

# DATA CARE (FKZ 01GP2112B) Kurzfassung

## Abschlussbericht

Das Karlsruher Institut für Technologie beteiligte sich mit seiner Kompetenz im technischen Datenschutz und der nutzerzentrierten Anwendungsentwicklung am Gesamtvorhaben des Projekts DATA CARE.

In enger Abstimmung mit den weiteren Projektpartnern wurde ein Szenario für eine DATA CARE Plattform entwickelt, die als technisches Rahmenwerk für das Projekt dient. Zusammen mit der Goethe Universität Frankfurt am Main wurde ein technisch-juristisches Konzept entwickelt, das die datenschutzkonforme Nutzung der medizinischen Daten unter Miteinbeziehungen der Patientinnen ermöglicht. Das technische-juristische Konzept zur Datensouveränität legt den Schwerpunkt auf ein digitales Einwilligungsmanagement, das für die sekundäre Nutzung von Gesundheitsdaten und die aktive Einbindung von Patient:innen unerlässlich ist. Die Datenschutzgrundverordnung (DSGVO) stellt hierfür den rechtlichen Rahmen bereit, insbesondere durch die Anforderungen an explizite und zweckgebundene Einwilligungen. Privatsphäre-wahrende Technologien wie k-Anonymität, Differential Privacy oder homomorphe Verschlüsselung sind essenziell, um das Re-Identifikationsrisiko zu minimieren. Studien und Softwarelösungen zeigen Ansätze zur Quantifizierung des Datenschutzrisikos, wobei ein standardisierter und anwendungsspezifischer Ansatz erforderlich ist. Ein innovativer Vorschlag ist die Einbindung einer externen Vertrauensstelle, die Populationsstatistiken nutzt, um das Re-Identifikationsrisiko vor einer Datenspende zu bewerten und Nutzer:innen eine fundierte Entscheidung zu ermöglichen. Dieses Konzept bildet die Grundlage des DATA CARE-Rahmenwerks, das eine interdisziplinäre und patientenorientierte Perspektive verfolgt.

Dieses Konzept wurde verwendet, um ein das Rahmenwerk der DATA CARE Architektur zu entwerfen. Das DATA CARE Rahmenwerk wurde entwickelt, um über den Prototyp hinaus die technisch-juristischen Aspekte des Projekts umfassend zu definieren und Szenarien abzustecken, die den Projektumfang überschreiten. Es beschreibt die Architektur der DATA CARE-App „HealthHub“, die Patient:innen Einblicke und Kontrolle über ihre medizinischen Daten ermöglicht, und beleuchtet die klinische Plattform „4D Klinik“, eine interdisziplinäre Einrichtung für die Behandlung und Forschung zu Immunerkrankungen. Die App setzt auf „Dynamic Consent“, um maximale Transparenz und Kontrolle zu gewährleisten, während eine KI unstrukturierte Daten analysiert und für Arzttermine oder Studien aufbereitet. Datenschutz wird durch eine vertrauenswürdige Stelle und Technologien wie attributbasierte Zugriffskontrolle und Differential Privacy gewährleistet. Insgesamt bildet das Rahmenwerk die Grundlage für die Verbindung von technischer Innovation, Nutzerfreundlichkeit und rechtlicher Sicherheit im Projekt DATA CARE.

Neben den konzeptionellen Architektur entwürfen wurde im Projekt auch eine Demonstrator für die Konzepte entwickelt. Der Demonstrator wurde nutzerzentriert und iterativ entwickelt, um die Bedürfnisse der Betroffenen in den Fokus zu rücken. In Workshops mit den Projektpartnern wurden Anforderungen erhoben und in einem zyklischen Prozess umgesetzt. Das App-Konzept ermöglicht eine intuitive Verwaltung und Freigabe medizinischer Daten mit vollem Nutzerfokus, ergänzt durch modulare Anpassungsfähigkeit. Datenschutz und Nachvollziehbarkeit standen dabei im Mittelpunkt, während Design und Funktionalität kontinuierlich optimiert wurden, um praxisnahe Szenarien wie Arzttermine oder Forschungsdatenspenden zu unterstützen.

Im ersten Halbjahr 2024 wurde der Demonstrator in einer Nutzenstudie mit sechs Teilnehmenden der Rheumaliga evaluiert (66 % männlich, 24 % weiblich, Alter 55–75). Die Teilnehmenden testeten remote drei Hauptfunktionen: Datenübersicht, Terminvorbereitung und Forschungsdatenfreigabe. Mit einem SUS-Wert von 72 wurde die Usability als gut bewertet, jedoch zeigten sich Optimierungspotenziale bei Navigation, Barrierefreiheit und Datenschutztransparenz. Besonders positiv wurden die KI-gestützte Terminvorbereitung und die Möglichkeit zur Forschungsdatenfreigabe wahrgenommen. Die Ergebnisse liefern wertvolle qualitative Erkenntnisse für die weitere Entwicklung.

Im Projekt wurde erfolgreich ein technisch-juristisches Rahmenwerk für eine medizinische Datenplattform entwickelt, das die Einbeziehung von Patientinnen ermöglicht. Dieses diente der iterativen Entwicklung eines mobilen Demonstrators, der die Einsicht, Verwaltung und Freigabe medizinischer Daten für Forschung ermöglicht. Die Evaluierung durch Probandinnen der Rheumaliga zeigte eine gute Usability und lieferte wertvolle Erkenntnisse für die Weiterentwicklung.

Das Rahmenwerk und die modulare Architektur des Demonstrators bieten eine solide Grundlage für zukünftige Projekte und Forschungen, wobei die gewonnenen Erfahrungen auch in andere Forschungs- und Industrievorhaben integriert werden können.



**Datensouveränität und informierte Zustimmung als Grundlage für  
eine patientenorientierte KI-gesteuerte klinische Forschung –  
DATA CARE**

**Schlussbericht**

01.12.2021 – 30.11.2024

**Zuwendungsempfänger:**  
Karlsruher Institut für Technologie

**Förderkennzeichen:**  
01GP2112B

**Gefördert durch:**



**Betreut vom Projektträger**



**Autoren:** Dr.-Ing. Arno Appenzeller, Prof. Dr.-Ing. habil. Jürgen Beyerer

# 1 Inhalt

<b>1 INHALT .....</b>	<b>2</b>
<b>2 KURZDARSTELLUNG .....</b>	<b>3</b>
2.1 AUFGABENSTELLUNG .....	3
<b>3 ERGEBNISBERICHT .....</b>	<b>5</b>
3.1 TECHNISCH-JURISTISCHE KONZEPTION.....	5
3.2 DATACARE RAHMENWERK .....	10
3.3 DEMONSTRATOR .....	15
3.4 NUTZENDENSTUDIE.....	19
3.5 FAZIT .....	23
<b>4 VORAUSSICHTLICHER NUTZEN &amp; VERWERTBARKEIT DER ERGEBNISSE .....</b>	<b>25</b>
<b>5 FORTSCHRITT AUF DIESEM GEBIET BEI ANDEREN STELLEN</b>	<b>26</b>
<b>6 VERÖFFENTLICHUNGEN .....</b>	<b>27</b>

## 2 Kurzdarstellung

### 2.1 Aufgabenstellung

Die verfügbare Menge elektronischer Daten aus der Gesundheitsforschung und -versorgung wächst rasant. Big Data und Künstliche Intelligenz (KI) bedeuten eine große Chance für die klinische Forschung. Gleichzeitig stärkt die neue Datenschutz-Grundverordnung der EU (DSGVO) die Souveränität jedes Einzelnen, über seine persönlichen Informationen selbst zu entscheiden. Selbstbestimmung und Einwilligung sind damit zentrale Schlüssel für die Datensouveränität von Patientinnen und Patienten. Die konsequente Umsetzung der Datensouveränität in der klinischen Forschung hat enorme Konsequenzen für die Prozesse der Datenerhebung und -verarbeitung in der Praxis. Das zentrale Ziel von DATACARE ist es, ein Konzept für Datensouveränität und informierte Einwilligung für die klinische Forschung unter Einbindung rechtlicher, klinischer, ökonomischer sowie technischer Forschungsexpertise zu entwickeln und in einem Prototyp umzusetzen. Dabei wird mit Blick auf die Akzeptanz der Ergebnisse von DATACARE besonderes Augenmerk auf die Aspekte der Nutzerorientierung und Partizipation gerichtet und die Beteiligung von Patientinnen- und Patientenvertretenden, Selbsthilfeorganisationen, Klinikerinnen und Klinikern sowie der Pharmaindustrie konsequent von Anfang an im Projekt verankert. Die Ergebnisse von DATACARE leisten einen wichtigen Beitrag für einen informierten wissenschaftlichen und gesellschaftlichen Diskurs und sollen entsprechend an die breite Öffentlichkeit, Politik und Wissenschaft kommuniziert werden.

In der heutigen klinischen Praxis werden viele Daten von Patientinnen und Patienten erhoben. Diese müssen stapelweise Einwilligungserklärungen lesen und unterschreiben – oft innerhalb eines kurzen Zeitrahmens und ohne angemessene Beratung durch das Klinikpersonal. Das hat zur Folge, dass die Patient\*innen nicht vollumfänglich informiert und selbstbestimmt sind und somit keine Datensouveränität über ihre eigenen medizinischen Daten haben, wie es die EU-Datenschutz-Grundverordnung fordert. Darüber hinaus werden die vorhandenen Daten für Forschungszwecke eingesetzt. Klinische Forschung, die KI bei der Datenanalyse einsetzt, erfordert große Datenbanken. Dafür ist es notwendig, die Erlaubnis der Patientinnen und Patienten zur Nutzung der bereits vorhandenen Daten für verschiedene Forschungszwecke einzuholen.

Das zentrale Ziel von DATACARE ist es, einen Beitrag zum wissenschaftlichen und gesellschaftlichen Diskurs zu leisten, indem eine Blaupause für Datensouveränität und informierte Zustimmung als Befähiger für patientenorientierte KI-getriebene klinische Forschung entwickelt und in einem Prototyp umgesetzt wird.

Das Karlsruher Institut für Technologie (KIT) unterstützt hierbei in den Arbeitspaketen der Kontextanalyse und der Kommunikation. Im größten Arbeitspaket zu Konzept und Prototyp verantwortet das KIT das technisch-juristische Konzept für Datensouveränität

und das Interaktionsdesign und Evaluation. Zusätzlich wird in den Paketen der Quantifizierungsmodelle und der KI unterstützt.

## 3 Ergebnisbericht

Die Arbeitspaket, die dem KIT zugeordnet sind lassen sich in die Themenblöcke technisch-juristische Konzeption, DATACARE Rahmenwerk, Demonstrator und Nutzenstudie einordnen. Im Folgenden werden die Ergebnisse dieser Themenblöcke dargestellt und ein abschließendes Fazit zu den Projektergebnissen gezogen.

### 3.1 Technisch-juristische Konzeption

Für das technische-juristische Konzept wurde zuerst der Begriff Datensouveränität aus technischer Sicht betrachtet.

Eine Komponente davon ist die Verwendung eines digitalen Einwilligungsmanagements. Diese ist vor allem für den Anwendungsfall der sekundären Nutzung von Gesundheitsdaten relevant. Gemäß Datenschutzgrundverordnung (DSGVO) Art. 9 Abs.1 zählen persönliche Gesundheitsdaten zur Kategorie der besonders sensiblen Daten deren Verarbeitung erstmal pauschal untersagt sind. Art. 9 Abs. 2 definiert allerdings Erlaubnistatbestände unter deren Umständen eine Verarbeitung dennoch möglich ist. Eine davon ist die explizite Einwilligung von Betroffenen. Durch die wachsende Datenmenge entsteht allerdings auch eine wachsende Anzahl von Einwilligungen, die ein Management benötigen. Des Weiteren sind Einwilligungen im Sinne der wachsenden Nachfrage nach aktiver Teilhabe von Patient:innen. Aktuell existierende Technologien bieten keinen automatischen Workflow, um Einwilligungen digital abzubilden und automatisiert auszuwerten. Dies ist Gegenstand von verschiedenen wissenschaftlichen Projekten und Arbeiten, die sich mit dem Thema beschäftigen. Hierbei gilt es allerdings festzustellen, dass dies selten aus der Sicht der Betroffenen erfolgt.

Weitere Aspekte für Datensouveränität sind sogenannte Privatsphäre wahrende Technologien. Diese sind erforderlich, da auch vollständig anonymisierte Daten ein Risiko der Re-Identifikation haben. Dies wird eindrucksvoll durch populäre Beispiele, wie von Latanya Sweeney im Jahr 2002<sup>1</sup> gezeigt. Hier wurde ein Datensatz aus dem Wählerregister des US Bundesstaates Massachusetts, der identifizierende Daten enthielt, mit einem scheinbar anonymen medizinischen Datensatz verknüpft. Da beide Datensätze Geschlecht, Postleitzahl und Alter gemeinsam hatten, ließ sich hierdurch die medizinische Akte des Gouverneurs von Massachusetts identifizieren. Dies stellt eine eindrucksvolle Re-Identifizierungsattacke dar, die auch die Notwendigkeit von Privatsphäre wahren Technologien unterstreicht. Mit diesen Verfahren kann das Re-Identifizierungsrisiko verringert werden. Im Projekt wurde eine solche Risikoeindämmung bzw. ein solches Risikobewusstsein als Notwendigkeit für Datensouveränität erachtet.

---

<sup>1</sup> SWEENEY, Latanya: „*k*-Anonymity: A Model for Protecting Privacy“. In: *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10.05 (Okt. 2002)

Beispiele für Technologien sind  $k$ -Anonymität,  $l$ -diversität, Differential Privacy oder auch Homomorphe Verschlüsselung<sup>2</sup>.

Darüber hinaus wurde aus technischer Sicht ein Blick auf die DSGVO und der Darstellung des Themas Datensouveränität geworfen. Hierbei wurde festgestellt, dass Datensouveränität als Informationelle Selbstbestimmung +  $X^3$  bezeichnet werden kann und die Digitalisierung generell zu einem enormen Zuwachs an personenbezogenen digitalen Daten führt. Wie zuvor erwähnt ist die Einwilligung für die Verarbeitung solcher Daten ein üblicher Weg. Für diese definiert die DSGVO auch Anforderungen. So müssen Einwilligungen beispielsweise stets zweckgebunden sein und dürfen nicht nach dem Opt-Out Prinzip funktionieren.

Im Rahmen der Publikation innerhalb des Projektverbundes wurde auf Basis dieser Betrachtung zur Technologische Perspektive Patientensouveränität eine Literaturanalyse erstellt. Diese wurde als „Data sovereignty requirements for patient-oriented AI-driven clinical research in Germany“<sup>4</sup> veröffentlicht und die Ergebnisse des KIT Beitrags wird hier verkürzt dargestellt:

Aus technologischer Sicht gibt es grundlegende Technologien, die die Datensouveränität und die informierte Zustimmung der Patienten ermöglichen: ein digitales Zustimmungsmagementsystem, Technologien zum Schutz der Privatsphäre und die Quantifizierung des Risikos für die Privatsphäre.

Eine Hauptanforderung ist ein digitales Einwilligungsmanagementsystem. Gemäß der europäischen Datenschutzgrundverordnung (GDPR) gelten medizinische Daten als sensible Daten, deren Verarbeitung daher standardmäßig nicht zulässig ist. Es gibt spezielle Ausnahmen, die die Datenverarbeitung erlauben, und eine davon ist die ausdrückliche Zustimmung der betroffenen Person. Für die medizinische Forschung ist dies die häufigste Rechtsgrundlage für die Verarbeitung medizinischer Daten. In den Erwägungsgründen 32 und 33 der Datenschutz-Grundverordnung werden weitere Anforderungen an die Einwilligung festgelegt. Die wachsende Menge an medizinischen Daten erfordert auch ein granulares Einwilligungsmanagement, um eine aktive Beteiligung und Einbeziehung der Patienten zu ermöglichen. In einer zuvor erschienenen Übersichtsarbeit über elektronische Einwilligungen untersuchten Verreydt et al.<sup>5</sup> 31 Publikationen über informierte und elektronische Einwilligungen, die zwischen 2010 und

---

<sup>2</sup> Siehe bspw. MACHANAVAJHALA, Ashwin; KIFER, Daniel; GEHRKE, Johannes und VENKITASUBRAMANIAM, Muthuramakrishnan: „ $l$ -diversity: Privacy Beyond  $k$ -Anonymity“. In: ACM Transactions on Knowledge Discovery from Data 1.1 (März 2007)

<sup>3</sup> Siehe: Datensouveränität für Patienten im Gesundheitswesen: Eine Chance für die medizinische Forschung und den Datenschutz; S Bretthauer, A Appenzeller, P Birnstil - Datenschutz und Datensicherheit-DuD, 2021

<sup>4</sup> Data sovereignty requirements for patient-oriented AI-driven clinical research in Germany; M Radic, J Busch-Casler, A Vosen, P Herrmann - Ethik in der Medizin, 2024

<sup>5</sup> Stef Verreydt, Koen Yskout, and Wouter Joosen. 2021. Security and Privacy Requirements for Electronic Consent: A Systematic Literature Review. ACM Trans. Comput. Healthcare 2, 2, Article 16 (April 2021), 24 pages. <https://doi.org/10.1145/3433995>

2019 veröffentlicht wurden. Die Autoren konzentrieren sich auf die Auswirkungen auf die Sicherheit und den Datenschutz bei digitalen Einwilligungen. Sie kommen zu dem Schluss, dass es keinen Konsens über die Anforderungen an Sicherheit und Datenschutz gibt, und schlagen vor, dass dies ein Schwerpunkt für künftige Systeme sein sollte. Aus technischer Sicht wird außerdem festgestellt, dass die eXtensible Access Control Markup Language (XACML) häufig verwendet wird und daher die empfohlene Technologie ist. Im Hinblick auf bestehende digitale Zustimmungsmanagementsysteme stellen Bahls et al.<sup>6</sup> ein System namens generic Informed Consent System (gICS) vor. gICS ist ein Softwaresystem, das Zustimmungserklärungen digital erstellt und verwaltet. Es bietet auch Vorlagen für die Erstellung neuer Einwilligungserklärungen. Das System erfordert jedoch nach wie vor das Ausfüllen von Formularen in Papierform durch die betroffene Person. Das Ergebnis dieser papiergestützten Einwilligung kann dann von gICS digital erfasst werden. Eine neuere Arbeit von Appenzeller et al.<sup>7</sup> schlägt einen vollständigen Arbeitsablauf für eine souveräne dynamische Einwilligung vor. Dieser Ansatz definiert die Anforderungen für diese Art der Einwilligung und stellt den Patienten in den Mittelpunkt des Prozesses, wodurch die Einbeziehung und Akzeptanz des Patienten in den Prozess der medizinischen Forschung verbessert wird. Darüber hinaus wird ein komplettes technisches System beschrieben, das eine dynamische digitale Einwilligung für granulare Entscheidungen ermöglicht und eine Quantifizierung der Auswirkungen auf die Privatsphäre der gemeinsam genutzten Daten bietet, so dass die betroffene Person eine informierte Einwilligungsentscheidung treffen kann.

Ein technologischer Baustein, der innerhalb des DATACARE Konsortiums für die Datensouveränität identifiziert wurde, ist die Quantifizierung des Datenschutzrisikos. Die digitale Einwilligung gibt dem Patienten zwar einerseits mehr Befugnisse, andererseits kann der Umfang der Wahlmöglichkeiten die betroffene Person aber auch überfordern. Daher sollten verständliche Quantifizierungen des Datenschutzrisikos die Patienten bei ihrer Entscheidung über die Weitergabe von Daten unterstützen. Es gibt bereits eine Vielzahl unterschiedlicher Ansätze zur Quantifizierung des Datenschutzrisikos für personenbezogene Daten, die in zwei Kategorien eingeteilt werden können.

Die erste sind datenbasierte Quantifizierungen, bei denen die gemeinsam genutzten Daten auf ein Preisschild für die Privatsphäre hin analysiert werden (z. B. Deusser et al.<sup>8</sup>). Dieses Preisschild wird auf der Grundlage verschiedener Faktoren berechnet. Dabei kann es sich um die Einzigartigkeit der Daten, die Größe der gemeinsam genutzten Datenbank oder die mögliche Geldstrafe im Falle eines Datenlecks handeln. Die andere Kategorie der Quantifizierung von Datenschutzrisiken sind regelbasierte Quantifizierungen, bei

---

<sup>6</sup> Bahls, T., Liedtke, W., Geidel, L., Langanke, M. (2015). Ethics Meets IT: Aspects and Elements of Computer-based Informed Consent Processing. In: Fischer, T., Langanke, M., Marschall, P., Michl, S. (eds) Individualized Medicine. Advances in Predictive, Preventive and Personalised Medicine, vol 7. Springer, Cham. [https://doi.org/10.1007/978-3-319-11719-5\\_11](https://doi.org/10.1007/978-3-319-11719-5_11)

<sup>7</sup> Appenzeller A, Hornung M, Kadow T, Krempel E, Beyerer J. Sovereign Digital Consent through Privacy Impact Quantification and Dynamic Consent. *Technologies*. 2022; 10():35. <https://doi.org/10.3390/technologies10010035>

<sup>8</sup> Deusser, Clemens, Steffen Passmann and Thorsten Strufe. "Browsing Unicity: On the Limits of Anonymizing Web Tracking Data." 2020 IEEE Symposium on Security and Privacy (SP) (2020)

denen z. B. Datenschutzrichtlinien oder andere textbasierte Richtlinien, die sich auf die Privatsphäre der gemeinsam genutzten Daten auswirken, analysiert werden (z. B. Kelley et al.<sup>9</sup>). Ein offenes Thema bei der Risikoquantifizierung ist die Schnittstelle zwischen technischen Quantifizierungen und dem Interaktionsdesign, das die Auswirkungen der medizinischen Daten eines Patienten auf die Privatsphäre erklärt und visualisiert. Aus technologischer Sicht generieren Kelley et al. menschenlesbare Datenschutzrichtlinien aus maschinenlesbaren Regeln, die auf einem Standard für Datenschutzrichtlinien (P3P) basieren. Sie bewerten verschiedene Visualisierungen für Datenschutzrichtlinien und identifizieren die Visualisierung als einen Schlüsselfaktor für die Verständlichkeit.

Im weiteren Projektverlauf wurde das Re-Identifikationsrisikos bei Datenweitergabe als eine weitere der wissenschaftlichen Kernfragen zum technisch-juristischen Konzept identifiziert. Hierzu wurde die Publikation „Das Re-Identifikationsrisiko bei der Weitergabe von Gesundheitsdaten“<sup>10</sup> in der Zeitschrift Datenschutz und Datensicherheit veröffentlicht. Der technische Beitrag konzentriert sich auf die technische Messung des Re-Identifikationsrisikos und wird hier verkürzt dargestellt:

Der bisherige technische Stand der Forschung bietet eine breite Auseinandersetzung mit der Thematik der Re-Identifikation und der Messung eines entsprechenden Risikos. Neben einer Vielzahl an Publikationen existieren auch einige teils frei verfügbare Softwarelösungen.

Eine relevante Publikation erschien bereits im Jahr 2012 von Dankar und Anderen<sup>11</sup>. In dieser Studie wird das Konzept der Einzigartigkeit als Maß für das Re-Identifikationsrisiko beachtet. Die Autoren weisen allerdings darauf hin, dass es in der Praxis häufig schwierig ist, Einzigartigkeit genau zu messen. Das Szenario der Studie ist das Angriffsmodell, in dem ein anonymisierter medizinischer Datensatz gegen das öffentliche Wählerregister gematched wird, um Daten zu linken. Im Weiteren werden vier verschiedene mathematische Metriken für Einzigartigkeit vorgestellt. Alle Varianten basieren darauf die Anzahl gleicher Quasi-Identifizierenden Merkmale in Relation zum Quell- oder Zieldatensatz zu setzen. Die Metriken wurden evaluiert und es zeigt sich, dass alle unterschiedlich gut in verschiedenen Szenarien abschneiden. Letztendlich plädieren die Autoren dafür, dass die Bestimmung des Re-Identifizierungsrisiko unerlässlich ist, sobald Gesundheitsdaten für sekundäre Zwecke genutzt werden. Eine weitere Publikation von Ratra et al. verwendet die Techniken  $k$ -Anonymität und Differential Privacy (DP) um das Re-Identifikationsrisiko von Gesundheitsdaten zu bestimmen<sup>12</sup>. Durch die kombinierte aufeinanderfolgende Anwendung von  $k$ -Anonymität und DP sollen Angreifer mit

---

<sup>9</sup> Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W. Reeder. 2009. A “nutrition label” for privacy. (2009).;

<sup>10</sup> Das Re-Identifikationsrisiko bei der Weitergabe von Gesundheitsdaten: Eine rechtliche und technische Analyse; A Appenzeller, B Orak - Datenschutz und Datensicherheit-DuD, 2024

<sup>11</sup> Dankar, F. K., El Emam, K., Neisa, A., & Roffey, T. (2012). Estimating the re-identification risk of clinical data sets.

<sup>12</sup> Ratra, R., Gulia, P., & Gill, N. S. (2022). Evaluation of Re-identification Risk using Anonymization and Differential Privacy in Healthcare

Hintergrundwissen daran gehindert werden, Betroffene anhand ihres sensitiven Attributes zu re-identifizieren. Zur Messung des Re-Identifizierungsrisiko wurde die ARX Software verwendet und es wurden gute Resultate erzielt, die einen Mehrwert darstellen. Neben den hier vorgestellten Publikationen existiert auch eine Reihe von Softwarelösungen die Möglichkeiten der Messung des Re-Identifikationsrisiko bieten. ARX<sup>13</sup> ist eine Anonymisierungssoftware, die quelloffen verfügbar ist. Neben verschiedenen Anonymisierungsfunktionen bietet ARX auch die Möglichkeit der Durchführung von Risikoanalysen. Eine mögliche Analyse ist die Re-Identifikationsrisikoanalyse. Dafür werden verschiedene Angreifermodelle betrachtet beispielsweise das sogenannte Prosecutor Model, dass ein spezifisches Ziel für den Re-Identifizierungsversuch betrachtet. Dabei wird bewertet, wie wahrscheinlich es ist anhand der quasi identifizierenden Merkmale eine Person zu erkennen, wobei dem Angreifer bekannt ist, dass das Ziel im Datensatz ist. Das quelloffene Softwarepaket sdcMicro<sup>14</sup> verwendet eine frequenzbasierte Risikomessung. Zusätzlich kann auch noch die bewährte Metrik l-diversity verwendet werden, um das Re-Identifizierungsrisiko zu messen. Neben den frei verfügbaren Lösungen existieren auch kommerzielle Programme wie Eclipse Risk<sup>15</sup>. Der Hersteller gibt Risikobewertungsmethoden an, aber spezifiziert diese nicht mehr als um die Angabe, dass statistische Methoden zur Messung des Re-Identifizierungsrisiko verwendet werden.

Die Übersicht über Publikationen und Softwarelösungen zeigt, dass eine Vielzahl an Lösungen gibt, um das Re-Identifizierungsrisiko zu messen. Es lässt sich allerdings feststellen, dass es sich bei den Bewertungen häufig um festdefinierte Modelle handelt oder Best-Practices, die für eine Reihe von Anwendungsfällen gut geeignet sind, aber nicht vollständig generalisieren und limitiert sind. Deswegen ist für jeden Anwendungsfall eine spezifische Analyse nötig. Für diese Schritte sollte systematisiert das Bedrohungsszenario definiert und mit welchen Maßnahmen dieses verringert werden kann. Auf dieser Basis kann dann auch entsprechendes Re-Identifizierungsrisiko berechnet werden.

Als eigener Vorschlag wird ein Verfahren zur Messung des Re-Identifikation medizinischer Daten, basierend auf der Messung der Seltenheit der Daten auf Basis öffentlicher populationsweiter statistischer Daten, wie Altersverteilung und Geschlechterverteilung, zu einer Krankheit, vorgeschlagen. Hierfür wird vorgeschlagen eine vertrauenswürdige Stelle für die Datenfreigaben zu etablieren. Vor einer möglichen Datenspende werden dafür der ICD-10 Code und die quasi-identifizierenden Merkmale an eine externe Vertrauensstelle gesendet. Diese Vertrauensstelle greift auf die populationsweiten Statistiken wie Krankheitsprävalenzen oder -inzidenzen zu, um Informationen über die Häufigkeit von Erkrankungen oder ähnliche Daten zu erhalten. Solche Daten können, wie zuvor erwähnt, beispielsweise vom ZfKD oder ähnlichen Stellen stammen.

---

<sup>13</sup> <https://arx.deidentifier.org>

<sup>14</sup> <https://sdctools.github.io/sdcMicro/index.html>

<sup>15</sup> <https://privacy-analytics.com/health-data-privacy/health-data-software/eclipse-risk/>



Abbildung 3.1.1 Schema zu Umsetzungsmöglichkeit für Re-Identifikationsrisikomessung für medizinische Daten

Das Vorgehen ist dann wie folgt:

1. Nutzende Person erwägt für ein Forschungsprojekt angefragte Daten freizugeben
1. Vor der Freigabe wird der mit den Daten korrespondierende ICD-10 Code an die Vertrauensstelle gesendet
2. Die Vertrauensstelle ruft Populationsstatistiken für ICD-10 Code ab (Quellen können Institutionen wie das ZfKD sein)
3. Auf Basis der Populationsstatistik wird die Risikoklasse bestimmt
4. Die betroffene Person erhält die Risikoklassifizierung der Datenspende und kann auf dessen Basis eine informierte Entscheidung treffen.

Auf Basis dieser interdisziplinären Betrachtungen ist das DATACARE Rahmenwerk entstanden, dass im folgenden Abschnitt beschrieben wird.

### 3.2 DATACARE Rahmenwerk

Das DATACARE Rahmenwerk wurde entworfen, da ein Prototyp immer nur einen begrenzten Umfang besitzen kann und gerade im Rahmen eines Forschungsprojekts einen sehr eng beschränkten Rahmen absteckt, dient ein solches Dokument als Rahmenwerk für die weiteren Aspekte des technisch-juristischen Konzeptes.

Das Rahmenwerk soll das Szenario rund um den Demonstrator, die DATACARE App, abstecken und technisch ausdefinieren. Dies sind in der Regel Aspekte deren technische Umsetzung den Rahmen des Projektes sprengen würde oder die aus diversen Gründen nicht realisierbar sind, aber dennoch eine genaueren Betrachtung benötigen.

Im Idealfall bildet das Rahmenwerk in Zusammenspiel mit dem Demonstrator die Voraussetzungen, um die technisch-juristischen Zielsetzungen des Projektes umzusetzen.

Das DATACARE Rahmenwerk besteht aus 5 inhaltlichen Kapiteln:

1. Architektur DATACARE
2. Klinische Plattform „4D Klinik“
3. Nutzende
  - I. Patient:innen
  - II. Ärzt:innen
  - III. Forschende
  - IV. Industrie
4. DATACARE Prototyp: „HealthHub“

- I. Funktionalität
  - II. Technische Umsetzung
  - III. Abgrenzung und Limitierungen
5. Forschungsschnittstellen

Zentraler Aspekt ist die Architektur von DATACARE, die im folgenden Schaubild dargestellt wird.

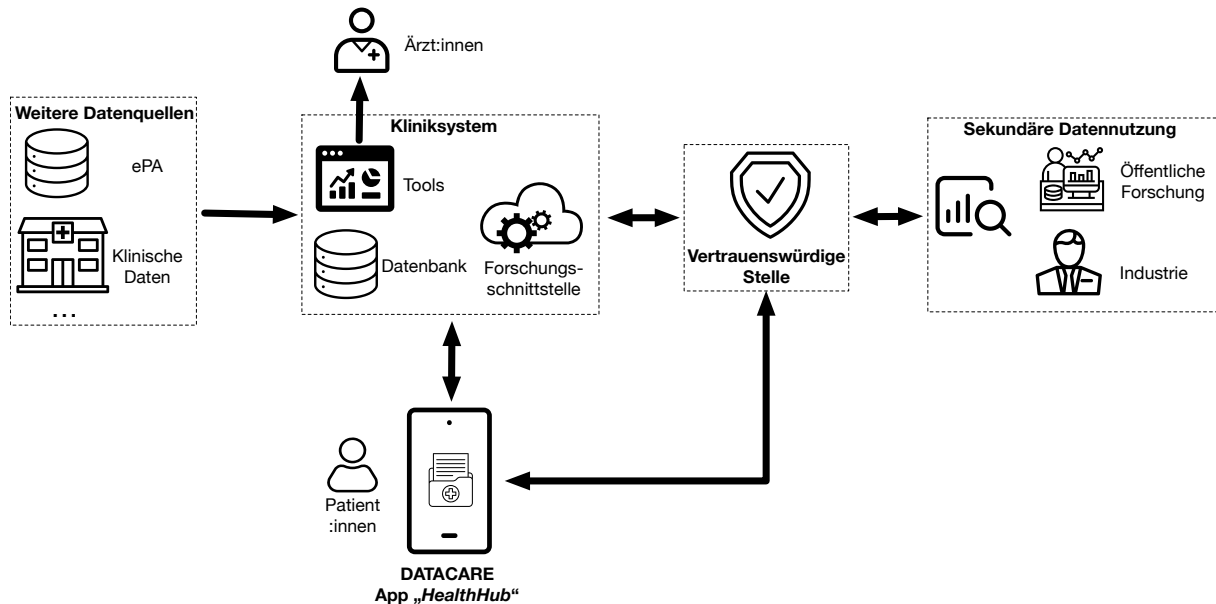


Abbildung 3.2.1: Schematische Darstellung der Architektur des DATACARE Konzeptes

Die Abbildung visualisiert die beteiligten Komponenten der Architektur, die möglichen Datenflüssen zwischen diesen Komponenten und die berücksichtigten Nutzenden.

Kernstück der DATACARE Architektur ist die sogenannte **DATACARE -App „HealthHub“**, die auch als Demonstrator des Projektes dient. Die Nutzenden der App sind die Patient:innen. Die App soll es ermöglichen, die eigenen medizinischen Befunde einzusehen und beispielsweise deren Freigaben zu verwalten. Dafür kommuniziert die App mit dem Kliniksystem und der Vertrauenswürdigen Stelle.

Das **Kliniksystem** ist das Zentrum der Datenhaltung im vorliegenden Rahmenwerk. Teil des Kliniksystems sind Datenbanken, die zum einen aus vorliegenden Datenquellen wie dem Krankenhausinformationssystem befüllt werden und auch Daten aus weiteren externen Datenquellen bezieht. Auf die eigenen Daten innerhalb dieser Datenbank können die Betroffenen über die DATACARE-App zugreifen. Anwendungen wie das Krankenhausinformationssystem oder nicht näher definierte Analyse-/Befundungswerkzeuge werden unter den Tools innerhalb des Kliniksystems zusammengefasst. Mit diesen interagieren auch die Ärzt:innen im Rahmen der Architektur. Die Tools greifen auf die vorliegende Datenbank zu und visualisieren diese oder erfassen Daten dafür. Die letzte Subkomponente innerhalb des Kliniksystems ist die Forschungsschnittstelle. Hier werden Daten aus dem Kliniksystem für die sekundäre Datennutzung aufbereitet. So werden hier zum einen Einwilligungen von Patient:innen bei der Datenfreigabe berücksichtigt, als auch das Re-Identifikationsrisiko bemessen.

Darüber hinaus können hier Anonymisierungs- und Pseudonymisierungsmaßnahmen vorgenommen werden. Grundsätzlich fließen alle Daten für die sekundäre Nutzung durch die Forschungsschnittstelle.

Die **weiteren Datenquellen**, die die Datenbank des Kliniksystems füllen, können beispielsweise die elektronische Patientenakte beziehungsweise die Telematik-Infrastruktur sein. Hier könnte eine Anbindung an deren Schnittstellen zum Datenaustausch vorgenommen werden. Weiterhin ist die Anbindung weiterer externer Kliniken denkbar. Die genauen weiteren Datenquellen werden für dieses Rahmenwerk nicht definiert, es gilt lediglich zu beachten, dass die Daten nicht aus derselben Quelle stammen müssen.

Die **Vertrauenswürdige Stelle** übernimmt die Rolle als Datentreuhand vor jeglicher Datenweitergabe für die Sekundärnutzung. Die Daten werden hier beispielsweise auf ihr Re-Identifikationsrisiko geprüft oder die Datenfreigabe auf ihre rechtliche Grundlage. Neben dem Datenfluss vom Kliniksystem durch die Forschungsschnittstelle, besteht auch die Möglichkeit eines direkten Datenflusses von den Betroffenen durch die DATACARE-App zu den Forschenden. Auch hier ist die Vertrauenswürdige Stelle die mittelnde Stelle.

Die letzte Komponente sind die Stellen, an denen die sekundäre Datennutzung ausgeführt wird. Dies soll Parteien darstellen, die Empfangende der Daten aus der Vertrauenswürdigen Stelle sind. Als mögliche Teilnehmenden werden im Rahmenwerk Forscher:innen, aber auch die Industrie genannt.

Im Kapitel klinische Plattform „4D Klinik“ wird das Anwendungsszenario für das DATACARE Rahmenwerk spezifiziert.

Die 4D-Klinik ist eine interdisziplinäre Versorgungs- und Forschungseinrichtung für Patient:innen mit Immunerkrankungen, die mehrere Organsysteme betreffen und daher eine umfassende medizinische Betreuung erfordern. Die Bezeichnung „4D“ steht für „Drugs, Devices, Diagnostics und Data“ und beschreibt die Zusammenarbeit von Mediziner:innen, Wissenschaftler:innen und Technikern. In der Klinik erfolgt die Untersuchung und Behandlung durch Fachkräfte der Rheumatologie, Dermatologie und Gastroenterologie, unterstützt durch eine enge Zusammenarbeit der beteiligten Fachbereiche zur Behandlung systemübergreifender Symptome.

Patient:innen, die an verschiedenen Organen betroffen sind, werden an die Klinik überwiesen. Die Teilnahme an einer Forschungslangzeitstudie ermöglicht zusätzliche molekularbiologische Analysen, um eine frühzeitige Diagnose und individuell zugeschnittene Therapien zu verbessern. Die Teilnahme an der Studie ist freiwillig, erfolgt pseudonymisiert und umfasst auch eine zusätzliche Blutentnahme sowie Erfassung klinischer und bildgebender Daten.

Die Klinik nutzt moderne Dokumentations- und Informationssysteme wie das Krankenhausinformationssystem (KIS) für medizinische Daten und zur Planung. Das eCRF-System wird verwendet, um Studienteilnehmerdaten zu speichern und eine Doppeldokumentation zu vermeiden. Daten werden mit Analysesoftware und externen

Tools verarbeitet und visualisiert, um Ärzten einen Überblick zu verschaffen und die Versorgung zu verbessern. Datenschutzmaßnahmen und ein Rollen-Rechte-System gewährleisten einen sicheren Umgang mit sensiblen Daten. Ziel ist die Identifizierung innovativer Biomarker zur Entwicklung neuer Therapien und unterstützende Tools für Therapieentscheidungen. Weitere Datenquellen wie Fragebögen, digitale Erfassung und potenziell Wearables könnten zukünftig integriert werden, um die Forschung und Patientenversorgung zu erweitern.

Im Kapitel Nutzende werden die verschiedenen Stakeholder DATACARE Architektur definiert. Die DATACARE-App richtet sich an Patient:innen, die ihre Behandlung und Datenverwaltung über die App nutzen können. Sie haben das Recht, ihre Betroffenenrechte wahrzunehmen. Ärzt:innen verwenden das Kliniksystem des DATACARE-Systems zur Behandlung und können verschiedene unterstützende Tools nutzen, die auch auf Daten aus weiteren Quellen zugreifen. Forschende erhalten Daten zur sekundären Nutzung und arbeiten mit einer vertrauenswürdigen Stelle zusammen, die sicherstellt, dass nur datenschutzkonform freigegebene Daten verwendet werden. Die Industrie, einschließlich Versicherungen und Pharmaunternehmen, kann ebenfalls auf Daten zugreifen, jedoch nur mit entsprechender Erlaubnis, wie einer Einwilligung. Die vertrauenswürdige Stelle sorgt auch hier für datenschutzkonforme Freigaben.

Kapitel 4 beschreibt den Demonstrator und die verschiedenen Aspekte der Entwicklung. Die Erarbeitung und der Entwicklungsprozess des Demonstrators wird detailliert in Abschnitt 3.3 dieses Abschlussberichts beschrieben.

Das Rahmenwerk geht zusätzlich auf die technisch juristische Analyse ein. Die technisch-juristische Analyse zur Einwilligung in der DATACARE-App basiert auf Art. 8 Abs. 2 Satz 2 der Europäischen Grundrechtecharta (GRCh) und Art. 4 Nr. 11 DSGVO. Beide Normen definieren die Einwilligung als freiwillige, informierte und zweckgebundene Willensbekundung der betroffenen Person. Die Verarbeitung personenbezogener Daten muss dabei nach Treu und Glauben sowie für festgelegte Zwecke erfolgen. Verschiedene Einwilligungsmodelle wie „Broad Consent“ und „Dynamic Consent“ bieten unterschiedliche Ansätze: Während „Broad Consent“ die Datenfreigabe für allgemeine Forschungszwecke erlaubt, bietet „Dynamic Consent“ eine flexible, fortlaufende Zustimmung mit der Möglichkeit zur Anpassung, was mehr Kontrolle und Transparenz für die Datengeber:innen schafft.

Die DATACARE-App setzt auf den „Dynamic Consent“, um die Datensouveränität und Selbstbestimmung der Nutzer:innen zu stärken. Dieses Modell ermöglicht es, die Einwilligung auch nachträglich an neue Verarbeitungszwecke anzupassen. Ein Opt-Out-Modell wird abgelehnt, da es die Anforderungen an Freiwilligkeit gemäß DSGVO nicht erfüllt. Wichtig ist zudem, dass der Widerruf der Einwilligung genauso einfach gestaltet ist wie ihre Erteilung, um Vertrauen und Nutzerfreundlichkeit zu gewährleisten. Dieses dynamische Einwilligungsmanagement unterstützt eine rechtssichere und transparente Handhabung der Nutzerdaten.

Darüber hinaus wird ein möglicher KI Einsatz in der App skizziert. Der Demonstrator simuliert den Einsatz einer KI zur automatisierten Datenvorauswahl für Arzttermine und Studien. Die App analysiert unstrukturierte Daten aus der elektronischen Patientenakte (ePA), extrahiert relevante Informationen und ordnet sie passenden Kategorien oder Labels zu. Für Studien gleicht die KI Einschlusskriterien mit Patientendaten ab und zeigt geeignete Optionen oder fehlende Informationen an. Transparenz und Kontrolle werden durch einsehbare Freigaben und anpassbare KI-Entscheidungen gewährleistet.

Das Konzept der Responsible AI stellt sicher, dass die KI fair, nachvollziehbar und ethisch handelt. Sorgfältig annotierte Trainingsdaten und Vergleichstests mit manuellen Prozessen minimieren Bias und stärken das Vertrauen. Feedback-Mechanismen fördern eine kontinuierliche Anpassung an Nutzerbedürfnisse, während Selbstbestimmung im Fokus bleibt.

Das letzte inhaltliche Kapitel beschreibt die Forschungsschnittstelle der DATACARE Architektur. Die Forschungsschnittstelle der DATACARE Architektur nutzt attributsbasierte Zugriffskontrolle, um die Einwilligungen der Betroffenen technisch umzusetzen.

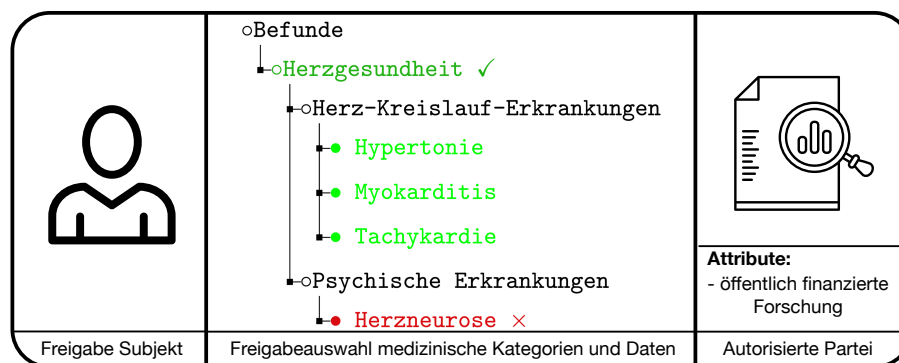


Abbildung 3.2.2: Schematische Darstellung einer attributsbasierten Zugriffskontrollrichtlinie

Abbildung 3.2.2 zeigt, wie Attribute genutzt werden, um die freizugebenden Daten, medizinischen Kategorien und autorisierten Parteien festzulegen. Die Zugriffskontrollrichtlinie wird dynamisch auf Basis der in der DATACARE App hinterlegten Einwilligungen erstellt und gewährleistet, dass nur die von den Betroffenen explizit freigegebenen Daten für Anfragen zugänglich sind.

Eine vertrauenswürdige Stelle übernimmt die Verifikation von Forschungsanfragen, die Messung des Re-Identifikationsrisikos und die korrekte Umsetzung der Einwilligungen. Datenschutzwahrende Technologien wie Differential Privacy und eine umfassende Risikoanalyse unterstützen die Betroffenen bei informierten Entscheidungen. Diese Kombination aus innovativen Ansätzen gewährleistet sowohl den Schutz sensibler Daten als auch die effiziente Nutzung für Forschungszwecke.

### 3.3 Demonstrator

Der Demonstrator sollte im Rahmen einer nutzerzentrierten, iterativen Entwicklung und durch Workshops entwickelt werden. Für eine nutzerzentrierte Anwendung, die die Bedürfnisse von Betroffenen in den Vordergrund stellt, sind Workshops, in denen die Anforderungen erhoben werden, unerlässlich. In diesen Workshops wird in verschiedenen Aufgaben zusammen mit den Stakeholdern Anforderungen für die iterative Entwicklung erhoben. Dies bedeutet, dass die verschiedenen Entwicklungsschritte immer wieder zyklisch durchgeführt werden. Das bedeutet üblicherweise zuerst die Erfassung der Nutzeranforderungen auf Basis des aktuellen Entwicklungsstandes und auf Basis dieser Anforderungen die weitere Entwicklung des Demonstrators. Dieser Zyklus wird dann wiederholt, bis der finale Prototyp erreicht ist.

Des Weiteren wurden App Konzepte vorgestellt, die am ehesten der Anforderungen der Zielgruppe entspricht. Durch ein App Konzept können Betroffene Daten jederzeit einsehen auf der Plattform, die vermutlich die meiste Zeit des Tages verwendet wird. Durch eine modulare Konzeption kann diese App auch auf verschiedene Use-Cases angepasst werden.

Im Rahmen dieses Entwicklungskonzeptes wurde ein Workshop durchgeführt, der zur Definition der Use-Cases und des Szenarios des DATACARE Demonstrators führte.

Der Demonstrator setzt folgendes Szenario um: *„Annegret de Vries, 45 Jahre alt hat kürzlich einen schweren Rheumaschub erlitten. Nun steht wieder der halbjährliche Termin beim Rheumatologen an. Hier möchte sie diesen Progress besprechen. Sie verfügt über digitale Grundkenntnisse. Wie üblich muss sie erfasste Daten und Daten, die an anderer Stelle erfasst worden sind, zusammentragen und für ihren Kontrolltermin vorbereiten.*

*Dies kann sie in der HealthHub App durchführen. Die App unterstützt Annegret beim Auswählen und Freigeben der Daten für ihren Rheumatlog:in. Zusätzlich kann die App anzeigen, für welche klinischen Studien Annegrets Rheumatlog:in Patienten rekrutiert. Für bestimmte Studien werden direkt bestimmte Daten angefragt die Annegret auf Wunsch mit der Studie zur Verfügung stellen kann. Hierfür wird Sie ebenfalls von der App unterstützt, in dem diese den Wert der freizugebenden Daten bemisst.“*

Daraus leiten sich folgende Funktionen ab

1. **Ansicht medizinischer Daten:** Nutzer:innen können ihre medizinischen Daten einsehen.
2. **Datenfreigabe:** Daten können an Rheumatolog:innen und Forschungsprojekte weitergegeben werden.
3. **Empfehlungen zur Datenfreigabe:** KI-gestützte Vorschläge zeigen, welche Daten für bestimmte Zwecke freigegeben werden sollten.
4. **Quantifizierung der Daten:** Der (monetäre) Wert der Daten für die Forschung wird geschätzt.

5. **Nachvollziehbarkeit:** Eine rudimentäre Anzeige gibt Einblick, welche Daten freigegeben wurden.

### Zentrale Szenarien der Datenfreigabe:

- **Für Arzttermine:** Daten werden für einen Kontrolltermin vorbereitet. Dies schließt die Möglichkeit ein, gleichzeitig Daten für Forschungsvorhaben freizugeben.
- **Als Forschungsdatenspende:** Freiwillige Datenfreigabe für spezifische Forschungsprojekte, mit Auswahlmöglichkeiten nach Kategorien oder einzelnen Datenpunkten.

### Datenschutz und Kontrolle:

- Es wird ein **Opt-In-Verfahren** verwendet, bei dem Nutzer:innen aktiv entscheiden, welche Daten freigegeben werden.
- Ein **Empfehlungsmechanismus** unterstützt die Auswahl basierend auf Zweck, Zeitraum oder Datenart.
- **Risikobewertungen** machen die potenziellen Auswirkungen der Freigabe verständlich.
- 

Die Gestaltung des Prototyps folgt Designprinzipien, die Nachvollziehbarkeit und Kontrolle für die Nutzer:innen gewährleisten.

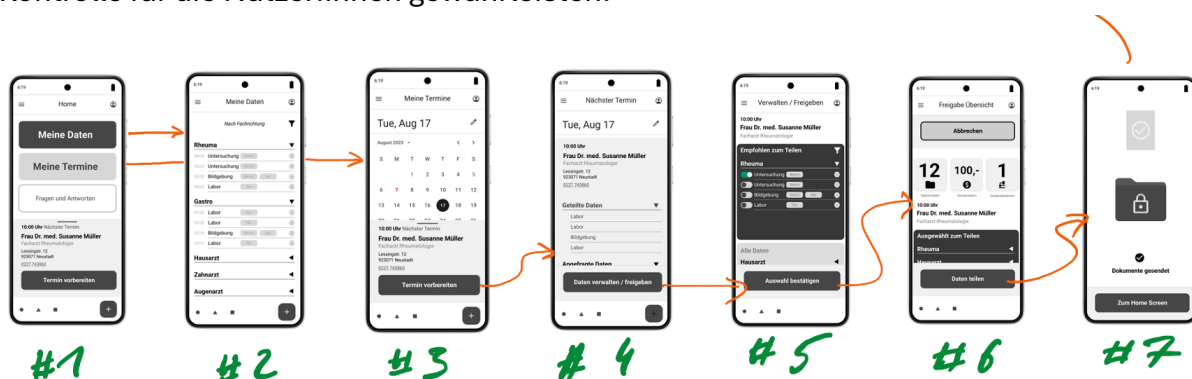


Abbildung 3.3.1: Design Prototypen mit schematischem Ablauf der Funktionen

Im iterativen Entwicklungsprozess wurden erste Design Entwürfe für den Demonstrator erstellt. Ein Beispiel für die Design Entwürfe zeigen die Mockups aus Abbildung 3.3.1. Diese Entwürfe wurde gemeinsam zur Diskussion gestellt und auf Basis des Feedbacks iteriert. Daraus resultierte zum Jahresende 2023 ein finales Set an Entwürfen auf deren Basis der Demonstrator angepasst wurde.

Der Demonstrator selbst wurde in einem iterativen Design Entwicklungsprozess stetig innerhalb der Projektlaufzeit vorangetrieben. Zum Ende des Jahres 2023 existierte beispielsweise noch ein größeres Delta zwischen technischem Demonstrator und Designentwürfen. Dieser Unterschied wurde dann für die Evaluation ausgeglichen und der Demonstrator auf denselben Stand gebracht.

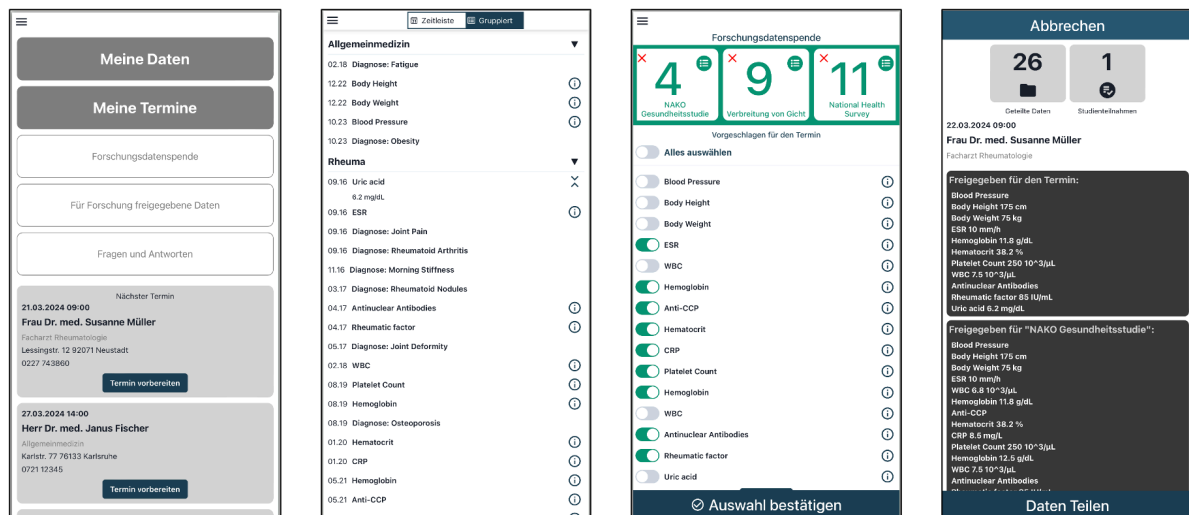


Abbildung 3.3.2: Screenshots aus einer technischen Iteration auf Basis der Design Entwürfe

Auf technischer Ebene ist der Demonstrator eine Webanwendung, die in Typescript<sup>16</sup> implementiert ist. Dadurch kann der Demonstrator grundsätzlich auf allen Geräten eingesetzt werden, auf denen ein moderner Webbrowser verfügbar ist, unter anderem auf Smartphones, Tablet PCs und Desktops. Das Layout des Demonstrators ist für Smartphones optimiert, um eine einfache Bedienung auch auf kleinen Bildschirmen und über Toucheingaben zu ermöglichen. Außerdem erfüllt der Demonstrator die Anforderungen an eine Progressive Web App (PWA). Dadurch kann das Verhalten von nativen Apps imitiert werden, zum Beispiel indem sich die Anwendung über eine Verknüpfung auf dem Home Screen starten lässt.

Abbildung 3.3.2 zeigt Screenshots aus einer technischen Iteration des Demonstrators. Die Implementierung basiert auf dem UI-Framework React-JS<sup>17</sup>, das die Aufteilung der Benutzeroberfläche in Komponenten und die Verwaltung von UI-Zuständen unterstützt. Auf architekturnaler Ebene ist die Implementierung der Benutzeroberfläche entsprechend den Anwendungsfällen in verschiedene Seiten unterteilt. Diese Seiten sind wiederum hierarchisch in Unterkomponenten aufgeteilt.

Im Back End nutzt der Demonstrator das FHIR-Format<sup>18</sup>, mit dem Gesundheitsdaten maschinenlesbar kodiert werden können. Die Daten werden über eine Webschnittstelle durch die FHIR-Serversoftware des HAPI<sup>19</sup> Open Source-Projekts bereitgestellt.

Die vorliegenden Demonstrator Daten werden durch ein mit einem Large Language Model generierten Skript im FHIR-Format erstellt, basierend auf einem umfangreichen Referenzdatensatz für medizinische und demografische Informationen aus der 4D Klinik. Dieser Referenzdatensatz enthält eine Liste von Variablen, die mögliche Werte und deren zulässige Bereiche vorgeben.

<sup>16</sup> <https://www.typescriptlang.org>

<sup>17</sup> <https://react.dev>

<sup>18</sup> <http://hl7.org/fhir>

<sup>19</sup> <https://github.com/hapifhir/hapi-fhir>

Der generative Prozess beginnt mit der Erstellung von Patientendaten, wobei der Referenzdatensatz als Leitfaden für mögliche Werte und Wertebereiche dient. Persönliche Informationen wie Name, Geburtsdatum, Geschlecht, Kontaktinformationen und demografische Details werden berücksichtigt.

Die medizinischen Behandlungsdaten, einschließlich Bedingungen wie rheumatoide Arthritis, Medikamentenangaben und Laborergebnisse, werden ebenfalls aus der Vorlage des Referenzdatensatzes generiert. Ein Python-Skript wurde entwickelt, um beliebig viele Daten zu erstellen. Aus dieser Datenmenge wird der passendste Beispielpatient ausgewählt, der für den Demonstrator verwendet wird.

Die generierten Daten werden zu einem FHIR-Transaktionsbündel zusammengeführt, das verschiedene Ressourcen wie Patient, Bedingung, Medikamentenaussage und Beobachtungen enthält. Dieses Bündel ermöglicht CRUD-Operationen auf einem FHIR-Server und ist strukturiert, um den Vorgaben des Referenzdatensatzes zu entsprechen.

Der Zweck der Datenverwendung liegt ausschließlich in der Evaluation eines Prototyps. Weitere Zwecke, insbesondere solche, die die Privatsphäre oder den Datenschutz beeinträchtigen könnten, sind ausgeschlossen. Die Orientierung am Referenzdatensatz gewährleistet eine hohe Qualität und Realitätsnähe der Daten für Evaluationszwecke.

Im Rahmen der Projektpublikation „Data sovereignty requirements for patient-oriented AI-driven clinical research in Germany“ wurden der Entwicklungsprozess als technologische Perspektive: Benutzerfreundlichkeit und Interaktionsdesign für Software im Gesundheitswesen beschrieben. Das folgende fasst diesen Beitrag zusammen:

Die Gestaltung von Benutzeroberflächen für Anwendungen im Gesundheitswesen ist eine komplexe Designaufgabe. Anwendungen im Gesundheitswesen können das gesamte Spektrum von hochspezialisierter Expertensoftware bis hin zu weit verbreiteter Verbrauchersoftware abdecken, was zu völlig unterschiedlichen Anforderungen führt, z. B. in Bezug auf die Einhaltung von Vorschriften und die Erwartungen der Nutzer. Wie Software im Allgemeinen sollten jedoch beide unter dem Gesichtspunkt der Menschenzentrierung konzipiert und entwickelt werden. Generell müssen wir zwischen Schnittstellen im Gesundheitswesen als Forschungsobjekt der Mensch-Computer-Interaktion (HCI) (Wissenserzeugung) und einem Software-Engineering-Problem (Wissensanwendung) unterscheiden. Hier konzentrieren wir uns auf Letzteres. Wir lassen technische Anforderungen wie Interoperabilität beiseite und fokussieren uns auf den zweiten, ebenso wichtigen Erfolgsfaktor von Software im Gesundheitswesen, nämlich die Benutzerfreundlichkeit (Usability).

Die Gebrauchstauglichkeit eines interaktiven Systems ist definiert als das Ausmaß, in dem bestimmte Benutzer ihre spezifischen Ziele in definierten Nutzungskontexten effizient, effektiv und zufriedenstellend erreichen können (ISO 9241-210<sup>20</sup>). Die Prinzipien der systematischen Erstellung von benutzbarer Software sind gut bekannt. Allerdings spielen diese Prinzipien in Softwareentwicklungsprozessen oft leider eine untergeordnete

---

<sup>20</sup> ISO 9241-210. Ergonomics of Human-System Interaction–Part 210: Human-Centred Design for Interactive Systems.

Rolle. Die Synchronisierung von Software-Engineering-Prozessen mit Human-centered Design bleibt eine Herausforderung auf operativer Ebene.

Wachter<sup>21</sup> und Gawande<sup>22</sup> beschreiben, dass Software für das Gesundheitswesen besonders unter mangelnder Benutzerfreundlichkeit zu leiden scheint, und es kann hinzugefügt werden, dass sie wenig Wert auf modernes Schnittstellendesign legt. Wachter weist darauf hin, dass "Ärzte in der Notaufnahme 44 Prozent ihrer Zeit mit der Eingabe von Daten in elektronische Patientenakten verbringen und während einer 10-Stunden-Schicht bis zu 4.000-mal klicken". Ihre Argumente werden durch wissenschaftliche Untersuchungen gestützt. Im Weiteren wird aufgezeigt, was der Stand der Technik in Bezug auf Design und Ingenieurwissen sein sollte, um Software für das Gesundheitswesen mit angemessener Benutzerfreundlichkeit herzustellen.

Das Dokument „ISO 9241-210 Ergonomie der Mensch-System-Interaktion: Menschenzentriertes Design“ stellt die Terminologie bereit, die notwendig ist, um auf benutzbare Software hinzuwirken, und beschreibt den menschenzentrierten Designprozess und seine Aktivitäten. Die wichtigsten Punkte sind:

„Menschenzentriertes Design ist ein Ansatz für Systemdesign und -entwicklung, der darauf abzielt, interaktive Systeme benutzerfreundlicher zu gestalten, indem er sich auf die Nutzung des Systems konzentriert und Wissen und Techniken aus den Bereichen Human Factors/Ergonomie und Benutzerfreundlichkeit anwendet.

Das Dokument „ISO 9241-110 Ergonomie der Interaktion zwischen Menschen und System: Grundsätze des Dialogs“ enthält Leitlinien zu den grundlegenden Qualitäten von benutzbarer Software. Diese Best Practices werden als Dialog- oder Interaktionsprinzipien bezeichnet. Software, die die folgenden grundlegenden Kriterien erfüllt, kann als benutzbare Software angesehen werden:

- Eignung für die Aufgabe des Nutzers
- Selbstbeschreibungsfähigkeit
- Erlernbarkeit
- Beherrschbarkeit
- Robustheit gegenüber Benutzerfehlern
- Engagement des Benutzers

### 3.4 Nutzendenstudie

Um die Benutzbarkeit und Akzeptanz der App zu evaluieren wurde im ersten Halbjahr 2024 eine Nutzendenstudie durchgeführt. Hierfür wurde auf die Vertreterinnen der Rheumaliga innerhalb des Projektes zugegriffen. Diese haben innerhalb der Patientinnenorganisation verschiedene Probandinnen rekrutiert.

---

<sup>21</sup> Wachter, R (2017). Digital Doctor: Hope, Hype, and Harm at the Dawn. McGraw-Hill Education. Reprint Edition.

<sup>22</sup> Gawande A, (2018). Why doctors hate their computers. The New Yorker. November 12 2018 Issue

Konkret wurde die Nutzendenstudie mit 6 Personen durchgeführt. Dabei waren 66% männlich und 24% weiblich. Die Altersspanne bewegte sich zwischen 55 und 75 Jahren. Hierbei besaß die Mehrheit der Probandinnen einen Hochschulabschluss (>65%). Alle Personen nutzen täglich Smartphone und PCs. Darüber hinaus war bei allen Probandinnen Erfahrung im Bezug zu Themen aus dem Gesundheitsbereich vorhanden (unter anderem durch die ehrenamtliche Tätigkeit innerhalb der Rheumaliga). Zum Thema Datenschutz gab allerdings nur eine Person an Kompetenzen zu besitzen.

Der Ablauf der Studie war wie folgt:

1. Einweisung (Briefing)
2. Ausfüllen eines Fragebogens zur Person
3. Bearbeiten von Aufgaben mit Prototyp
4. Fragebogen
5. Optionales Interview
6. Verabschiedung (De-Brief)

Als Dauer wurden ca. 50min pro Durchlauf veranschlagt. Die Probandinnen sollten innerhalb dieser Zeit drei Kernfunktionen der App testen. Die Durchführung erfolgt aufgrund der breiten Verteilung der Probandinnen innerhalb Deutschlands remote. Hierbei wurden die Probandinnen angehalten die Studie nach dem Think Aloud Verfahren durchzuführen. Dazu sollen die Teilnehmenden alle Gedanken, die sie bei der Durchführung der Studie haben, laut aussprechen.

Technisch wurde der Demonstrator auf einem Serversystem innerhalb der KIT Infrastruktur projektintern zur Verfügung gestellt. Hierdurch konnten die Probandinnen den Demonstrator mit ihrem eignen Smartphone über einen Weblink im Browser abrufen. Dieser Studienablauf erwies sich als sehr gut geeignet. Die Systematik für diesen Studienablauf wurde bereits im Jahr 2023 durch die Publikation „*Usability for Data Sovereignty - Evaluation of Privacy Risk Quantification Interfaces*“<sup>23</sup> erarbeitet. Diese Publikation evaluiert in einer Nutzerstudie verschiedene Nutzerinterfaces, die eine Privatsphärenrisikoquantifizierung darstellen. Diese Interfaces sind in Teilen verwandt mit der DATACARE App, werden so aber nicht eins zu eins umgesetzt.

Als Studienszenario wurde, angenommen, dass die Probandinnen den Demonstrator auf ihrem Smartphone nutzen und die darin vorhandenen Beispieldaten beispielsweise aus der elektronischen Patientenakte stammen. Den Probandinnen wurde mitgeteilt, dass sie Zugriff auf neuartiges Tool zur Vorbereitung von Arztterminen und Forschungsfreigabe von Daten erhalten.

---

<sup>23</sup> Arno Appenzeller, Falk Balduf, and Jürgen Beyerer. 2023. Usability for Data Sovereignty - Evaluation of Privacy Risk Quantification Interfaces. In Proceedings of the 16th International Conference on Pervasive Technologies Related to Assistive Environments (PETRA '23). Association for Computing Machinery, New York, NY, USA, 206–214.

Zusätzlich wurde den Teilnehmenden folgende Eigenschaften zu ihrer Rolle angegeben:

- Die elektronischen Patientenakte enthält eine signifikante Anzahl an Einträgen (Diagnosen, verschriebene Medikamente, Laborergebnisse, Scans, ...). Die Einträge beziehen sich auf vergangene und aktuelle Erkrankungen und Therapien.
- Unter den Einträgen der elektronischen Patientenakte befinden sich auch Daten, die persönlich als vertraulich eingestuft sind.
- Die App bietet die Möglichkeit bevorstehende Arzttermine vorzubereiten und Daten und Werte den Ärzt:innen zur Verfügung zu stellen oder diese darauf aufmerksam zu machen. Im Rahmen dieser Terminvorbereitung können Daten für diverse vorgeschlagene Studien freigegeben werden.
- Zusätzlich bietet die App die Möglichkeit Daten gesondert für die Forschung freizugeben

Folgende Aufgaben in drei Kategorien sollten von den Probandinnen bearbeitet werden:

#### **Datenübersichts Modul**

1. Wählen Sie die Datenübersicht an.
2. Lassen Sie sich die Daten in gruppierter Form anzeigen
3. Blenden Sie alle Daten außer den Daten zu Rheuma aus
4. Welcher CRP Wert wurde am 22.01.20 gemessen?  
Bewerten Sie die Funktion zur Anzeige der Detailansicht.
5. Vergleichen Sie die Zeitleiste und die gruppierte Ansicht.
6. Wie bewerten Sie die Ansicht?

#### **Grundfunktionen Modul**

1. Verschaffen Sie sich eine Übersicht der bevorstehenden Termine
2. Betrachten Sie die Kalender Funktion und die bevorstehenden Termine dort
3. Wann stehen die nächsten Termine an?
4. Bereiten Sie die Daten für bevorstehenden Termin beim Rheumatologen vor.

#### **Forschungsdatenfreigabe Modul**

1. Wählen Sie den Menüpunkt für Forschungsfreigabe an
2. Informieren Sie sich über die angebotenen Forschungsprojekte
3. Wählen Sie die Forschungsprojekte, für die Sie Daten freigeben möchten
4. Nehmen Sie die Datenfreigabe vor und bestätigen Sie die Auswahl
5. Analysieren Sie die Datenschutzbewertung der Freigaben
6. Ändern Sie die Auswahl der Forschungsprojekte oder senden Sie die Daten
7. Betrachten Sie nach der erfolgreichen Freigabe ihre für die Forschung freigegebenen Daten
8. Ziehen Sie eine Freigabe zurück falls gewünscht.

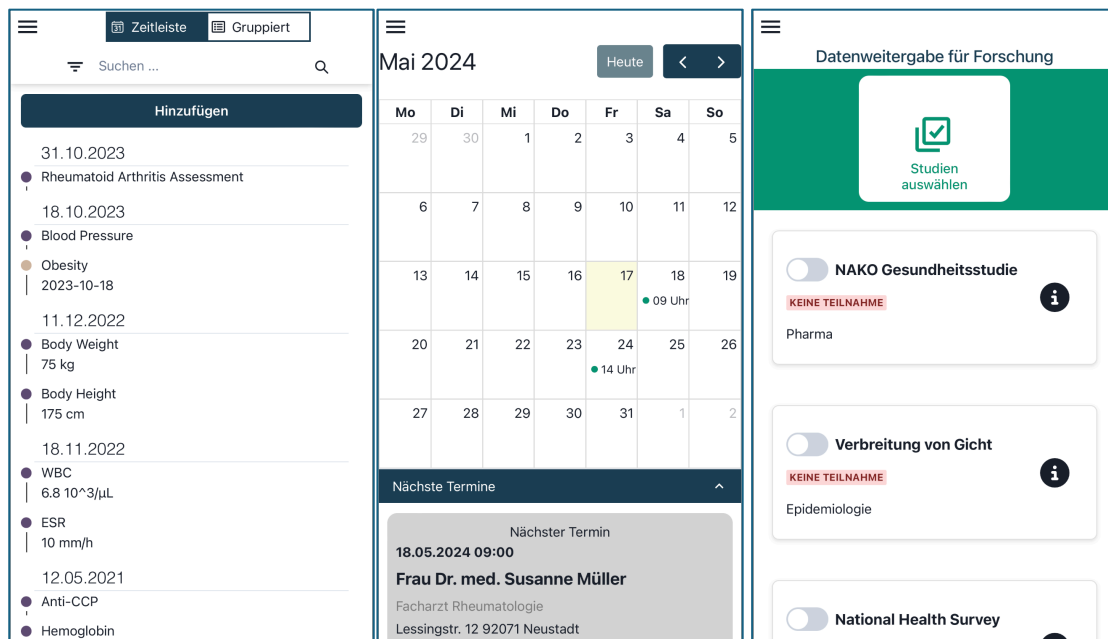


Abbildung 3.4.1: Exemplarische Ansicht zu den verschiedenen Aufgabekategorien (in der Reihenfolge der Auflistung von links nach rechts).

Die Studienergebnisse teilen sich in quantitative und qualitative Ergebnisse auf. Für die quantitativen Ergebnisse wurde ein standardisierter Fragebogen auf Basis der System Usability Scale (SUS)<sup>24</sup> zur Messung der Nutzbarkeit verwendet.



Abbildung 3.4.2: Ergebnisse der SUS Fragebögen.

Als gute Usability wird ein Wert größer 70 bezeichnet, der in diesem Fall mit dem Durchschnitt von 72 erreicht wird. Somit kann die Nutzbarkeit des Demonstrators als gut beschrieben werden. Allerdings gilt die geringe Teilnehmendenzahl zu beachten, die die Aussagekraft der quantitativen Ergebnisse einschränkt.

Bei kleineren Teilnehmendenzahlen sind qualitative Ergebnisse aussagekräftiger, die in dieser Studie durch das Think-Aloud Feedback und die Interviewfragen gesammelt worden sind.

Beim Think-Aloud zeigte sich Optimierungsbedarf in der Benutzerfreundlichkeit und Verständlichkeit. Interaktionen wie Kacheln wischen oder Schalter umlegen wurden oft als nicht intuitiv empfunden, und die Navigation, besonders im Browser, führte zu

<sup>24</sup> Usability Evaluation In Industry, von Patrick W. Jordan, B. Thomas, Ian Lyall McClelland, Bernard Weerdmeester, London: Taylor and Francis, 1996

Verwirrung. Barrierefreiheit war ein weiteres Problem, da die App mit Anzeigezoom und Schriftgrößenanpassungen schlecht zurechtkam. Die Datenschutzbewertungen und Projektinformationen wurden als zu technisch wahrgenommen, was ein Onboarding und klare Begriffserklärungen notwendig macht. Die Nutzerinnen wünschten sich eine verständlichere Darstellung und betonten die Wichtigkeit von Transparenz über die Datenverwendung.

In dem Abschlussinterview wurde Fragen zur App allgemein, der Nutzung von KI und zur Datenschutzbewertung innerhalb der App gestellt. Die Interviews zur DATACARE App zeigten, dass die Arztterminvorbereitung in Kombination mit passenden Datenquellen als nützlich empfunden wurde. Gleichzeitig empfanden einige Nutzer:innen die App als unübersichtlich und hatten Schwierigkeiten, Funktionen intuitiv zu finden.

Die Nutzung von KI wurde positiv aufgenommen, insbesondere die Vorschläge zur Vorbereitung von Arztterminen. Wichtig war den Teilnehmenden jedoch, dass die KI transparent, sicher und nachvollziehbar arbeitet. Es wurde bevorzugt, dass die Datenverarbeitung lokal auf dem Gerät stattfindet.

Beim Thema Datenschutz äußerten sich einige Teilnehmende besorgt über die Weitergabe von Daten, insbesondere an Forschungsprojekte. Während Datenschutz unterschiedlich priorisiert wurde, wünschten sich alle eine verständlichere Erklärung der Bewertungen und deren Bedeutung, um die Informationen auch für Laien zugänglich zu machen.

Zusammenfassend zeigt die Studie eine insgesamt positive Bewertung der Usability des DATACARE Demonstrators, mit einem SUS-Wert von über 70. Dennoch wurden Verbesserungspotenziale bei der Navigation und Steuerung der App sowie bei der Erklärung der Datenschutzhinweise identifiziert. Besonders die Funktion zur Arztterminvorbereitung mit KI-gestützten Vorschlägen wurde als äußerst hilfreich wahrgenommen. Zudem äußerten die Nutzer:innen eine grundsätzliche Bereitschaft zur Freigabe von Forschungsdaten, die in Kombination mit der Terminvorbereitung sogar noch gesteigert wurde.

### 3.5 Fazit

Im Rahmen des Projekts wurde in interdisziplinärer Zusammenarbeit erfolgreich ein technisch-juristisches Rahmenwerk für eine medizinische Datenplattform mit Patientinnenmitembeziehung entworfen. Erkenntnisse aus diesem Prozess sind in diverse wissenschaftliche Publikationen miteingeflossen.

Das Rahmenwerk diente in einem iterativen nutzerzentrierten Prozess den Projekt-Demonstrator zu entwickeln. Dieser existiert in Form einer mobilen Applikation mit dem Betroffene ihre persönlichen medizinischen Daten einsehen, verwalten und für die Forschung verfügbar machen können.

Der Demonstrator wurde durch Probandinnen der Patientenvertretung der Rheumaliga innerhalb des Projekts evaluiert. Die Studie mit einer Fokusgruppe aus möglichen echten Nutzenden der Anwendung liefert wertvolle qualitative Erkenntnisse für die weitere

Entwicklung ähnlicher Verfahren und zeigt, dass die Ansätze der App eine gute Nutzbarkeit bieten.

Das Rahmenwerk und die modulare Architektur des Demonstrators ermöglicht eine Verwendung beziehungsweise die Erweiterung von Teilen oder auch des ganzen Systems in ähnlichen Nachfolge- oder auch neuen Projekten.

## 4 Voraussichtlicher Nutzen & Verwertbarkeit der Ergebnisse

Der entwickelte Demonstrator bietet eine gute technische Grundlage für weitere ähnlich gelagerte Forschungsarbeiten. Dies gilt vor allem in Bezug auf die weitere Entwicklung und Präzisierung des Demonstrators und für weitere Nutzbarkeitsstudien.

Das entwickelte technisch-juristische Konzept kann auch in zukünftigen Projekten als Grundlage dienen.

Somit können die im Projekt erworbenen Erfahrungen in weitere Forschungs- oder auch Industrieprojekte einfließen.

## 5 Fortschritt auf diesem Gebiet bei anderen Stellen

Während der Projektlaufzeit haben sich vor allem Fortschritte auf im Bereich der Verbreitung und Nutzungsgrundlage der elektronischen Patientenakte ergeben.

So ist nach dem Softlaunch der ePA nach Einwilligung zu Beginn 2021 inzwischen Gesundheitsdatennutzungsgesetz (GDNG)<sup>25</sup> beschlossen worden. Dieses ermöglicht die ePA für alle und sieht vor allem eine Widerspruchslösung vor. Das heißt die ePA wird automatisch für jede gesetzlich versicherte Person angelegt<sup>26</sup>. Neben dem reinen Anlegen der ePA ist zukünftig auch eine Nutzung der Daten der ePA zu sekundären Zwecken wie beispielsweise innerhalb des Forschungsdatenzentrum Gesundheit vorgesehen. Die Details zu dieser Weitergabe waren aber zum Zeitpunkt, zu dem dieser Bericht verfasst worden ist, noch nicht bekannt.

Zum einen kann dies als politische Richtungsentscheidung in Richtung von Widerspruchslösungen gewertet werden zum anderen ist allerdings die Entscheidung über die Erstellung der ePA anders gelagert als die Datenweitergabe aus dieser.

Eine weitere Entwicklung ist die weitere Verfügbarkeit des Forschungsdatenzentrum Gesundheit, das verschiedene Gesundheitsdaten der Bevölkerung im Laufe des Projekts für die forschende Öffentlichkeit weiter zugänglich gemacht hat. Diese Entwicklung kann auch ergänzenden zu den Bestrebungen innerhalb des DATACARE Projektes gesehen werden.

Auch andere Projekte wie der europäische Gesundheitsdatenraum EHDS sind in der Projektlaufzeit weiter vorangeschritten. So wurde für den EHDS im Jahre 2024 eine politische Einigung<sup>27</sup> erzielt und auch die deutsche Medizin Informatik Initiative ist an dem Vorhaben beteiligt<sup>28</sup>.

Zusammenfassend sind vor allem die Entwicklungen hinsichtlich der ePA und die weiteren Bestrebungen Forschungsdaten verfügbar zu machen, ein Zeichen für die gesellschaftliche Relevanz der Themen, die in DATACARE beforscht worden sind. Auch wenn viele der Entwicklungen etwaige Ansätze aus DATACARE überholen könnten, ist vor allem zu betrachten, dass Konzepte wie die Datenspende und das nutzerzentrierte datenschutzfreundliche Einwilligungsmanagement potenzielle Erweiterungen und Synergien zu bestehenden Ansätzen und Technologien wie der ePA bieten kann.

---

<sup>25</sup> <https://www.bundesgesundheitsministerium.de/ministerium/gesetze-und-verordnungen/guv-20-lp/gesundheitsdatennutzungsgesetz.html>

<sup>26</sup> <https://www.bundesgesundheitsministerium.de/presse/pressemitteilungen/bundestag-verabschiedet-digitalgesetze-pm-14-12-23.html>

<sup>27</sup> [https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space\\_de](https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_de)

<sup>28</sup> <https://www.medizininformatik-initiative.de/de/medizininformatik-initiative-legt-grundlagen-fuer-ehds>

## 6 Veröffentlichungen

### Veröffentlichungen:

Titel	Datum	Link
Datenschutzkonforme Weitergabe von Versichertendaten aus dem Forschungsdatenzentrum	26.09.22	<a href="https://dspace.gi.de/handle/20.500.12116/39545">https://dspace.gi.de/handle/20.500.12116/39545</a>
Usability for Data Sovereignty - Evaluation of Privacy Risk Quantification Interfaces	06.07.23	<a href="https://dl.acm.org/doi/10.1145/3594806.3594816">https://dl.acm.org/doi/10.1145/3594806.3594816</a>
Das Re-Identifikationsrisiko bei der Weitergabe von Gesundheitsdaten	01.05.24	<a href="https://link.springer.com/article/10.1007/s11623-024-1930-1">https://link.springer.com/article/10.1007/s11623-024-1930-1</a>

### Vorträge:

Titel	Datum
Datenschutzkonforme Weitergabe von Versichertendaten aus dem Forschungsdatenzentrum	26.09.22
Usability for Data Sovereignty - Evaluation of Privacy Risk Quantification Interfaces	06.07.23

*Hinweis: Die hier aufgelisteten Veröffentlichungen sind aus Gründen der Redundanz nicht um die Veröffentlichungen des Gesamtkonsortiums erweitert. Für diese Veröffentlichungen und Vorträge sei auf die entsprechenden Partner Abschlussberichte vor allem der Projektleitung unter Fraunhofer Gesellschaft verweisen.*