

Datenräume für KMU im Gesundheitswesen: Zwischen politischem Anspruch und praktischer Nutzbarkeit

Katharina Dassel, Benedikt Krieger, Maxie Lutze,
Felicitas Muth, Mona Stebner



Datenräume für KMU im Gesundheitswesen: Zwischen politischem Anspruch und praktischer Nutzbarkeit

Für unsere Gesundheitswirtschaft sind Gesundheitsdaten eine zentrale Voraussetzung für Innovation, regulatorische Anschlussfähigkeit und nachhaltige Wettbewerbsfähigkeit. Ob digitale Therapeutika, vernetzte Medizintechnik oder neue Versorgungsmodelle; viele Produktentwicklungen beruhen auf der Verfügbarkeit qualitativ hochwertiger, realitätsnaher und regulatorisch verwertbarer Daten – ebenso wie zahlreiche KI-basierte MedTech-Anwendungen. Daten sind damit für kleine und mittlere Unternehmen (KMU) sowie Start-ups eine kritische Ressource entlang des gesamten Innovationszyklus: von der frühen Entwicklung über Validierung und Zulassung bis hin zu Marktzugang, Erstattung und Skalierung.

Gleichzeitig hat sich die politische und institutionelle Landschaft in den vergangenen Jahren spürbar verändert. Vor der Perspektive des gemeinsamen europäischen Gesundheitsdatenraums (EHDS) entstehen erstmals verbindliche, systematische Strukturen, um den Zugang zu und die Sammlung von Gesundheitsdaten zu ordnen und rechtssicher zu gestalten. Ferner ist mit dem Gesundheitsdatennutzungsgesetz (GDNG) seit 2025 erstmals ein regulatorischer Rahmen geschaffen, der die industrielle Sekundärnutzung von und die Zugangsmöglichkeiten zu Gesundheitsdaten zentral regelt. Diese Entwicklungen markieren einen wichtigen Fortschritt gegenüber der bisherigen Situation fragmentierter, projektbasierter Einzellösungen. Das radikal Innovative dabei ist, dass im Sinne eines bidirektionalen Ökosystems Gesundheitsdatenräume Daten nicht nur Nutzenden zur Verfügung stehen sollen, sondern auch Datenhalter dazu verpflichtet werden, Daten bereitzustellen. Auch wenn KMU nicht im engeren Sinne als Datenhalter zu verstehen sind, so müssen Daten zumindest aus ihren Produkten in den EHDS fließen können: EHDS Art. 33(1) „Data holders shall make the following categories of electronic data available for secondary use“ (k): „electronic health data from medical devices [...]“.

In der Praxis bleibt für KMU vor allem in der Datennutzung eine Lücke: Der Zugang zu Routinedaten aus Versorgung, Registern oder Abrechnungssystemen ist heute zwar grundsätzlich möglich, jedoch noch keine von allen Datenhaltern gelebte und zentral zugängliche Realität. Vielmehr erschweren zahlreiche entkoppelte und kleinteilige Zugangswege die Nutzung, sodass sie zeitlich und

wirtschaftlich schwer planbar ist. Damit stehen die vorhandenen Zugangswege im Konflikt mit der Logik iterativer Produktentwicklung. Es resultiert ein bekanntes Muster: KMU arbeiten zunächst mit kleinen, selektiven Datensätzen aus bilateralen Kooperationen, oft mit dem späteren Kunden als einzigem Datengeber. Robuste Real-World-Evidenz wird erst aufgebaut, wenn Designentscheidungen, Kohortendefinitionen und regulatorische Pfade bereits feststehen. Das erhöht Entwicklungsrisiken, verlängert die Time-to-Market und kann kurz vor der Zulassung zu erheblichen Nacharbeiten führen, insbesondere wenn die späteren Breitdaten eine andere Datenqualität oder Kodierungstiefe aufweisen als die frühen Entwicklungsdaten.

Vor diesem Hintergrund stellt sich die Frage der Gestaltung des Ökosystems, um Gesundheitsdaten für innovative Unternehmen in Forschung und Entwicklung von KI-basierter Medizintechnik realistisch nutzbar zu machen. Dieser Artikel fokussiert die Perspektive von KMU, die Gesundheitsdaten als strategische Ressource für ihre Produktentwicklung benötigen, ebenso wie Fördergeber und Infrastrukturverantwortliche, die die Gestaltung von Datenräumen politisch und finanziell steuern und dabei die Möglichkeiten von KMU berücksichtigen wollen. Die Analyse kombiniert eine Auswertung relevanter politischer, regulatorischer und strategischer Dokumente mit qualitativen leitfadengestützten Gesprächen mit Datennutzenden und Akteuren aus Datenrauminiciativen. Im Zentrum stehen drei Leitfragen: Welche externen Daten sind für KMU im Entwicklungszyklus relevant? Wo liegen zentrale Hürden der Datennutzung in der aktuellen Praxis? Und welche Förder- und Governance-Ansätze sind geeignet, Datenräume von starren Infrastrukturen zu wirksamen Innovationsräumen weiterzuentwickeln?

1. Welche Daten sind für KMU relevant?

Aus Sicht von KMU im Gesundheitswesen bemisst sich die Relevanz von Daten nicht allein an ihrer Menge, sondern an ihrer Nutzbarkeit sowie an ihrer wissenschaftlichen Validierbarkeit und regulatorischen Verwertbarkeit. [Hierfür eignet sich die Betrachtung entlang des Produktentwicklungszyklus](#). Der Entwicklungsprozess verläuft dabei iterativ. Erkenntnisse aus späteren Phasen, insbesondere aus

Validierung und Post-Market-Überwachung, führen häufig zu Anpassungen früherer Design- und Modellentscheidungen und erfordern eine erneute Nutzung und Bewertung von Daten.¹ Im Folgenden wird dies entlang der einzelnen [Entwicklungsphasen von Medizinprodukten](#) konkretisiert:

In der **Konzeptphase** dienen externe Daten der explorativen Analyse und zum Aufstellen von Hypothesen. Dabei wird beispielsweise auf epidemiologische Daten, Register- und Studiendaten sowie Daten aus der klinischen Praxis zurückgegriffen. KMU betreiben aufgrund der hohen Kosten selten eigenständige langfristige Grundlagenforschung, sondern greifen häufig auf [Drittmittel-finanzierte Kooperationen mit Hochschulen oder Forschungseinrichtungen](#) zurück. In der Forschung und Entwicklung zu KI-basierter Medizintechnik werden in der Konzeptphase weniger neue medizinische Theorien formuliert als vielmehr explorative Datenanalysen, Feature-Identifikationen und die Entwicklung erster algorithmischer Ansätze durchgeführt. Versorgungsdaten sind dabei methodisch besonders relevant, weil sie es ermöglichen, die Häufigkeit des Zielphänomens in der realen Versorgung (Prävalenz) abzuschätzen, die tatsächliche Zielpopulation zu definieren und zu prüfen, welche Datenpunkte (Features) in klinischen Systemen verlässlich und vollständig vorliegen. Ein KMU, das diese drei Fragen nicht auf Basis realer Versorgungsdaten beantwortet, trifft die grundlegenden Designentscheidungen seines Produkts ohne empirische Grundlage.

Im **Design und Prototyping** werden Anforderungen an das spätere Produkt identifiziert, technische Machbarkeitsanalysen durchgeführt und Trainingsdatensätze zur Modellbildung aufgebaut. Hier stehen klinische Routinedaten aus elektronischen Patientenakten sowie Krankenhausinformationssystemen (KIS) und Praxisverwaltungssystemen (PVS) im Vordergrund. Je nach Anwendungsfall umfasst dies Diagnosen, Laborwerte und klinische Prozeduren. Anhand dieser im Regelfall KMU-externen Daten können außerdem initiale Modelle trainiert und Fehlannahmen erkannt werden. In dieser Phase sind neben der Datenerhebung insbesondere Datenaufbereitung, Annotation, Qualitätskontrolle und die Trennung in Trainings-, Validierungs- und Testdatensätze entscheidend. Zudem muss die Repräsentativität der Daten hinsichtlich Zielpopulation und Anwendungsszenario überprüft werden. Relevant ist dafür unter anderem, dass KMU nicht ausschließlich mit Daten eines einzigen Versorgungsstand-

orts arbeiten, weil dies durch mangelnde Repräsentativität zu Overfitting und Biases führen kann. Ebenfalls ist die Sicherstellung der Interoperabilität der eigenen Daten unerlässlich, da Datensätze von Drittparteien die eigenen Datensätze nur dann sinnvoll ergänzen, wenn gemeinsame Standards genutzt werden – sowohl syntaktisch (zum Beispiel FHIR, HL7) als auch semantisch (SNOMED, LOINC).

In der Phase der **Verifikation und Validierung** werden primär interne Entwicklungsdatensätze sowie unabhängige Validierungsdatensätze genutzt. Deren Unabhängigkeit ist besonders wichtig, damit algorithmische Biases, zum Beispiel durch Daten-Standort-spezifische Artefakte, erkannt und behoben werden können. Bei KI-basierten Medizinprodukten umfasst dies retrospektive Datensätze, multi-zentrische externe Validierungskohorten sowie, sofern erforderlich, Daten aus prospektiven klinischen Machbarkeitsstudien. Während dieser Phase werden auch Registerdaten (zum Beispiel Krankheits-, Implantat- und Qualitätsregister) genutzt, um die eigene Modellleistung im Kontext von Informationen über Krankheitsverläufe, Therapieeffekte, Subgruppenunterschiede und langfristige Ergebnisse zu testen. Für KMU sind diese Daten essenziell, um Modelle zu kalibrieren, Bias zu erkennen und die Güte der Modelle für real existierende Heterogenität zu bewerten. Dabei werden für die eigentliche Leistungsvalidierung kuratierte klinische Datensätze mit validierten Referenzstandards² genutzt und Routinedaten ergänzend zur Generalisierbarkeitsprüfung herangezogen. Die Berücksichtigung heterogener und diverser Datensätze ist sowohl methodisch erforderlich, um robuste und generalisierbare Modelle zu entwickeln, als auch regulatorisch und ethisch geboten.

Für die **Zulassung und Zertifizierung** gemäß Medical Device Regulation ([MDR](#)) ist eine klinische Bewertung verpflichtend³. Bei KI-basierten Medizinprodukten kann die Leistungsbewertung auf retrospektiven Datensätzen beruhen. Die dafür genutzten Daten dienen mehreren Zwecken: klinische Validierung und Leistungsbewertung, Erstellung der regulatorischen Dokumentation (MDR, In Vitro Diagnostic Regulation [IVDR](#)), Nachweis von Sicherheit, Leistungsfähigkeit und klinischem Nutzen sowie Vorbereitung des Health Technology Assessments ([HTA](#)). Auch hier sind Register- und Versorgungsdaten relevant, um die eigene Leistungsbewertung sowie eine Abschätzung von Zusatznutzen und Kosteneffektivität durchführen zu können.

1 Siehe [WHO, 2025](#) als Einschätzung des Status-Quo der Datennutzung im Gesundheitswesen. Ein benannter Mangel ist u. a. die Zulassung von Medizinprodukten mit begrenzter klinischer Evidenz.

2 Der Referenzstandard (Ground Truth) ist die definierte Zielantwort, die das Modell lernen soll, zum Beispiel die von zwei unabhängigen Experten konsentrierte Diagnose oder ein histologisch bestätigter Befund. Regulatoren prüfen im Zulassungsverfahren explizit, nach welchem Verfahren dieser Referenzstandard erstellt wurde (IMDRF, 2025)

3 KMU, die KI-basierte Medizinprodukte entwickeln, müssen ab August 2026 zusätzlich die Anforderungen des EU AI Acts erfüllen, sofern ihr Produkt als Hochrisiko-KI eingestuft wird. Dies umfasst Dokumentationspflichten zur Datenherkunft, Nachweise zur Bias-Kontrolle sowie Anforderungen an Umfang und Qualität der Trainingsdaten. Diese Pflichten überlagern sich mit der MDR, sind aber nicht deckungsgleich.

Für Letzteres sind insbesondere GKV-Abrechnungsdaten geeignet, wie sie seit Oktober 2025 über das Forschungsdatenzentrum Gesundheit zugänglich sind. Dabei hängen die konkreten Anforderungen an die Daten wesentlich von der Risikoklasse des Medizinprodukts ab. Zunächst gilt, dass die [ausschließliche Nutzung retrospektiver Studien nur bei Verfügbarkeit adäquater Daten in ausreichender Menge und Qualität möglich ist](#). Wenn das Produkt den zukünftigen Zustand eines Patienten vorhersagt, klinische Ergebnisse beeinflusst oder Patientenmanagement-Entscheidungen begründet, können zusätzlich prospektive Leistungsstudien erforderlich sein, was eine aufwendige Erhebung eigener Daten bedeutet, – ein Aspekt, den KMU frühzeitig klären müssen. Für Produkte der Klasse IIb und III (mittleres bis hohes respektive höchstes Risiko) schreibt die MDR grundsätzlich vor, dass die klinische Bewertung Daten aus einer klinischen Prüfung enthalten muss. Retrospektive Daten allein reichen hier in der Regel nicht aus; Ausnahmen sind nur unter engen Bedingungen möglich (Artikel 61 MDR). Je stärker ein KI-System Behandlungsentscheidungen beeinflusst, desto wahrscheinlicher wird eine prospektive Leistungsstudie verpflichtend; mit entsprechendem Zeit- und Kostenaufwand. Über die Zulassung hinaus spielen in dieser Phase insbesondere Versorgungs- und Abrechnungsdaten auch im Kontext von Erstattungs- und Preisentscheidungen eine wichtige Rolle, da sie für die Bewertung von Zusatznutzen und Kosteneffektivität herangezogen werden.

In der **Post-Market-Überwachung** müssen Hersteller gemäß MDR die Produktleistung in der realen Versorgung systematisch überwachen. Dies umfasst die kontinuierliche Sammlung, Auswertung und Bewertung von Leistungs- und Sicherheitsdaten. Bei KI-Systemen kann ein Post-Market Clinical Follow-up (PMCF) erforderlich sein, um frühzeitig Model Drift, Data Drift und Leistungsverlechterungen in spezifischen Subgruppen zu erkennen. Während dieser Phase werden neben den bereits aufgelisteten Formen der Real-World-Daten (RWD) auch Nutzungs- und Interaktionsdaten, Systemprotokolle sowie Fehlermeldungen ausgewertet. Sie stammen aus der Versorgung, nationalen Registern und Qualitätsdatensätzen sowie herstellereigenen Vigilanzsystemen und ermöglichen die Identifikation technischer Fehlfunktionen, atypischer Anwendungsszenarien und potenzieller Bias-Entwicklungen. Ferner werden sie genutzt, um Real-World-Evidence zur Produktverbesserung zu sammeln und Zertifikatserneuerungen vorzubereiten.

Es wird deutlich, dass KMU entlang des gesamten Produktentwicklungszyklus auf strukturierten, interoperablen und rechtssicheren Zugang zu Gesundheitsdaten angewiesen sind. Insbesondere für explorative Analysen, externe Validierung und Post-Market-Überwachung sind föderierte oder zentral organisierte Gesundheitsdatenräume poten-

ziell ein zentraler Enabler für KMU, da sie skalierbaren und standardisierten Zugang zu repräsentativen Daten aus der Versorgung ermöglichen.

2. Datenräume heute: Warum bestehende Modelle für digital innovative KMU nur begrenzt funktionieren

Dass von den KMU im Gesundheitswesen trotz wachsender Dateninfrastrukturen [noch ein deutlicher Anteil nicht mit realen Daten aus der Versorgung arbeitet](#), ist kein individuelles Versäumnis. Um den zugrundeliegenden strukturellen Hürden nachhaltig entgegenzuwirken und Gesundheitsdaten für die Akteure des Innovationsökosystems nutzbar zu machen, sind in den vergangenen Jahren in Deutschland und Europa vielfältige Dateninfrastrukturen entstanden. Dabei lassen sich grob drei Typen unterscheiden, die jeweils unterschiedliche Stärken und aus KMU-Sicht unterschiedliche Grenzen der Nutzbarkeit aufweisen: nationale öffentlich-rechtlich gerahmte Zugänge zu Routinedaten (zum Beispiel über nationale Datenhalter und Datenzugangsstellen wie das Forschungsdatenzentrum Gesundheit (FDZ)), wissenschaftsgetriebene Integrationsinfrastrukturen (zum Beispiel Datenintegrationszentren (DIZ), Forschungsdatenportal für Gesundheit (FDPG) und Forschungsverbünde) sowie industrie- bzw. servicegetriebene Datenökosysteme und Datenrauminiziativen (wie zum Beispiel sphin-X, Honic, Gaia-X-Health). Übergeordnet entsteht in den nächsten Jahren mit dem EHDS ein europäischer Ordnungsrahmen, der Sekundärnutzung über definierte Zugangsstellen, Genehmigungsprozesse und sichere Verarbeitungsumgebungen operationalisiert. Ein zentrales Antragsverfahren soll dabei sicherstellen, dass KMU nicht für jeden Datenzugang mehrfach die gleichen Prozesse durchlaufen müssen; es bleibt aber ein formaler Genehmigungspfad mit entsprechendem Aufwand für Registrierung und Antragstellung (siehe Whitepaper von [SRD Rechtsanwälte](#)). Aus Sicht von Medizintechnikunternehmen ist der Nutzen dieser drei Arten von Datenzugängen insbesondere daran zu messen, ob sie entlang des Produktentwicklungszyklus Mehrwert schaffen. Genau hier zeigt sich jedoch, dass formale Zugänglichkeit nicht automatisch praktische Nutzbarkeit bedeutet und die Vielfältigkeit in Strukturen und Governance Modellen momentan Hürden darstellen.

Öffentlich-rechtliche Zugänge (insbesondere FDZ Gesundheit) können prinzipiell hochrelevant sein, weil sie populationsnahe Versorgungs- und Abrechnungsdaten bereitstellen. Gleichzeitig stoßen KMU hier besonders häufig auf operative Engpässe: Der Zugang hängt von Antragsprozessen (Usability, Bearbeitungszeit) und Zugangsbeschränkungen ([10 Analysetage innerhalb eines 30 Tagesfensters, Skriptentwicklung in der Regel auf synthetisierten Daten-](#)

sätzen, nur in eng begründeten Ausnahmefällen RWD, pro Antrag eine Datennutzung auf einem Datensatz) ab, die nicht auf iterative, reproduzierbare Produktentwicklung ausgelegt sind. Zudem ersetzt ein zentraler Koordinationsmechanismus zwar Einzelverträge mit einzelnen Datenhaltern, der Mehrwert der Daten ist aufgrund der aktuell geringen klinischen Datentiefe sowie der noch im Aufbau befindlichen Datenbreite (zum Beispiel Anbindung an elektronische Patientenakte (ePA) ab voraussichtlich Q4 2026) bisher aber eingeschränkt.⁴ Dadurch sind bestimmte Formen der Produktentwicklung, insbesondere iteratives KI-Training auf klinisch tiefen Daten, über das FDZ bisher nicht realisierbar. In den Gesprächen wird deutlich, dass in der Konzeptionsphase oder zur Validierung von Versorgungseffekten sich die Daten gut eignen. Als Ableitung der Versorgungseffekte kann das zum Beispiel auch in der Verhandlung von Vergütungen innovativer Produkte helfen. Positiv zu bewerten ist die hohe Transparenz der verfügbaren Daten mit einem Datenmodell, Public Use File sowie einem Statistikportal, die zur Antragsvorbereitung genutzt werden können. Zudem werden entwickelte Skripte auf pseudonymisierten Echtdaten getestet, sodass repräsentative statistische Ergebnisse resultieren. Das grundsätzliche Interesse der Forschungsgemeinschaft zeigt sich in den mittlerweile [82 Institutionen, die am FDZ Gesundheit registriert sind, den 53 eingegangenen Anträgen sowie 18 Forschungsteams, die aktuell gleichzeitig die Analyseräume des FDZ nutzen](#) (Stand: 19.3.2026).

Wissenschaftsgetriebene Integrationsinfrastrukturen (zum Beispiel DIZ, Forschungsdatenportal, Forschungsverbünde) sind häufig am stärksten in der Datenharmonisierung, methodischer Qualität sowie klinischer Tiefe der Daten. Sie liefern klinisch tiefe Rohdaten wie Laborwerte, Bilddaten, Freitexte, Genomik und longitudinale Verläufe und sind damit aktuell der zugänglichste Infrastrukturfad in Deutschland, der die für KI-Modellentwicklung und -validierung erforderliche Datenqualität bereitstellt. Sie sind darüber hinaus prädestiniert für Kohortenbildung und – je nach Zugang – auch für multizentrische Analysen, die für Generalisierbarkeit und Bias-Prüfungen in der Produktentwicklung sowie die klinische Validierung wichtig sind. Im Gegensatz zu FDZ-Daten ermöglichen DIZ-Kooperationen zudem iterative Datenzugriffe, die für agile Entwicklungszyklen unerlässlich sind. Für KMU und innovative Unternehmen ist hierbei allerdings problematisch, dass mit jedem Datenhalter eigene Verträge und Zugangsmöglich-

keiten verhandelt werden müssen. In der Praxis arbeiten KMU entsprechend nur mit einzelnen Partnern zusammen. Das führt strukturell zu einem Single-Site-Problem. Denn Modelle, die auf Daten eines einzigen Standorts entwickelt werden, unterschätzen die reale klinische Variabilität, da sie zum Beispiel unterschiedliche Geräte, Kodiergewohnheiten und Patientenpopulationen nicht abbilden. Dies schränkt sowohl die Generalisierbarkeit als auch die regulatorische Validierbarkeit ein. Diese Beobachtung wird von einer [Umfrage von Bayern Innovativ](#) gestützt, in der deutlich mehr als die Hälfte der Unternehmen angibt, Partnerschaften mit Kliniken und Praxen als wichtigste Datenquelle zu nutzen. Während einzelne KMU mit etablierten Klinikpartnerschaften kurzfristig profitieren, bleibt das Gesamtsystem langsam, weil Erkenntnisse aus bilateralen Kooperationen häufig proprietär bleiben und nicht ins breitere Ökosystem zurückfließen. Neue Marktteilnehmer ohne bestehende Partnerschaften sind strukturell benachteiligt. Eine Verbesserung verspricht hier das Forschungsdatenportal für Gesundheit (FDPG), was die Nutzung von Daten aus über 40 DIZ zentralisiert und föderierte Anfragen an über 40 DIZ-Standorte ermöglicht. Damit wird das Single-Site-Problem teilweise adressiert. Gleichzeitig ist der Zugang jedoch nur eingeschränkt mit der iterativen Entwicklungslogik von KMU kompatibel. Genehmigungsprozesse können mehrere Monate dauern und die Entscheidung über die Bereitstellung der Daten liegt weiterhin bei den einzelnen DIZ. Damit ist auch über das FDPG kein iterativer Datenzugang möglich, die Prozesslogik bleibt auf einmalige, vorab definierte Anfragen ausgelegt. Zur Verbesserung der rechtssicheren Nutzung der Daten durch industrielle Partner ist für die Zukunft eine Partnerschaft mit sphin-X geplant ([LinkedIn](#), 23.3.2026). Zudem wird die Usability weiterentwickelt, unter anderem durch die Möglichkeit [Machbarkeitsanfragen](#) zu stellen, um die Verfügbarkeit relevanter Daten zur Beantwortung der Forschungsfragen zu prüfen (siehe [SiBB Health-IT Talk „MII Update: 10 Jahre“ vom 13.4.2026](#))

Industrie- und servicegetriebene Datenökosysteme und Datenrauminiziativen adressieren gezielt praktische Probleme der Datennutzung – etwa Standardisierung, Datenqualität, sichere Analyseumgebungen und Vertrauen sowie Interoperabilität und Regelungen zu IP. Beispiele sind ein Fraunhofer-naher Datenraumansatz ([Fraunhofer Medical Data Space](#)), eine Health-Instanz von [Gaia-X](#) oder servicebasierte Plattformen wie [Honic](#) oder [sphin-X](#). Auf dem Weg zum EHDS können solche Umgebungen als „Ver-

4 Perspektivisch sollen die FDZ-Daten mit weiteren Datenquellen verknüpft werden. Gesetzlich bereits angelegt ist die Verknüpfung mit Krebsregisterdaten über das GDNG. Der Datensatz des Zentrums für Krebsregisterdaten (ZfKD) umfasst seit 2023 auch klinische Verlaufsdaten zu Therapien und Überlebensverläufen und ist über das ZfKD heute bereits separat auf Antrag für onkologische Versorgungsforschung und Therapiemusteranalysen zugänglich, jedoch nicht für bildbasierte KI. Die NAKO Gesundheitsstudie mit über 200.000 Teilnehmenden – mit tiefer Phänotypisierung, Bildgebung und Biomaterialien – steht als eigenständige Infrastruktur daneben: Seit September 2025 ist über eine Trusted Research Environment auch ein formaler Unternehmenszugang möglich, allerdings mit der Erfordernis eines öffentlichen Interesses verbunden und ohne iterativen Datenzugang.

trauensräume“ wirken, in denen Unternehmen das Datenteilen und gemeinsame Standards praktisch erproben (zum Beispiel bei [sphin-X](#)). Gleichzeitig sind diese Modelle in der Regel auf selbst aufgebaute Netzwerke von Datengebern oder auf Fördermittel angewiesen, mit dem Ziel sich langfristig selbst tragen zu können. Der Zugang zu solchen Plattformen ist grundsätzlich offen. Das zentrale Problem liegt daher weniger im Datenzugang selbst als in der strukturellen Datenknappheit: Datenbestände sind auf die datenhaltenden Institutionen begrenzt und selektiv. Entsprechend ist der initiale Aufwand für solche Datenräume groß, ein attraktiv großes Netzwerk aufzubauen, damit der Aufwand für einzelne datennutzende KMU sinkt. Während im Vergleich zu den heutigen Daten im FDZ die Daten in solchen Initiativen eine höhere klinische Tiefe aufweisen, ist unklar, wie repräsentativ sie sind. Für bevölkerungsweite Fragestellungen oder breit angelegte Subgruppenanalysen sind sie daher weniger geeignet, für fokussierte klinische Validierung und produktnahe Evidenz hingegen ein realistischer Weg als öffentlich-rechtliche Antragspfade. Aus einer Usability-Perspektive sind sie an einer [schnelleren Datenbereitstellung](#) sowie [geeigneten Datenaufbereitung und produktnahen Services](#) bestrebt. Ihr Mehrwert kann damit nicht nur in der Aufbereitung und Standardisierung, sondern insbesondere in der frühen Produktentwicklungsphase liegen: schneller Datenzugang, transparenter Antragsprozess, aufbereitete und annotierte Datensätze sowie produktnahe Services, die iterative Entwicklungszyklen unterstützen. Für KMU, die klinisch tiefe Daten für einen spezifischen Indikationsbereich suchen und bereit sind, aktiv Datengeber einzubinden, können diese Ökosysteme ein strategisch sinnvoller erster Schritt sein, solange die Einschränkungen hinsichtlich Repräsentativität und Skalierbarkeit in der regulatorischen und klinischen Strategie berücksichtigt werden.

Übergeordnet bestehen für KMU bei allen Arten von Datenräumen zudem vier weitere zentrale Hürden:

Die strukturellen Hürden beim Zugang zu sekundären Gesundheitsdaten unterscheiden sich je nach Infrastrukturtyp grundlegend und sind Ausdruck einer **fragmentierten Datenlandschaft**: Beim FDZ Gesundheit liegt das Hemmnis in der Prozesslogik, bei den DIZ fehlen Standardverträge und skalierbare Mehrpartnerlösungen, bei industrie-/servicegetriebenen Ökosystemen entsteht eine strukturelle Datenknappheit durch Freiwilligkeit der Datenbereitstellung. Gemeinsam ist allen drei Typen, dass klinische Routinedaten, Registerdaten und Abrechnungsdaten unterschiedlichen Datenhaltern, Zweckbindungen und Governance-Strukturen unterliegen. Mit dem im Aufbau befindlichen FDZ Gesundheit entsteht zwar nun ein formaler Koordinationsmechanismus, der Antragstellung und Vertragsrahmen künftig auch insbesondere über den EHDS (über die Datenzugangs- und Koordinie-

rungsstelle, DACO beim Bundesamt für Arzneimittel und Medizinprodukte) bündeln soll. Für KMU ersetzt das FDZ jedoch keine Kooperationen mit Leistungserbringern – nicht nur wegen des aktuellen Datenbestands, sondern strukturell: Iteratives KI-Training, Bildgebungsdaten und Freitextanalyse bleiben mit bisher geplanter Ausgestaltung außerhalb des FDZ-Pfades. Fraglich ist zudem, inwiefern sich die verschiedenen Datenräume künftig in den EHDS integrieren. Trotz insgesamt hoher Datenverfügbarkeit, erfordert eine einzelne Sekundärdatennutzung derzeit häufig viele Einzelvereinbarungen und Anträge. Separate Freigaben, Datenschutzprüfungen und gegebenenfalls Ethikvoten bleiben bestehen. Es entsteht zwar ein Single Point of Contact, aber häufig kein durchgängiger Entscheidungsprozess – insbesondere bei der Kombination mehrerer Datentypen. Dieses Spannungsverhältnis ist strukturell bedingt: Die Datenhalter behalten ihre Freigabekompetenz aus datenschutz- und haftungsrechtlichen Gründen; sodass zentrale Koordination und dezentrale Verantwortung dauerhaft nebeneinander bestehen.

Ein zweites strukturelles Problem sind **lange Verfahren**, die abhängig vom Datentyp und den beteiligten Institutionen die Dauer zwischen erster Datenanfrage und tatsächlicher Datennutzung unvorhersehbar in die Länge ziehen können: Ein FDZ-Antrag durchläuft ein mehrstufiges Prüfverfahren mit langen maximal möglichen Bearbeitungszeiten und vorher notwendiger Institutionsregistrierung. Eine bilaterale DIZ-Kooperation kann bei bereits bestehender Partnerschaft schneller sein. Industrie-/servicegetriebene Plattformen versprechen zeitlich stärker begrenzte Bearbeitungszeiten – allerdings mit den beschriebenen Einschränkungen bei Datenvielfalt und Repräsentativität. Für KMU ist diese Varianz strategisch relevant, weil sie die Wahl der Datenquelle nicht nur nach Datenqualität, sondern auch nach verfügbarem Zeitbudget treffen müssen. Verbindliche Bearbeitungsfristen, garantierte Durchlaufzeiten oder standardisierte Kombinationsszenarien sind bislang die Ausnahme. Für KMU mit begrenztem finanziellem Runway und iterativen Entwicklungszyklen sind diese Zeitachsen schwer vereinbar mit produktgetriebener Innovation. Vielfach müssen Entwicklungsentscheidungen getroffen werden, bevor reale Versorgungsdaten verfügbar sind – mit entsprechendem Risiko späterer Fehlinvestitionen.

Drittens dürfte aktuell auch die fehlende **Planbarkeit der Kosten** KMU eher abschrecken. Trotz Gebührenordnungen für einzelne Datenzugriffe (zum Beispiel [beim FDZ Gesundheit](#)) lassen sich die Kosten kaum vorab abschätzen, da nicht klar ist, wie oft Daten abgefragt werden müssen bis hin zur Frage ob gegebenenfalls ergänzende Kooperationsverträge geschlossen werden müssen. Hinzu kommen Transaktionskosten wie juristische Beratung, Antragsanpassungen, Datenschutz- und Ethikprüfungen,

technische Integration und Datenharmonisierung. Dabei besteht eine strukturelle Asymmetrie: Große Unternehmen können diese Kosten als Overhead absorbieren, für KMU und Start-ups wirken dieselben Prozesse aufgrund ihrer relativen Belastung oft prohibitiv. Ergänzend wirken hohe Sicherheitsanforderungen – insbesondere dort, wo im EHDS-Kontext sichere Verarbeitungsumgebungen (Secure Processing Environments) zum Standardpfad werden – als weitere Zugangshürde.⁵

Ein letzter limitierender Faktor ist die fehlende **Lifecycle-Governance**: Heutige Datenzugänge sind strukturell auf einmalige, projektförmige Nutzung ausgelegt. Dabei braucht ein KMU, das ein KI-Medizinprodukt entwickelt, Datenzugang über den gesamten Produktlebenszyklus: für Training, Validierung, Post-Market Clinical Follow-up, Re-Training bei Modell-Drift und regulatorisch geforderte Nachbeobachtung. Kein heutiger Zugangspfad – weder FDZ noch DIZ-Kooperation noch industrie-/servicegetriebenes Ökosystem – scheint für diese Anforderung konzipiert. Dauerhafte Nutzungsrechte, Service Level Agreements und den entsprechenden Regelungen zur Ergebnisverwertung fehlen weitgehend. Wie [dieses Beispiel](#) zeigt sind es vielmehr Einzelvereinbarungen, die auf Basis von Standards getroffen werden müssen. Neben diesen strukturellen Hürden spielen rechtlich-regulatorische Unsicherheiten eine zentrale Rolle, etwa divergierende Auslegungen der Datenschutzgrundverordnung auf Bundeslandebene. Das GDNG hat mit der Einführung einer federführenden Datenschutzaufsicht bei länderübergreifenden Forschungsvorhaben einen ersten Schritt zur Reduktion dieser Fragmentierung unternommen. Gelöst ist das Problem damit aber noch nicht, sondern lediglich für einen spezifischen Anwendungsfall strukturiert worden.

3. Ansätze für eine Stärkung der Gesundheitsdaten-Landschaft in Deutschland

Interoperabilität

Das Fundament sekundärer Datennutzung sind interoperable Daten. Wie bereits oben ausgeführt, reicht dies bis zur semantischen Interoperabilität. Entsprechend wichtig ist es, dass zum Beispiel ePA-Daten nach Medizinischen Informationsobjekten (MIOs) strukturiert vorliegen. Neben regulatorischen Bemühungen sind besonders Medizintechnik- und Digital-Health-Unternehmen gefordert, ihre Systeme interoperabel zu gestalten. In der Praxis gestaltet sich Interoperabilität weniger als technisches Inno-

vationsproblem, sondern vielmehr als Umsetzungs- und Betriebsproblem: Die Implementierung interoperabler Schnittstellen – insbesondere in Primärsystemen Krankenhausinformationssystemen (KIS) oder Praxisverwaltungssystemen (PVS) – ist mit erheblichen Aufwänden und Kosten verbunden. Neben Medizintechnik-Herstellern sind daher auch Leistungserbringer gefordert, Interoperabilität in der Praxis tatsächlich umsetzbar zu machen. Herr Kirchhoff von medicalvalues unterstreicht dies aus Sicht eines KMU mit medizinischen KI-Produkten: „Ohne Daten aus einem KIS oder LIS⁶ hätten wir unser Produkt nicht entwickeln können und könnten unseren Service nicht anbieten. Für die Schnittstellen müssen wir teilweise mehrere Tausende Euro an die Hersteller bezahlen.“

Konsolidieren bestehender Datenquellen durch zentrale Zugriffsmöglichkeiten

Die Übersicht der Datenraumbenutzungsinitiativen zeigt, dass es bereits viele Möglichkeiten zur Datennutzung gibt. Ein echter Mehrwert im Sinne einer innovationsfreundlichen und gleichzeitig Datenschutz-wahrenden Sekundärdatennutzung würde jedoch erst durch die Kombination aus einem öffentlich zugänglichen Metadatenkatalog und einem durchgängigen, standardisierten Zugangsprozess entstehen – idealerweise ergänzt durch verbindlichen Bearbeitungsfristen für Datenanfragen. Auf dem Weg zu mehr Transparenz über verfügbare Daten stellt der [EU Gesundheitsdatenkatalog \(HealthDCAT-AP, entwickelt im Rahmen des HealthData@EU-Pilotprojekts\)](#) einen ersten wichtigen Schritt der Mitgliedsstaaten dar. Die zuverlässige, responsive und einheitliche Zugangsstruktur, müsste dann die Daten nicht alle selbst halten, sondern sollte vielmehr bestehende Strukturen (zum Beispiel DIZ, ambulante Versorgung, Pflegeeinrichtungen) integrieren und Datenanfragen orchestrieren, so Herr Kirchhoff. Auch wenn die datenschutzkonforme Datenfreigabe weiterhin bei den jeweiligen Datenhaltern verbleibt, könnte eine zentrale Governance den Zugang für KMU handhabbarer machen. Diese Perspektive wird auch von Herr Dr. Müller, Lead Data Innovation Solution & Collaboration der Bayer Vital GmbH geteilt, der zentral orchestrierte Zugänge als Voraussetzung für die Erschließung heterogener Datenquellen sieht, ein Ansatz der unter anderem auch zu den Zielen von sphin-X zählt. Mit dem Europäischen Gesundheitsdatenraum (EHDS) sollen entsprechende Strukturen in Form sogenannter Health Data Access Bodies (HDAB) etabliert werden, die als nationale Zugangsstellen für die Sekundärdatennutzung fungieren und den Zugang zu Gesundheitsdaten koordinieren. Die nutzerfreundliche Ausgestal-

5 Secure Processing Environments bedeuten, dass KMU nicht ihre eigene Analyseinfrastruktur nutzen können, sondern in einer vorgegebenen Umgebung beim Datentreuhänder mit eingeschränkter Toolauswahl arbeiten. Das erhöht den Einarbeitungsaufwand und macht iterative Entwicklungszyklen schwerer realisierbar.

6 Laborinformationssystem

tung und konkrete Leitlinien für die HDAB werden aktuell im Projekt TEHDAS2 ausgearbeitet.

Öffentliche Wirkung mitdenken

Um die gesellschaftliche Akzeptanz umfassender Sekundärdatennutzung – auch für industrielle Produktentwicklung – zu sichern, muss über technische und ökonomische Aspekte hinausgedacht werden. [Die laufende Klage der Gesellschaft für Freiheitsrechte gegen das FDZ Gesundheit verdeutlicht, dass fehlende gesellschaftliche Legitimation auch technisch und rechtlich funktionierende Systeme blockieren kann](#). Es bedarf daher eines umfassenden Kommunikationsansatzes, der zu einem gesellschaftlichen Diskurs und schlussendlich zu mündigen Entscheidungen über akzeptierte Datennutzungsformen führt. Für KMU bedeutet das konkret: Wer Gesundheitsdaten für Produktentwicklung nutzt, sollte von Anfang an eine transparente über Zweck, Nutzen und Schutzmaßnahmen kommunizieren; nicht nur als regulatorische Anforderung, sondern als strategische Voraussetzung für gesellschaftliche Akzeptanz und damit für die langfristige Nutzbarkeit der Dateninfrastruktur. Voraussetzung dafür ist, dass die datenbasierten Innovationen, nachvollziehbaren gesellschaftlichen Mehrwert schaffen und einen Beitrag zur Verbesserung der Gesundheitsversorgung leisten. Zugleich benötigen Governance-Strukturen von Datenräumen (zum Beispiel FDZ, HDAB, sphin-X) gezielt Ressourcen für Öffentlichkeitsarbeit und Informationskampagnen, um solche Beispiele sichtbarer zu machen und Vertrauen aufzubauen.

Keine ökonomische Exklusivität vs. Incentivierung des Datenteilens

Wichtig ist in der Entwicklung einer solchen Zukunftsvision ebenfalls die Adressierung ökonomischer Aspekte. Wie der Bundesverband Medizintechnologie (BVMed) im Gespräch deutlich macht, besteht aktuell keine Incentivierung für Datenteilung oder Datenbereitstellung. Gleichzeitig kann eine Incentivierung von Datenbereitstellung nicht dazu führen, dass die Datennutzung prohibitiv teuer wird für Innovatoren im Gesundheitssystem.⁷ Der Bedarf für ökonomisch gerechte und gleichzeitig incentivierende Strukturen muss entsprechend adressiert werden. Ein weiterer Aspekt ist, inwiefern die Nutzung von Daten als öffentliches Gut aller Bürger:innen relevante Effekte für deren Versorgung hat: Wenn Gesundheitsdaten für kommerzielle Produktentwicklung genutzt werden, sollte ein nachweisbarer Rückfluss in die Versorgungsqualität Teil der Legitimationsbasis sein. Über eine politische Forderung hinaus, ist dieser Nachweis als mögliches Kriterium für Datenzugangsgenehmigungen denkbar. Wie in der

Öffentlichkeitsarbeit oben sind hier KMU gefragt, entsprechende Begründungen und Nachweise liefern zu können. Gleichzeitig können Fördermittel helfen entsprechende Nutzeneffekte zunächst zu fördern und damit in Proof-of-Concept-Evaluationen zu verankern. Für Datenräume wiederum bringt Herr Dr. Müller die Frage ein, welche Mechanismen etabliert werden können, um die Datennutzung sowohl gemeinwohlorientiert als auch unternehmerisch attraktiv zu gestalten.

Use-Cases fördern um Mechanismen zu etablieren

Auf dem Weg zur Vision gerechter und innovativer Datennutzung können öffentliche Förderinstrumente wichtige Impulse setzen. Um entsprechende Mechanismen des Datenteilens und Datennutzens auf Basis von Datenräumen zu validieren, eignet sich die Förderung spezifischer medizinischer Use-Cases – im Idealfall entlang der gesamten Patient-Journey für spezifische Indikationen. Ein Beispiel ist die BMFTR-geförderte Medizininformatik Initiative (MII), mit der der Grundstein für eine sektoren- und standortübergreifende Erschließung und Nutzung von Daten für die medizinische Forschung gelegt wurde. Ein Beispiel daraus ist der [MiHUB](#) als geförderter DigiHub, der die Vorteile eines Patient-Journey fokussierten Datenökosystems am Beispiel des Lungenkrebses aufzeigt. Auch in der [Plattform Lernende Systeme](#) wurde ein ähnliches Beispiel, wenngleich ein auf Anamnesen beschränktes Datenökosystem, skizziert und die verschiedenen Vorteile aufgezeigt. Für KMU stellt sich dabei eine kritische Frage, die in bisherigen Förderansätzen zu wenig adressiert wird: Große Use-Case-Förderungen wie die MII sind akademisch dominiert, ihre Ergebnisse, Datenstrukturen und Erkenntnisse verbleiben häufig bei Unikliniken und Konsortialpartnern. KMU, die an solchen Verbundprojekten teilnehmen, müssen sicherstellen, dass IP-Regelungen und Datenzugangsrechte auch nach Projektende fair gestaltet sind. Zukünftige Förderprogramme sollten explizit regeln, unter welchen Bedingungen KMU aus geförderten Use-Cases Nutzungsrechte an entstehenden Datenstrukturen oder Trainingskorpora ableiten können.

Experimentieren mit bestehenden Datenräumen

Weniger umfassend und gleichzeitig relevant wäre das Inkludieren von Vouchern in öffentliche Förderung für die Nutzung von bestehenden Datenräumen. Wie oben gezeigt wurde, entstehen für KMU mitunter die größten Hürden nicht durch die Datennutzung selbst, sondern durch juristische Klärungen, Antragsverfahren, Ethikprüfungen, Vertragsverhandlungen, Datenharmonisierung und technische Integration. Diese Vorlaufkosten sind

7 Der EHDS adressiert dieses Dilemma ansatzweise: Er sieht das Recht der Datenhalter auf Vergütung für die Bereitstellung von Daten zur Sekundärnutzung vor, kombiniert mit der Möglichkeit für Mitgliedstaaten, Gebührenobergrenzen für Datennutzer festzulegen (Verordnung (EU) Nr. 2025/327 Art. 42). Dieses Modell ist ein möglicher Referenzpunkt für eine ausgewogene Lösung, seine praktische Wirksamkeit hängt jedoch von der nationalen Umsetzung ab, die in Deutschland noch aussteht.

nach heutiger Förderlogik kaum förderfähig, obwohl sie über Erfolg oder Scheitern datengetriebener Projekte entscheiden. Europäische Programme zeigen Lösungsansätze auf, indem sie nicht nur Use-Cases fördern, sondern diese Kosten explizit bezuschussen: [EHDEN](#) finanzierte zum Beispiel über einen sogenannten Harmonisation Fund gezielt das Mapping von Daten in ein gemeinsames Datenmodell (OMOP) – also genau die Harmonisierungskosten, die in der Praxis oft den Zugang blockieren. Ähnlich bieten [TEF-Health Open Calls](#) subventionierten Zugang zu Test- und Validierungsleistungen einschließlich Infrastruktur-/Datenressourcen. Solche voucher-ähnlichen Instrumente könnten auch für die Nutzung bestehender Datenräume systematisch eingesetzt werden. Gleichzeitig sollte nach den Prinzipien „public money – public data“ und „public money – public knowledge“ das Teilen und Bereithalten von Daten für Datenräume sowie das Teilen von Erkenntnissen zur Datennutzung aus Datenräumen unter Einhaltung von Datenschutzbestimmungen in die Förderbedingungen eingebettet werden. [Für KMU arbeitet die EU gerade daran, dass das Teilen nicht zum Verlust von geistigem Eigentum beziehungsweise Geschäftsgeheimnissen führt.](#) Dazu zählt auch die Verpflichtung zur Nutzung von Interoperabilitätsstandards in öffentlichen Forschungsprojekten beziehungsweise die Förderung von offenen Schnittstellen zu bestehenden Software-Landschaften.

Verzahnung: AI Act Regulatory Sandboxes, MDR Sandboxes und Datenräume

Zur Stärkung der Innovationslandschaft sind Datenräume ein Element. Zwei weitere regulatorische Instrumente, die auch insbesondere innovativen KMU helfen sollen, sind in Kombination mit Datenräumen relevant:

1. Regulatory Sandboxes nach AI Act (Art. 57–63 AIA) ermöglichen es Unternehmen, Hochrisiko-KI-Systeme unter Aufsicht der zuständigen Behörden unter realen Bedingungen zu entwickeln und zu testen – mit temporären Ausnahmen von bestimmten regulatorischen Anforderungen. Diese Sandboxes entfalten jedoch nur dann ihre volle Wirkung, wenn der Datenzugang für den Testbetrieb gesichert ist. Ohne Klärung, welche Gesundheitsdaten in einer Sandbox unter welchen Bedingungen genutzt werden dürfen, bleibt das Instrument für datengetriebene Medizintechnik-KMU weitgehend wirkungslos.
2. MDR Sandboxes werden im aktuellen Entwurf zur MDR-Reform erstmals vorgesehen. Analog zu den Sandboxes im AI Act sollen sie Herstellern von Medizinprodukten – ermöglichen, neue Produkte und Technologien unter regulatorischer Begleitung zu erproben, bevor ein vollständiges Konformitäts-

bewertungsverfahren durchlaufen wird. Auch hier gilt: Der Nutzen dieser Sandboxes hängt maßgeblich davon ab, ob der Zugang zu klinischen Real-World-Daten für den Erprobungszeitraum strukturiert ermöglicht wird. Andernfalls können Produkte zwar regulatorisch erprobt, aber nicht evidenzbasiert weiterentwickelt werden.

Die Verzahnung dieser drei Instrumente – Datenräume, AI-Act-Sandboxes und MDR- Sandboxes – ist damit keine optionale Ergänzung, sondern eine strukturelle Voraussetzung dafür, dass jedes einzelne Instrument seine Wirkung vollständig entfalten kann. Diese Notwendigkeit wurde auch in aktuellen Arbeiten zur Ausgestaltung von Regulatory Sandboxes im Kontext des AI Act hervorgehoben. [Sie unterstreichen, dass diese Sandbox-Maßnahmen nur ihre volle Wirkung entfalten, wenn sie mit Strukturen für Datenzugang zusammen gedacht werden.](#)

Für die Veröffentlichung haben wir Einzelgespräche mit folgenden Personen geführt:

Jan Kirchhoff, Co-Gründer und Geschäftsführer bei medicalvalues. medicalvalues bringt klinische Disziplinen, Forschungswissen und Datenpunkte zusammen, um Ärzt:innen noch bessere Entscheidungen zu ermöglichen. Dabei werden Labor-, Radiologie- und Anamnese-Daten im Kontext der individuellen Patientensituation verknüpft. medicalvalues hat dazu eine nachvollziehbare, KI-basierte Lösung für den klinischen Alltag entwickelt.

Dr. Christian Müller, Lead Data Innovation Solution & Collaboration bei der Bayer Vital GmbH. Neben seiner Rolle bei Bayer Vital bringt sich Herr Müller auch bei sphin-X als Vorsitzender des Ausschuss Projektideen und beim Verband Forschender Arzneimittelhersteller (VfA) als Befürworter eines einheitlichen Antragsprozesses für die Sekundärdatennutzung in Deutschland ein.

Natalie Gladkov, Leiterin Referat Digitale Medizinprodukte beim Bundesverband Medizintechnologie (BVMed). Der BVMed repräsentiert mehr als 300 Hersteller, Zulieferer und Händler der MedTech-Branche sowie Hilfsmittel-Leistungserbringer und Homecare-Versorger.

Herausgeber:

VDI/VDE Innovation + Technik GmbH

Steinplatz 1 | 10623 Berlin

www.vdivde-it.de

Bildnachweis:

elenabs/istockphoto

© VDI/VDE-IT 2026