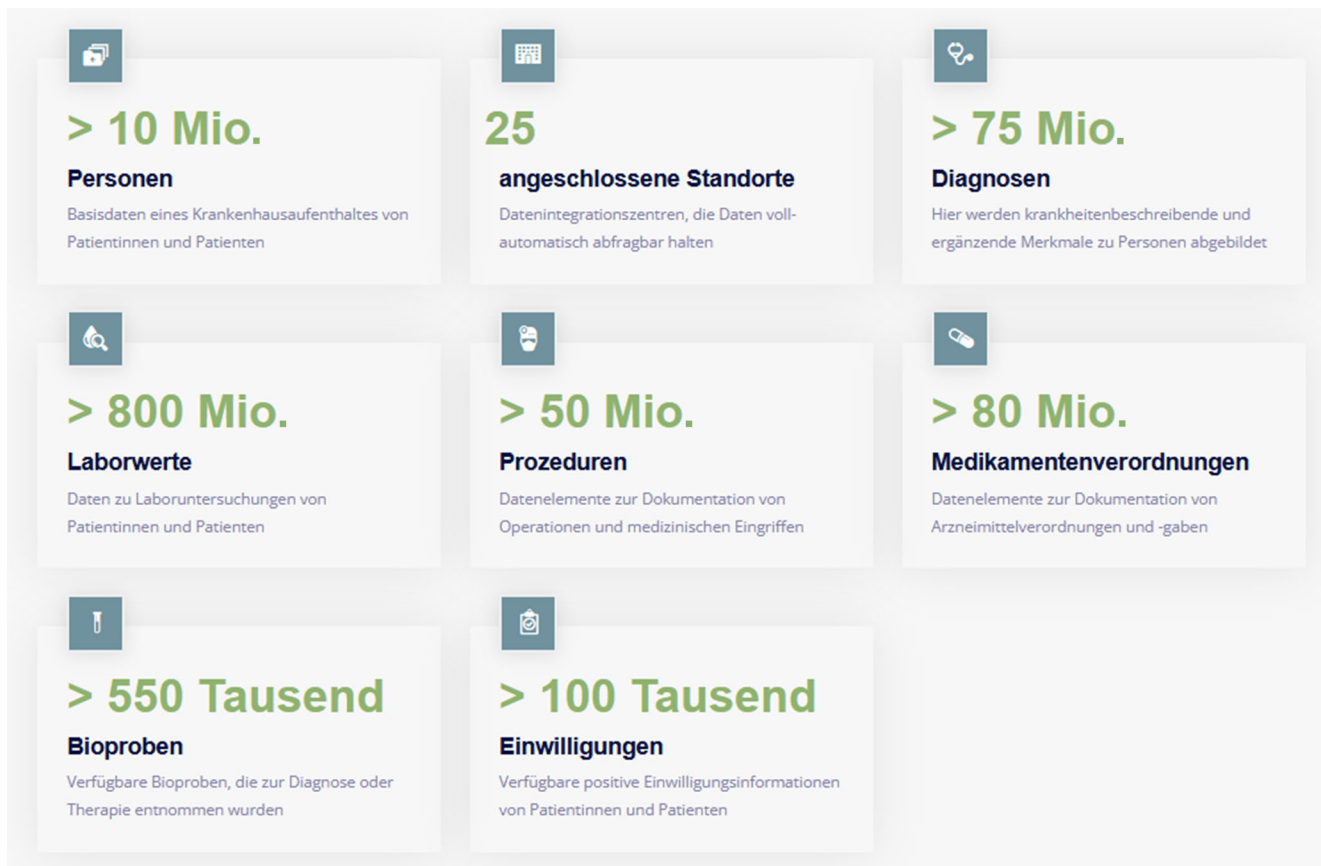
### Machbarkeitsanfragen und Datenbeantragung mit dem Forschungsdatenportal für Gesundheit

Mit der dritten Stufe des Data Sharing sollen Forschende die Möglichkeit haben, konkrete Anfragen zum Datenbestand<sup>8</sup> zu formulieren. Dahinter steht die Frage, ob der verfügbare Datenbestand genügend Patientinnen und Patienten umfasst, um die konkrete Forschungsfragestellung adressieren zu können. Dazu wurde das Forschungsdatenportal für

<sup>6</sup> <https://www.hl7.org/fhir/overview.html>. Zugegriffen am 20.12.2023.

<sup>7</sup> <https://www.medizininformatik-initiative.de/de/der-kerndatensatz-der-medizininformatik-initiative>. Zugegriffen am 20.12.2023.

<sup>8</sup> Der im Folgenden skizzierte Beantragungsprozess für Daten gilt mit kleinen Unterschieden ebenso für Bioproben.



**Abb. 2** ▲ Das „Schaufenster“ des Forschungsdatenportals für Gesundheit (FDPG) zeigt übersichtlich an, welche Kategorien an Daten in welcher Gesamthäufigkeit an den teilnehmenden Universitätskliniken verfügbar sind. (Quelle: forschender-gesundheit.de; zugegriffen am 10.01.2024)

Gesundheit (FDPG) etabliert, in dem Forschende Ein- und Ausschlusskriterien spezifizieren und vollautomatisch Abfragen an angeschlossene Standorte schicken können, um die Machbarkeit der Studie vor Beantragung zu überprüfen. Die Kriterien (z. B. Geschlecht = „weiblich“) können mit logischen Operatoren AND und OR verbunden werden [12].

Für derart ausgeführte Machbarkeitsabfragen wurden restriktive Datenschutzmaßnahmen umgesetzt. Dazu gehören die Limitierung der Anzahl an Abfragen sowie das Verrauschen der ermittelten Häufigkeit von Patientinnen und Patienten am Klinikum. Geringe Häufigkeiten sowohl von Patientinnen und Patienten als auch von Kliniken werden nicht angezeigt, um das Risiko einer Reidentifizierung weiter zu verringern. Die Kliniken werden dabei verschleiert (nur als „Klinikum 1“ usw. gekennzeichnet), sodass kein Rückschluss auf

einzelne Universitätskliniken genommen werden kann. Der Zugang zu den Machbarkeitsabfragen ist nicht öffentlich; er setzt eine Forschungstätigkeit sowie Registrierung voraus.

Im Gegensatz zu den 3 vorgenannten Stufen des Data Sharing, in denen Metadaten ausgetauscht wurden, bezieht sich die vierte Stufe auf das Teilen der medizinischen Daten selbst. Dazu müssen Forschende den Antragsprozess vollständig durchlaufen. Als technische und organisatorische Basis dient das FDPG, in dem Informationen und Prozesse zur Datennutzung gebündelt werden.

Im Antrag enthalten sind die wissenschaftliche Zielstellung, ein Analyseplan, die Machbarkeitsanfrage sowie die zum Erreichen der Ziele notwendigen Variablen. Darüber hinaus erfordert jeder Datennutzungsantrag ein Ethikvotum, das zusammen mit dem eingereichten Datennutzungsantrag an alle angebundene Datenintegrationszentren übermit-

telt wird. An jedem DIZ werden die Datennutzungsanträge von einem Use & Access Committee (UAC) geprüft. Im Ergebnis melden die DIZ die Zustimmung oder die begründete Ablehnung über das FDPG zurück. Rückfragen werden im Portal zentral koordiniert.

Im Anschluss an die Beantragung und Begutachtung wird ein Vertrag zwischen den datengebenden Standorten und der Institution des Antragsstellenden geschlossen. Zeitgleich wird das Projekt im Projektregister<sup>9</sup> des FDPG veröffentlicht. Das Projektregister fungiert als öffentlich zugängliches Transparenzportal; damit können sowohl Patientinnen und Patienten, Forschende als auch die interessierte Öffentlichkeit einen Überblick bekommen, welche Projekte durchgeführt werden oder wurden. Die datengebenden Einrichtungen und die

<sup>9</sup> [https://forschender-gesundheit.de/menu\\_register.php](https://forschender-gesundheit.de/menu_register.php). Zugegriffen am 20.12.2023.

beantragende Person werden aufgeführt. Letztere ist die Ansprechperson für weitere Nachfragen, z. B. in Hinblick auf gemeinsame Forschungsprojekte.

### Vorlagen für die Beantragung von Daten und Bioproben und die Vertragsgestaltung

Die Beantragung von Daten und Bioproben aus verschiedenen Universitätskliniken wird umso einfacher, je mehr die Prozesse und die benötigten Daten harmonisiert und standardisiert sind. Dies schließt ebenso die Bereitstellung der Vorlage eines Datennutzungsvertrags mit allen notwendigen Anlagen ein, die mit allen Justizariaten der an der MII teilnehmenden Universitätskliniken diskutiert und abgestimmt wurde; ein konkreter Datennutzungsvertrag bedarf nur noch der Aufnahme aller Partner und eventuell erteilter Auflagen sowie als Anlage des Datennutzungsantrags. Damit vereinfacht und verkürzt sich der Antragsprozess bis zum Vertragsabschluss, insbesondere wenn viele Partner inkludiert sind. Breit geprüfte Vorlagen für den Nutzungsvertrag sind auf den Webseiten der MII verfügbar.<sup>10</sup>

### Übergreifende Nutzungsordnung als Vorlage für lokale Nutzungs- und Geschäftsordnungen

Grundlage des Handelns der Datenintegrationszentren bezogen auf die lokalen und übergreifenden Datennutzungsprojekte sind vorab verbindlich getroffene Regelungen und Verfahrensweisen. Damit alle Datenintegrationszentren nach denselben Grundsätzen handeln, ist eine Harmonisierung der Regelungen in diesen Geschäftsordnungen notwendig. Dazu wurde eine übergreifende Nutzungsordnung<sup>11</sup> konzipiert und abgestimmt. Diese definiert den grundsätzlichen Antragsprozess (s. oben) und alle Folgeprozesse und benennt die regula-

torischen Grundlagen sowie die Rechte und Pflichten der am Prozess Beteiligten. Darüber hinaus gibt sie Empfehlungen zum Umgang mit Abweichungen (z. B. Ergänzungen in Studienprotokollen). Lokale Nutzungs- oder Geschäftsordnungen können darauf Bezug nehmen und Erweiterungen mit lokalspezifischen Regelungen vornehmen.

Alle Vereinbarungen, die für die Durchführung von Datennutzungsprojekten zwischen den Standorten der MII erforderlich sind, werden im Teilnehmerahmenvertrag der MII getroffen. Alle vorgenannten Bausteine (Datennutzungsantrag, Nutzungsordnung, Standardnutzungsvertrag und Teilnehmerahmenvertrag) ergeben zusammen das MII-Vertragswerk, das gemeinsam mit allen organisatorischen und technischen Komponenten der MII stetig weiterentwickelt und optimiert wird.

### Strukturen zur Durchführung von verteilten Analysen

Aufgrund der föderalen Datenhaltung und Datenschutzbestimmungen wurden innerhalb der MII frühzeitig verschiedene Ansätze und Strukturen zur verteilten Analyse bearbeitet. Im Gegensatz zur zentralen Datenzusammenführung („Bringe die Daten zur Analyse“) verlassen bei der verteilten Analyse die Daten den Standort nicht („Bringe die Analyse zu den Daten“). Damit wird die Souveränität des Datengebenden in Bezug auf seine Daten gewahrt. Es muss gewährleistet werden, dass die Daten an den datengebenden Einrichtungen verbleiben und dort verarbeitet (analysiert) werden.

Neben dem Ausführen von geprüften, öffentlich einsehbaren Skripten zur Analyse der Projektdaten am Standort pilotiert die MII Prozesse und Rahmenbedingungen für verteilte Analysen mit der Plattform DataSHIELD<sup>12</sup>. Dazu wurde begleitend ein Dokumentationskonzept entworfen, das verschiedene Dokumentationen in sich vereint, die zum Betrieb der Verteilten-Analyse-Infrastruktur notwendig sind. Es ist generisch, so dass die aktuelle Dokumentation auf wei-

tere Infrastrukturen übertragen werden kann.

Aufbauend auf den Erfahrungen, die mit den Infrastrukturen zur verteilten Analyse von medizinischen Daten gemacht wurden, zu denen sowohl DataSHIELD als auch verschiedene Implementierungen des Konzepts „Personal Health Train“ (PHT) zählen, wird mit dem Projekt PrivateAIM<sup>13</sup> in der aktuellen Förderphase der MII eine Implementierung entlang der Anforderungen der Datenintegrationszentren vorgenommen. Neben dem grundlegenden Ziel, so datenschutzkonformes maschinelles Lernen und Datenanalytik innerhalb der Plattform der MII zu ermöglichen, möchte das Projekt auch maßgeschneiderte KI-Methoden, Förderationsmechanismen und „Privacy-Enhancing-Technologies“ für die medizinische Datenanalytik der Zukunft entwickeln.

### Qualitätsindikatoren und -prozesse für den Aufbau von Datenintegrationszentren

Letztlich bedarf es einer beständigen Kontrolle und Steuerung, inwieweit die im Aufbau befindlichen Strukturen und Prozesse zu einer Unterstützung des Data Sharing beitragen. Dazu wurden von der TF Audit Indikatoren festgelegt, die sich grob in die Kategorien Struktur-, Prozess- und Ergebnisqualität einteilen lassen. Zu vielen dieser Prozesse wurden Standard Operating Procedures (Standardverfahrensanweisungen, SOP) an den Datenintegrationszentren etabliert, zu denen die Mitarbeitenden geschult und beständig trainiert werden. Indikatoren der Ergebnisqualität richten sich auf die erzielten Ergebnisse eines Datenintegrationszentrums, z. B. die Möglichkeiten des DIZ, die unterschiedlichen Formen des Datenzugriffs (Datenherausgabe vs. verteilte Analyse) zu bedienen. Entlang dieser Indikatoren wurden schon während der ersten Förderperiode Begutachtungen durchgeführt. Eine externe Begutachtung fand zum Ende der ersten Förderperiode durch eine un-

<sup>10</sup> <https://www.medizininformatik-initiative.de/de/nutzungsvertrag>. Zugegriffen am 20.12.2023.

<sup>11</sup> <https://www.medizininformatik-initiative.de/de/nutzungsordnung>. Zugegriffen am 20.12.2023.

<sup>12</sup> <https://www.datashield.org>. Zugegriffen am 20.12.2023.

<sup>13</sup> <https://privateaim.de/>. Zugegriffen am 20.12.2023.

abhängige Institution statt. Im Frühjahr 2021 wurden als Ergebnis des komplexen Auditierungsverfahrens 90% der Standorte als Datenintegrationszentren „im fortgeschrittenen Aufbaustadium“ bewertet.

## Durchgeführte Datennutzungsprojekte

In der Pilotphase wurde eine Reihe von Datennutzungsprojekten in Form von „Projectathons“ initiiert und durchgeführt [13]. Im Bereich der Kardiologie wurden Analysen zu NT-proBNP<sup>14</sup>-Werten bei Vorhofflimmern umgesetzt und die unterschiedlichen technischen Aspekte der MII-Prozesse evaluiert: (1) zentrale Datenzusammenführung, (2) verteilte Analyse mit Skripten und (3) verteilte Analyse mittels DataSHIELD. Die Ergebnisse wurden aufgearbeitet und verbessern die etablierten Prozesse kontinuierlich [13]. Das neurologische Projekt WE-STORM zielt auf die Entwicklung und Validierung statistischer und maschineller Lernmodelle zur Vorhersage von Schlaganfallhäufigkeiten unter Einbeziehung meteorologischer und klinischer Risikofaktoren ab. Im Rahmen des endokrinologischen und diabetologischen Projekts DiaClusT werden Subphänotypen von Typ-2-Diabetes untersucht, um personalisierte Behandlungsansätze zu entwickeln. POLAR-DABI fokussiert auf nephrologische und pharmakologische Fragestellungen. Untersucht wird die Verordnungspraxis des Antikoagulans Dabigatran und seine Auswirkung auf die Nierenfunktion. Unter dem Dach von CORD-MI werden Studien zu Mukoviszidose, Phenylketonurie, zum Kawasaki-Syndrom und PIMS (Pediatric Inflammatory Multi-system Syndrome) sowie zu weiteren ausgewählten Seltenen Erkrankungen bearbeitet und Datenqualitätsprüfungen durchgeführt. Neben diesen standortübergreifenden Projekten wurden auch lokale Datennutzungsprojekte mittels der MII-Strukturen abgestimmt. Beispiele umfassen ein Register für Patienten

mit hepatobiliären Systemtumoren sowie regelmäßige Machbarkeitsanfragen für multizentrische Forschungsprojekte.

Das FDPG steht seit dem 16.05.2023 allen Forschenden zur Verfügung und ist nicht mehr auf MII-interne Projekte eingeschränkt. Stand April 2024 nutzen 405 aktive Forschende das Portal. Mehr als 10.000 Machbarkeitsanfragen wurden durchgeführt und automatisiert von bis zu 25 angeschlossenen Standorten beantwortet. 19 Projekte sind bereits gestartet und wurden im Transparenzregister veröffentlicht. Ein Projekt ist vollständig abgeschlossen. 12 weitere Projekte wurden beantragt und durchlaufen die etablierten Prozesse (vor der Veröffentlichung im Register) aktuell.

## Diskussion und Schlussfolgerungen

Die Medizininformatik-Initiative (MII) ermöglicht Zugriff auf standortübergreifende klinische Routinedaten durch Standardisierung der Datenhaltung und Datenbeantragung. Die AG Data Sharing hat hier entscheidend zu Entwicklung und Aufbau effizienter Prozesse beigetragen, die heute multizentrische Datennutzung erlauben und weiter ausbauen.

Mit dem Aufbau von Datenintegrationszentren an deutschen Universitätskliniken ist ein Meilenstein zur Unterstützung der medizinischen Forschung gelungen. Die Datenintegrationszentren stellen den einheitlichen Zugang zu den Daten in den Universitätskliniken dar. Sie werden die Digitalisierung und den Auf- und Ausbau von interoperablen Informationssystemen in den Universitätskliniken in den nächsten Jahren vorantreiben. Dabei werden sie stets die mit steigender Nutzungszahl immer häufigeren Anfragen auf noch unerschlossene oder nicht digital vorliegende Daten berücksichtigen. Der Fokus der MII auf die Universitätskliniken war für den ersten Schritt von Vorteil; sie sind ausnahmslos Maximalversorger mit einem großen Patientenanteil an Gesundheitsdaten. In nächster Zeit sollen auch kleinere Kliniken und andere Gesundheitsdienstleister sukzessive in die aufgebauten übergrei-

fenden Infrastrukturen zum Data Sharing inkludiert werden.

Die Qualität von Routinedaten aus der medizinischen Versorgung kann mäßig sein, da diese Daten nicht primär zum Zwecke der Forschung erhoben werden. Die standardisierten Prozesse der Datenhaltung (AG Interoperabilität), aber insbesondere auch die der AG Data Sharing tragen dazu bei, die Datenerfassung zu verbessern und damit auch die Qualität der Daten für die Versorgung und Forschung zu erhöhen. Dies ist notwendig, um aussagekräftige evidenzbasierte Forschung (unter anderem auch mit maschinellem Lernen) an den immens umfangreichen Datensätzen der Routineversorgung durchführen zu können.

Im Fokus der standardisierten Prozesse befindet sich stets auch die Verbesserung der Versorgung von Patientinnen und Patienten. Diese profitieren von der Forschung mit Routinedaten nicht nur durch verbesserte Therapiemöglichkeiten, sondern auch durch die verbesserten Prozesse in der Versorgung selbst, die durch die Vereinheitlichung und Standardisierung, wie zum Beispiel der Datenerfassung, ermöglicht werden.

Ungeklärt ist, wie diese Gesundheitsdateninfrastruktur dauerhaft finanziert wird; Betrieb und Wartung von IT-Infrastrukturen sowie organisatorische und vertragliche Administration von Projekten kosten Geld und bedürfen daher einer nachhaltigen Ausfinanzierung.

Obwohl die etablierten Data-Sharing-Prozesse bereits produktiv genutzt werden, bedarf es weiterer Optimierung und Verschlankeung, um kürzere Bearbeitungszeiten bis zur Erteilung des Datenzugriffs für Forschende zu erreichen. Dies inkludiert den weiteren Auf- und Ausbau von Infrastrukturen, die im Kleinen (z. B. elektronische Vertragsunterzeichnung) wie im Großen wirken. Für Letzteres werden in der aktuellen Förderperiode verschiedene Projekte, sogenannte Methodenplattformen (z. B. PrivateAIM, GemTex), gefördert. Dadurch wird die Nutzung der aufgebauten Infrastrukturen weiter vorangetrieben, die Forschungsgemeinschaft und somit die medizinische Forschung werden gestärkt.

<sup>14</sup> NT-proBNP: „N-terminal pro-B-type natriuretic peptide“ (Hormon, das als Herzinsuffizienzmarker dient).

Letztlich bleibt festzuhalten, dass auf der Grundlage der aufgebauten Infrastrukturen und Regelungen ein Zugang zu FAIR-Daten in der medizinischen Forschung möglich ist. Sie sind eine ideale Ausgangsbasis für die Einbindung und Verknüpfung von Routinedaten mit weiteren verfügbaren Datenquellen. Studiendaten, Kassendaten und Registerdaten, wie z. B. Krebs-Registerdaten, sind erweiternde Daten, die die medizinische Forschung auf Routinedaten bereichern. Die Zusammenarbeit mit diesen Dateninfrastrukturen muss stetig verbessert werden. Die etablierten Prozesse der AG Data Sharing ermöglichen dies nun, sodass Deutschland eine gute Ausgangsbasis geschaffen hat, um in den europäischen Datenraum einzutreten.

## Korrespondenzadresse

### Prof. Dr.-Ing. Oliver Kohlbacher

Institut für Biomedizinische Informatik,  
Universität Tübingen  
Sand 14, 72074 Tübingen, Deutschland  
oliver.kohlbacher@uni-tuebingen.de

**Danksagung.** Wir bedanken uns bei allen aktiven und ehemaligen Mitgliedern der AG Data Sharing und der Task-Forces für die lebhaften Diskussionen und die aktive Gremienarbeit. Besonderer Dank gilt Farid Tehrani aus der MII-Koordinationsstelle für die Arbeiten am Vertragswerk sowie den Mitarbeiterinnen und Mitarbeitern an den DIZ für die Umsetzung der Prozesse, die von der AG Data Sharing entwickelt wurden.

**Förderung.** Die Medizininformatik-Initiative und das Netzwerk Universitätsmedizin werden gefördert vom Bundesministerium für Bildung und Forschung (BMBF). Förderkennzeichen der an der Arbeit der AG Data Sharing beteiligten Autorinnen und Autoren der Konsortien und Projekte der MII: *DIFUTURE*: 01ZZ1804D, 01ZZ2304C; *HiGHmed*: 01ZZ1604A, 01ZZ1802X; *MIRACUM*: 01ZZ1608D, 01ZZ1801L, 01ZZ2301B; *SMITH*: 01ZZ2303A, 01ZZ1803A; *Koordinationsstelle der Medizininformatik-Initiative*: 01ZZ2305A, 01ZZ1805, 01ZZ1601; *NUM-DIZ*: 01KX2121.

**Funding.** Open Access funding enabled and organized by Projekt DEAL.

## Einhaltung ethischer Richtlinien

**Interessenkonflikt.** T. Kirsten, P. Kleinert, M. Gebhardt, J. Drepper, A.-K. Andreff, F. Prasser und O. Kohlbacher geben an, dass kein Interessenkonflikt besteht.

Für diesen Beitrag wurden von den Autor/-innen keine Studien an Menschen oder Tieren durchgeführt. Für die aufgeführten Studien gelten die jeweils dort angegebenen ethischen Richtlinien.

**Open Access.** Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

## Literatur

- Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A et al (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3:160018. <https://doi.org/10.1038/sdata.2016.18>
- Kalkman S, van Delden J, Banerjee A, Tyl B, Mostert M, van Thiel G (2022) Patients' and public views and attitudes towards the sharing of health data for research: a narrative review of the empirical evidence. *J Med Ethics* 48:3–13. <https://doi.org/10.1136/medethics-2019-105651>
- Völzke H, Schössow J, Schmidt CO, Jürgens C, Richter A, Werner A et al (2022) Cohort Profile Update: The Study of Health in Pomerania (SHIP). *Int J Epidemiol* 51:e372–83. <https://doi.org/10.1093/ije/dyac034>
- Loeffler M, Engel C, Ahnert P, Alfermann D, Arelin K, Baber R et al (2015) The LIFE-Adult-Study: objectives and design of a population-based cohort study with 10,000 deeply phenotyped adults in Germany. *BMC Public Health* 15:691. <https://doi.org/10.1186/s12889-015-1983-z>
- Celesti A, Fazio M, Romano A, Bramanti A, Bramanti P, Villari M (2018) An OAIS-Based Hospital Information System on the Cloud: Analysis of a NoSQL Column-Oriented Approach. *IEEE J Biomed Health Inform* 22:912–918. <https://doi.org/10.1109/JBHI.2017.2681126>
- Wolfson M, Wallace SE, Masca N, Rowe G, Sheehan NA, Ferretti V et al (2010) DataSHIELD: resolving a conflict in contemporary bioscience—performing a pooled analysis of individual-level data without sharing the data. *Int J Epidemiol* 39:1372–1382. <https://doi.org/10.1093/ije/dyq111>
- Beyan O, Choudhury A, Van Soest J, Kohlbacher O, Zimmermann L, Stenzhorn H et al (2020) Distributed Analytics on Sensitive Medical Data: The Personal Health Train. *Data Intelligence* 2:96–107. [https://doi.org/10.1162/dint\\_a\\_00032](https://doi.org/10.1162/dint_a_00032)
- Welten S, Mou Y, Neumann L, Jaberansary M, Yediel Ucer Y, Kirsten T et al (2022) A Privacy-Preserving Distributed Analytics Platform for Health Care Data. *Methods Inf Med* 61:e1–e11. <https://doi.org/10.1055/s-0041-1740564>
- Ballhausen H, Hinske LC (2023) Federated Secure Computing. *Informatics* 10:83. <https://doi.org/10.3390/informatics10040083>
- Wettstein R, Kussel T, Hund H, Fegeler C, Dugas M, Hamacher K (2022) Secure Multi-Party Computation Based Distributed Feasibility Queries—A HiGHmed Use Case. In: Röhrig R, Grabe N, Hoffmann VS, Hübner U, König J, Sax U et al (Hrsg) *Stud. Health Technol. Inform. IOS Press*, <https://doi.org/10.3233/SHTI220802>
- Wirth FN, Kussel T, Müller A, Hamacher K, Prasser F (2022) EasySMPC: a simple but powerful no-code tool for practical secure multiparty computation. *BMC Bioinform* 23:531. <https://doi.org/10.1186/s12859-022-05044-8>
- Gruendner J, Deppenwiese N, Folz M, Köhler T, Kroll B, Prokosch H-U et al (2022) The Architecture of a Feasibility Query Portal for Distributed COVID-19 Fast Healthcare Interoperability Resources (FHIR) Patient Data Repositories: Design and Implementation Study. *JMIR Med Inform* 10:e36709. <https://doi.org/10.2196/36709>
- Prokosch H-U, Gebhardt M, Gruendner J, Kleinert P, Buckow K, Rosenau L et al (2023) Towards a National Portal for Medical Research Data (FDPG): Vision, Status, and Lessons Learned. In: Hägglund M, Blusi M, Bonacina S, Nilsson L, Cort Madsen I, Pelayo S et al (Hrsg) *Stud. Health Technol. Inform. IOS Press*, <https://doi.org/10.3233/SHTI230124>

**Hinweis des Verlags.** Der Verlag bleibt in Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutsadressen neutral.