VDI/VDE Innovation + Technik GmbH
Steinplatz 1, 10623 Berlin

Data In Motion Consulting GmbH
Kahlaische Str. 4, 07745 Jena

Abschlussbericht

zum Teilvorhaben

Nutzung verteilter, dynamische, modulare Systemarchitektur für Multi-Modell-basierte Anonymisierung und Privacy-Enhancement Ansätze im Bereich Gesundheitsdaten

im Verbundprojekt

Anonymisierung persönlicher Gesundheitsdaten durch Erzeugung digitaler Avatare in Medizin und Pflege - AVATAR -

Projektlaufzeit: 15.11.2022 - 31.12.2025
FKZ: 16KISA027
Datum: 26.01.2026

Mark Hoffmann

Inhalt

1. Vorbemerkung **3**
2. Kurzbericht **4**
2.1 Ursprüngliche Aufgabenstellung und wissenschaftlich-technischer Stand 4
2.2 Ablauf des Vorhabens 4
2.3 Wesentliche Ergebnisse und Zusammenarbeit 5
3. Eingehende Darstellung **6**
3.1 Überblick und Strukturierung der durchgeführten Arbeiten 6
3.2 Eingehende Darstellung der Teilprojektarbeiten 7
3.2.1 Teilprojekt 1 7
3.2.2 Teilprojekt 3 8
3.2.3 Teilprojekt 4 9
3.2.4 Teilprojekt 5 12
4. Erfolgskontrollbericht **14**
5. Kurzfassung zum Verwertungsplan **19**

1. Vorbemerkung

Der vorliegende Abschlussbericht orientiert sich in Aufbau und Inhalt an der Hinweisseite zum Verwendungsnachweis auf Kostenbasis AZK (https://www.vdivde-it.de/de/verwendungsnachweis-auf-kostenbasis-azk). Er enthält den Sachbericht (Kurzbericht und Eingehende Darstellung), den Erfolgskontrollbericht sowie die Kurzfassung zum aktuellen Verwertungsplan im Sinne der dort aufgeführten Nachweisbestimmungen.

2. Kurzbericht

2.1 Ursprüngliche Aufgabenstellung und wissenschaftlich-technischer Stand

Das Forschungsprojekt AVATAR hatte zum Ziel, eine technische Lösung zu entwickeln, die die Nutzung sensibler Gesundheitsdaten für die Forschung ermöglicht, ohne den Datenschutz der Patient:innen zu gefährden. Dies sollte durch die Erzeugung digitaler „Avatare“ geschehen – aggregierte Datensammlungen, welche die statistischen Eigenschaften der Realdaten bewahren, aber durch die Anwendung verschiedener Anonymisierungsverfahren keinen Rückschluss auf Einzelpersonen zulassen.

Die Data In Motion Consulting GmbH (DIM) übernahm im Projekt vornehmlich die Rolle der softwaretechnischen Architektur, Umsetzung und Integration. Die ursprüngliche Aufgabenstellung umfasste die Konzeption und Umsetzung einer verteilten, dynamischen und modularen Systemarchitektur. Zum Projektstart war der Stand der Technik durch monolithische oder container-basierte Systeme geprägt, die zwar Skalierbarkeit boten, jedoch bei dynamischen Abhängigkeiten und der technischen Interpretation komplexer rechtlicher Datenschutzvorgaben an ihre Grenzen stießen. Hier hat Data In Motion angesetzt („Model-Driven-Privacy“), um Interoperabilität zwischen heterogenen Teilsystemen sicherzustellen und ein Framework für die automatisierte Modellierung von komplexen Anonymisierungs- und Datensynthetisierungsprozessen zu schaffen.

2.2 Ablauf des Vorhabens

Das Vorhaben wurde von uns in drei wesentlichen Phasen durchgeführt:

- **Architektur- und Konzeptionsphase:** Kernbestandteil dieser Phase war die Schaffung von Grundlagen für eine föderierte Systemumgebung. Hierbei wurden Discovery-Mechanismen definiert, die es ermöglichen, Dienste und Datenmodelle systemübergreifend auffindbar zu machen.
- **Entwicklungs- und Implementierungsphase:** Im Fokus stand der Aufbau eines modularen Datenpools auf Basis des OSGi-Standards. Dafür wurde eine Model-Registry entwickelt und entsprechende Schnittstellen programmiert, um Gesundheitsdaten aus quasi beliebigen Quellen integrieren zu können. Darüber hinaus wurde, über den ursprünglichen Projektantrag hinausgehend, durch Data In Motion ein **komplettes UI (User Interface)** für die Frontenddatenabfrage an das Avatarsystem entwickelt, welches die Konfiguration komplexer Datenabfragen intuitiv ermöglicht.
- **Integrations- und Validierungsphase:** In der finalen Phase wurden die Teilmodule der Partner (v.a. Datenquellen der beteiligten Praxispartner) integriert. DIM unterstützte hierbei die Orchestrierung der Einzelkomponenten zu einem funktionsfähigen Gesamtsystem.

2.3 Wesentliche Ergebnisse und Zusammenarbeit

Die wesentlichen durch uns erreichten Ergebnisse des Avatarprojektes umfassen folgende Bereiche:

- **Standardbasierte Interoperabilität (HIMSA & FHIR):** Um eine nahtlose Integration in die klinische Praxis und bestehende Gesundheitsdatenverwaltungssysteme zu gewährleisten, hat Data In Motion dedizierte Schnittstellen für den internationalen, im klinischen Betrieb führenden FHIR-Standard (Fast Healthcare Interoperability Resources) sowie exemplarisch für den in der Hörgesundheitsversorgung gängigen HIMSA-Standard entwickelt. Dadurch konnte im Projektrahmen eine breite Palette an Datensätzen in das Avatarsystem eingebunden und die Grundlage geschaffen werden, in künftigen Weiterentwicklungen Daten von Praxispartnern nahtlos einzubinden.
- **Model Atlas Technologie:** Die Modellierung dieser Schnittstellen sowie der zugrundeliegenden Datenschutzmodelle erfolgte mit Hilfe der innovativen Model Atlas-Technologie von Data In Motion. Diese erlaubt eine hochflexible und wartbare Abbildung technischer und rechtlicher Anforderungen in maschinenlesbaren, interoperablen Datenmodellen.
- **Modulare Systemarchitektur:** Es wurde in Zusammenarbeit mit den anderen Projektpartnern eine verteilte Architektur realisiert, die auf einem semantischen Katalog (RDF) basiert. Alle Services und Metadaten sind dadurch im Netzwerk föderierbar und dynamisch nutzbar. Die Realisierung dieser Systemarchitektur hat darüber hinaus die Grundlage geschaffen, dass das Avatarsystem prinzipiell kompatibel mit GaiaX-Infrastrukturen ist und Bestandteil von Datenföderationen im Gesundheitsdatenbereich sein beziehungsweise auf diese zugreifen kann.
- **Frontend User-Interface (UI):** Obwohl im Antrag ursprünglich nicht vorgesehen, wurde für das Avatar-Gesamtsystem ein komplettes User-Interface erstellt, das es ermöglicht, entsprechend der Systemfunktionalität Datenanfragen zu formulieren und zu übermitteln, ohne über ein über das übliche Maß hinausgehendes technisches Wissen zu verfügen. Dieses User-Interface bietet vollumfängliche Individualisierungsoptionen für die Datenabfrage und war im Rahmen des Projektes der letzte Schritt, um Gesundheitsdaten, die über das Avatarsystem aggregiert wurden

Die Ergebnisse wurden in enger Abstimmung und Zusammenarbeit mit den teilnehmenden Forschungseinrichtungen, den klinischen sowie den technischen Praxispartnern erzielt. Data In Motion stellte dabei sicher, dass die wissenschaftlichen Konzepte der Partnerinstitutionen in eine industrietaugliche, modular erweiterbare Softwarearchitektur überführt wurden. Die erfolgreichen Demonstrationen auf Fachmessen wie der MEDICA unterstrichen die praktische Relevanz der von Data In Motion mitentwickelten Datenplattform.

3. Eingehende Darstellung

3.1 Überblick und Strukturierung der durchgeführten Arbeiten

Die Data In Motion Consulting GmbH war als technischer Umsetzungs- und Integrationspartner maßgeblich in die Teilprojekte 1, 3, 4 und 5 eingebunden.

Teilprojekt 1 bildete das koordinative Rückgrat für die technische Zusammenarbeit der verschiedenen Projektbeteiligten. Unser Schwerpunkt lag darauf, die Vielzahl an dezentral entstehenden Modulen zu einem funktionierenden Gesamtsystem zusammenzuführen. Wir haben dabei maßgebliche Aufgaben zur Architekturabstimmung und Sicherstellung der Dateninteroperabilität übernommen. Uns oblag dabei die technische Definition für den Informationsaustausch sowie die Entwicklung von Best-Practices für interne Schnittstellen. Mit unserer Modell-Registry haben wir sichergestellt, dass die im Projekt zahlreichen Modelle in unterschiedlichen technischen Formalisierungen zentral abgelegt und referenziert wurden. Darüber hinaus haben wir zur Überprüfung der Interoperabilität sogenannte "Test-Compatibility-Kits" (TCK) spezifiziert und in einem Test-Repository bereitgestellt, um die Qualitätssicherung zu gewährleisten.

In Teilprojekt 3 lag der Fokus auf der Erschließung sowie der rechtlich-technischen Bewertung von Datenquellen, die für die Generierung von Gesundheits-Avataren genutzt werden konnten. Wir brachten hier vor allem Expertise bei der Evaluation technischer Schnittstellen ein. Gemeinsam mit den Partnern wurde geprüft, inwiefern kommerzielle oder alternative Datenquellen, wie beispielsweise Wearables oder Fitness-Tracker, für den Gesundheitsbereich geeignet waren. DIM analysierte die zugehörigen Schnittstellen-Dokumentationen und bewertete die jeweilige Datenqualität. Ein wesentlicher Aspekt war zudem die lizenzrechtliche Verwertung sowie die Prüfung, ob die Datenfreigabemechanismen den Anforderungen des Datenschutzes und der Dateneigner entsprachen. Diese theoretischen Analysen mündeten schließlich in konkrete Anbindungskonzepte, um die Eignung der Datenquellen in einem Proof-of-Concept praktisch nachzuweisen.

Teilprojekt 4 befasste sich mit der technologischen Basis des Gesamtsystems. Data In Motion setzte hierbei auf moderne, modulare Architekturparadigmen, um die hohe Komplexität und die strengen Datenschutzanforderungen zu bewältigen. Es wurde ein Demonstrator für eine dynamische Datenpool-Architektur auf Basis des OSGi-Standards entwickelt, der eine flexible Kommunikation zwischen Komponenten verschiedener Programmiersprachen ermöglicht. Da das System verteilt konzipiert war, wurden Mechanismen zur Service-Discovery und Distribution implementiert. Ein besonders innovativer Ansatz war die durch uns entwickelte „Model-Driven-Privacy“ (MDP), bei der rechtliche Vorgaben als technisch interpretierbare Modelle abgebildet wurden. Diese fungierten als eine Art Filter, den nur datenschutzkonforme Informationen passieren konnten. Zudem wurden mittels des RDF-Standards semantische Zusammenhänge vereinheitlicht, um die Interoperabilität im europäischen Datenraum sicherzustellen.

Teilprojekt 5 stellte die finale Phase dar, in der die theoretischen Modelle und technischen Komponenten zur tatsächlichen Erstellung von Avataren zusammengeführt wurden. Hier

leiteten wir die zentralen Arbeitspakete zur Spezifikation und Systembereitstellung. Es wurden Anforderungen für die Schnittstellen definiert, über die Forschende und Interessierte Avatare abfragen konnten, und entsprechende Vorlagen (Templates) auf Basis von Praxis-Use-Cases entwickelt. Das System gleicht die Anfragen mit den verfügbaren Datenquellen im Datenpool ab und orchestriert die Daten (inkl. synthetisierter Daten) zu einem Antwortpaket. Um das System bereits frühzeitig prüfen zu können, wurde zudem verschiedene Testszenarien entwickelt und Testdaten generiert, mit deren Hilfe synthetische Avatare erzeugt werden konnten.

Als federführendes Projektmitglied bei der Softwarearchitektur und Programmierung lag unser Arbeitsschwerpunkt insbesondere auf den Teilprojekten 4 und 5, die inhaltlich wie organisatorisch im Projektverlauf eng miteinander verzahnt wurden. In unterstützender und beratender Funktion waren wir darüber hinaus auch maßgeblich in die Aktivitäten der anderen Teilprojekte eingebunden.

Im Sinne einer proaktiven Projektbeteiligung haben wir auch organisatorische Verantwortung übernommen. Als Teilprojektleitende für Teilprojekt 5 haben wir unseren Personalaufwand dazu genutzt, an den Quartalstreffen, den monatlichen Projektgesamttreffen sowie an 2-3-wöchigen Teilprojekttreffen teilzunehmen und den ungebrochenen Informationsfluss zwischen den zahlreichen Akteur:innen zu unterstützen. Mit unserer Expertise als Softwareunternehmen im Bereich verteilte Architekturen standen wir dabei teilprojektübergreifend auch für inhaltliche Beratungen zur Verfügung und unterstützten die Umsetzung von Public Relations Kommunikation durch unseren Input.

3.2 Eingehende Darstellung der Teilprojektarbeiten

3.2.1 Teilprojekt 1

Im Rahmen des Teilprojekts 1 haben wir mit der Spezifizierung und praktische Umsetzung der Test-Compatibility-Kits (TCK) in Arbeitspaket 1.3 einen Beitrag zur technischen Qualitätssicherung geleistet. Um eine hohe technische Qualität und Interoperabilität sicherzustellen, wurden zunächst etablierte Test-Frameworks wie das Mock Framework und JUnit Jupiter eingehend evaluiert und auf ihre spezifische Leistungsfähigkeit innerhalb der Projektarchitektur geprüft. Ein besonderes Augenmerk lag dabei auf der Kompatibilität mit OSGi-Tests, um eine nahtlose Unterstützung dieser modularen Umgebung zu gewährleisten. Parallel dazu wurden die Bedingungen für die Veröffentlichung von Test-Kit-Artefakten analysiert und ein demonstratives Setup für eine automatisierte Testumgebung im Sinne des Continuous Integration und Continuous Testing (CI/CT) Prinzips erstellt. Sämtliche Ergebnisse dieser Spezifikationsphase wurden umfassend dokumentiert. Dazu wurde auf GitHub ein Beispielprojekt als Vorlage für die TCKs aufgesetzt und mit beispielhaften Implementierungen die Testausführung für das Projektteam greifbar gemacht. Diese Arbeiten mündeten in den Meilenstein 1.3.3, wobei die konzeptionellen Beschreibungen und die Dokumentation im entsprechenden GitHub-Repository (https://github.com/DataInMotion/avatar-dataspaces-demo) hinterlegt wurden, um den Projektpartnern die aktive Arbeit mit den Kits zu ermöglichen. Dabei wurden abschließend

auch Best Practices für die internen AVATAR-Schnittstellen dokumentiert und mit Fortschreiten der Gesamtsystemorchestrierung fortlaufend reviewed.

Parallel dazu konzentrierten sich die Aktivitäten im Arbeitspaket 1.4 auf die Konzeption und Implementierung einer zentralen Model Registry, die als technologischer Dreh- und Angelpunkt für Metamodelle im AVATAR-System fungierte. Dieses Konzept basierte maßgeblich auf UML- und Ecore-Metamodellen. Um die operative Funktionalität dieser Registry zu demonstrieren, wurden exemplarische Microservices implementiert, welche sowohl die Registrierung neuer Modelle als auch deren Bereitstellung für Konsumierende über eine REST-Schnittstelle übernahmen. Diese bereits frühzeitig aufgesetzte Registry ermöglichte es dann später, verschiedene Metamodelle (z.B. HL7 FHIR, HIMSA) zu mappen und in das Avatarsystem zu laden.

Darauf aufbauend konzipierten wir die automatisierte clientseitige Nutzung der Model Registry. Hierbei wurde ein konfigurierbares URI-Mapping entwickelt, welches die Verwendung von Model-URIs und deren Zuweisung zu den entsprechenden Endpunkten in der Registry regelte, was eine nahtlose Integration für die Endnutzenden sicherstellte. Diese Entwicklungen wurden als Meilenstein 1.4.1 dokumentiert und über ein dediziertes GitHub-Repository (https://github.com/DataInMotion/avatar-modelrepo) zugänglich gemacht. Ergänzend dazu wurde ein Best-Practice-Dokument für die interoperable Modellierung verfasst und als Meilenstein 1.4.2 bereitgestellt. Dieses Dokument diente als zentrales Werkzeug für den Wissenstransfer und bildete die verbindliche Grundlage für die gemeinsame Modellierungsarbeit innerhalb des gesamten Projektteams.

Obwohl im ursprünglichen Antrag als Möglichkeit bei Inkompatibilitäten vorgesehen, mussten keine intermediären Modelle entwickelt werden. Stattdessen wurden im Umfang der Model Registry Modelltransformatoren bereitgestellt, um Daten zwischen verschiedenen Modellen konvertierbar zu machen und aus den Modellen Softwareartefakte generieren zu können.

3.2.2 Teilprojekt 3

Im Rahmen von Teilprojekt 3, das sich mit Datenspende und Einwilligung befasste, lag der Arbeitsschwerpunkt auf der Identifikation und Analyse alternativer sowie kommerzieller Gesundheitsdatenquellen. Gemäß Arbeitspaket 3.1 haben wir an einer umfassenden Evaluierung verschiedener Datenquellen mitgewirkt. Hierbei standen insbesondere die Plattformen Apple Health und Google Health sowie die dazugehörigen Software Development Kits (SDKs) im Fokus der Untersuchung. Um eine potentielle Integration in das AVATAR-Gesamtsystem zu bewerten, wurden umfassende Kriterien wie Datenexportmöglichkeiten, interne Datenmodellstrukturen und insbesondere die Kompatibilität zum internationalen HL7 FHIR-Standard aufgestellt, anhand derer die Plattformen dann geprüft wurden.

Parallel zur technischen Analyse wurden auch die rechtlichen und organisatorischen Rahmenbedingungen der Datengebenden intensiv evaluiert. Dies umfasste eine detaillierte Prüfung der jeweiligen Lizenzen sowie der Nutzungsbedingungen, denen die Anwendenden bei der Preisgabe ihrer Informationen zustimmen müssen. Im Fokus stand dabei die Analyse

der Mechanismen zur Weitergabe und Spende von Gesundheitsdaten, um sicherzustellen, dass die Prozesse den hohen Anforderungen des Datenschutzes und der Datensouveränität entsprachen.

Die Ergebnisse dieser technischen und rechtlichen Evaluationen wurden systematisch dokumentiert. Diese Vorarbeiten dienten als essenzielle Grundlage für die Bewertung der Datenqualität und die nachfolgende Konzeption technischer Anbindungsschnittstellen. Durch diese fundierte Analyse konnten wir sicherstellen, dass nur solche Datenquellen für die Generierung der Gesundheits-Avatare in Betracht gezogen wurden, die sowohl technisch interoperabel als auch datenschutzrechtlich einwandfrei nutzbar waren. Damit wurde die Basis für die praktischen Anbindungskonzepte geschaffen, die die potentielle Grundlage für die datenschutzkonforme Datenaggregation bildeten. Die Umsetzung dieses Ansatzes wurden allerdings nicht über die konzeptionelle Ebene hinaus weiter verfolgt, da die lizenzrechtliche Bewertung der so gewonnenen Daten komplex war und sich durch Praxispartner im Projekt genügend andere Datenquellen zur Fertigstellung eines funktionsfähigen Avatardemonstrators gefunden haben.

3.2.3 Teilprojekt 4

In Teilprojekt 4 konzentrierten sich unsere Aktivitäten auf die Schaffung einer hochmodernen, föderierten Datenpool-Architektur sowie die technische Umsetzung von Datenschutzmechanismen.

Im Rahmen des **Arbeitspakets 4.1** lag der Fokus auf der Konzeption und Realisierung einer zukunftsweisenden Datenpool-Architektur, die sowohl dynamische Modularität als auch eine föderierte Verteilung sicherstellte. Die Arbeiten gliederten sich dabei in die zwei wesentlichen Säulen der modularen Systemarchitektur sowie deren Überführung in eine verteilte Infrastruktur.

Zu Beginn der Arbeiten stand die Konzeption einer dynamischen Microservice-Architektur auf Basis des OSGi-Standards im Vordergrund (UAP 4.1.10). Die Data In Motion Consulting GmbH brachte hierbei ihre Expertise aus der Mitarbeit im Steering Committee der OSGi Working Group der Eclipse Foundation ein, um sicherzustellen, dass die Architektur höchsten Ansprüchen an Modularität und Interoperabilität genügte. Durch eine intensive Spezifikationsrecherche und die Evaluierung von Implementierungsansätzen wurde ein System entworfen, bei dem unterschiedliche Datenpools durch multiple Implementierungen desselben Services repräsentiert wurden. In einem evolutionären und iterativen Entwicklungsprozess wurden hierzu beispielhafte Demonstratoren nach dem Whiteboard-Pattern konzipiert, welche die Grundlage für die Service-Orchestrierung bildeten.

Der Meilenstein wurde durch die Etablierung eines Open-Source-Code-Repositorys dokumentiert, in dem die Realisierung der Architektur als transparentes Projekt erfolgte. Um die Integrität des Gesamtsystems zu wahren, wurden bereits in dieser frühen Phase Mechanismen zur Schnittstellen- und Modellvalidierung integriert, wobei gezielt auf die im

Arbeitspaket 1.3 entwickelten Prüfverfahren zurückgegriffen wurde. Daraus resultierte ein funktionsfähiger Demonstrator einer modularen Architektur, deren Einzelmodule nahtlos über eine zentrale Orchestrierung interagieren konnten.

Aufbauend auf diesen Ergebnissen wurde im weiteren Verlauf die Föderierung der Datenpool-Architektur vorangetrieben (UAP 4.1.22). Das Ziel bestand darin, die zuvor physisch gebundenen Komponenten in eine verteilte Systemarchitektur zu überführen. Hierzu wurde ein Konzept für eine föderierte Struktur entwickelt, die auf Service-Discovery und Service-Distribution basierte. Durch die Implementierung dynamischer Service-Announcements konnte sichergestellt werden, dass neue Systeminstanzen im Netzwerk automatisch erkannt und eingebunden wurden. Für die Verteilung der Funktionsaufrufe über verschiedene Instanzen hinweg wurde eine robuste Infrastruktur auf Basis von Container-Virtualisierung (Docker) geschaffen.

Die Arbeiten gipfelten in Meilenstein 4.1.3, mit dem ein Demonstrator für eine föderierte Datenpool-Architektur bereitgestellt wurde. Das System wurde so erweitert, dass eine Föderierung über mehrere Instanzen hinweg möglich war, die dezentral über eine einheitliche Topologie kommunizieren konnten. Die technische Validierung dieser Architektur erfolgte unter anderem durch die Umsetzung des Use-Case-Beispiels „ISMA“. Sämtliche Schnittstellendefinitionen, Implementierungen und die zugehörigen Docker-Container wurden im öffentlichen Repository (https://github.com/DataInMotion/avatar-dataspace) dokumentiert und für das Projektteam sowie externe Parteien zugänglich gemacht. Damit wurde eine skalierbare und technologisch hochmoderne Basis für die Verwaltung komplexer Gesundheitsdaten im AVATAR-Projekt geschaffen.

Im Rahmen des **Arbeitspakets 4.3** widmeten wir uns der technologischen Umsetzung der „Model-Driven Privacy“ (MDP). Unser primäres Ziel bestand darin, komplexe datenschutzrelevante Regularien in ein formalisiertes Modell zu überführen, das für technische Systeme unmittelbar interpretierbar war. Damit schufen wir eine Brücke zwischen den rechtlichen Anforderungen der DSGVO und der softwaretechnischen Implementierung im AVATAR-Gesamtsystem.

Zu Beginn unserer Arbeiten führten wir eine fundierte Recherche zu juristischen Kommentaren und rechtlichen Vorgaben durch, um die theoretische Basis für die Formalisierung zu legen. In enger Abstimmung mit Fachexpert:innen aus dem Projekt entwickelten wir allgemeine Abläufe für datenschutzrechtliche Ereignisse und leiteten daraus eine spezifische Domänensprache (DSL) ab. Diese Sprache ermöglichte es uns, datenschutzrechtliche Sachverhalte präzise zu beschreiben und anschließend in technisch verwertbare Modelle zu transformieren.

Ein zentraler technologischer Meilenstein war die Konzeption und Umsetzung eines Systems zur dynamischen Analyse von Metamodellen. Wir evaluierten verschiedene Algorithmen, um den Prozess der Identifikation DSGVO-relevanter Informationen innerhalb der Datenstrukturen zu automatisieren. Parallel dazu untersuchten wir spezifische Vokabularien für sensible medizinische Informationen, die als Grundlage für das Training von Machine-Learning-Algorithmen dienten.

Die daraus resultierende Architektur wurde so konzipiert, dass sie einen zweistufigen Scan-Prozess ermöglichte: Zunächst wurden die Metamodelle der Daten auf Datenschutz- und Transparenzrelevanz geprüft. In einem zweiten Schritt erfolgte die Untersuchung der tatsächlichen Modellinstanzen, also der konkreten Daten. Als wesentlichen Output dieses Prozesses implementierten wir die automatisierte Erstellung von Privacy-Reports, die den Status der Datenschutzumsetzung detailliert dokumentierten.

Zur Validierung unserer Konzepte schufen wir einen „Dummy-Case“ zur technischen Interpretation und setzten einen konkreten Use-Case technisch um (siehe Meilenstein 4.3.2). Die Ergebnisse und der zugehörige Quellcode wurden im Repository (https://github.com/DataInMotion/avatar-mdp) dokumentiert und für das Projektkonsortium bereitgestellt. Abschließend stellten wir sicher, dass die entwickelten Datenschutz-Modelle in die Modell-Registry (AP 1.4) zurückgespiegelt wurden, um eine projektweite Konsistenz der Datenschutzvorgaben zu gewährleisten. Durch diesen integrierten Ansatz konnten wir nachweisen, inwieweit bestimmte Datenschutz-Regularien für eine automatisierte technische Umsetzung geeignet waren und ein tragfähiges Modell für datenschutzrelevante Ereignisse im Gesundheitsbereich etablieren, das sich auch problemlos auf die gesamte Projektarchitektur skalieren ließ.

Im Rahmen des **Arbeitspakets 4.4** konzentrierten wir uns auf die technologische Befähigung des AVATAR-Systems zur Integration in (supra-)nationale Datenräume wie GAIA-X, den darauf aufbauenden European Health Data Space (EHDS) und die European Open Science Cloud (EOSC). Unser vorrangiges Ziel war die semantische Katalogisierung von Services, Schnittstellen und Modellen, um deren gegenseitige Bezüge maschinenlesbar und interoperabel abzubilden.

Zu Beginn unserer Arbeiten führten wir eine umfassende Standard-Recherche durch, um die Eignung verschiedener Metadaten-Standards und Kataloglösungen zu prüfen. Dabei identifizierten wir das Resource Description Framework (RDF) als das maßgebliche Format für die semantische Beschreibung der Ressourcen. Aufbauend auf diesen Erkenntnissen evaluierten wir „Fraunhofer Piveau“ als technologische Basis für einen Open-Data-Katalog. In diesem Kontext untersuchten wir intensiv den Standard DCAT (Data Catalog Vocabulary) und dessen spezifische Anwendung in IDS- (International Data Spaces) und GAIA-X-Infrastrukturen.

Ein wesentlicher Teil unserer praktischen Tätigkeit bestand in der Evaluierung geeigneter Schnittstellen und Modelle, die in einem direkten fachlichen Bezug zueinander standen. Um die Praxistauglichkeit des gewählten Ansatzes nachzuweisen, setzten wir einen demonstrativen Katalog um. Hierbei führten wir eine beispielhafte Registrierung von Datenmodellen aus der im Projekt entwickelten Modell-Registry (AP 1.4) in Fraunhofer Piveau unter konsequenter Verwendung von DCAT durch. Dieser Schritt war entscheidend, um die Kompatibilität mit GAIA-X-Datenräumen sicherzustellen, da diese DCAT als zentrale Grundlage für ihre Service- und Datenmodellregistrierung nutzen.

Parallel zur technischen Implementierung bewerteten wir die Eignung des Katalogs für den Einsatz im AVATAR-Gesamtsystem und definierten die notwendigen Mindestvoraussetzungen für die systemweite Integration. Die daraus gewonnenen

Erkenntnisse und Spezifikationsvorgaben flossen direkt in das Arbeitspaket 1.3 ein, um eine konsistente Architekturabstimmung zu gewährleisten. Durch die Dokumentation und Bereitstellung dieses Katalogs im RDF-Format schufen wir die notwendige Transparenz und technische Grundlage, damit die im Projekt entwickelten Services und Modelle über die Grenzen des Vorhabens hinaus in föderierten Dateninfrastrukturen auffindbar und nutzbar sind.

3.2.4 Teilprojekt 5

Für Teilprojekt 5 übernahmen wir eine zentrale Rolle bei der Leitung und technischen Ausgestaltung der Generierung dynamischer Avatare.

Im Rahmen von **Arbeitspaket 5.1** konzentrierten wir uns auf die Spezifikation der technischen Anforderungen für die Schnittstelle zu Avatar-Nutzenden. Wir erarbeiteten ein Vorschlagsdesign für eine modellierte Schnittstelle und legten dabei HL7 FHIR als die grundlegende Basis-Datenstruktur innerhalb des Avatarsystems fest (Meilenstein 5.1.1). Auf dieser Grundlage entwickelten wir ein Query-Modell unter Anwendung von Best-Practices für die verteilte, modulare Modellierung, um HL7-Daten gezielt abfragen zu können. Entgegen dem ursprünglichen Umfang des Projektantrages übernahmen wir darüber hinaus auch die Entwicklung eines korrespondierenden Frontend. Das durch uns entwickelte Abfrage-UI ermöglicht komplexe Spezifizierung von Datenabfragen aus dem Avatarsystem. Im Zuge dessen implementierten wir die dahinterliegende Schnittstelle zur Weiterleitung der Requests an den Avatar-Orchestrator.

Ein weiterer wesentlicher Bestandteil war die Erarbeitung der technischen Spezifikation des Avatar-Generators (siehe Meilenstein 5.1.3). Hierfür erstellten wir ein Generatormodell, welches Definitionen verschiedener Datenausgabeformate sowie Metadatendefinitionen für die Datenqualität, Anonymisierungsinformationen und Details zu den Konnektoren der Datenquellen enthielt. Zudem legten wir die Paketstruktur für die finale Avatarpaketierung sowie die entsprechenden Validierungskriterien für Metadaten und Inhalte fest. Im Rahmen des Unterarbeitspakets 5.1.8 setzten wir einen Testdatengenerator um, der auf Basis der definierten Schnittstellen und Templates die technische Verifizierung der erwarteten Avatar-Daten ermöglichte.

Im Rahmen von **Arbeitspaket 5.2** beteiligten wir uns an der Integration der Matching-Funktionalitäten in die Datenpool-Architektur (UAP 5.2.6). Wir extrahierten und abstrahierten hierbei Schnittstellen zur Anbindung an den Datenpool und setzten die notwendigen Spezifikationen für die Metadaten-Discovery um.

Innerhalb des **Arbeitspakets 5.3** wirkten wir an der technischen Validierung der Avatarinstanzen sowohl vor (UAP 5.3.4) als auch nach der Generierung (UAP 5.3.6) mit. Unser Fokus lag auf der Konzipierung und Umsetzung informationstechnischer Flows zur datenschutzrechtlichen und statistisch-qualitativen Bewertung der Daten. Dabei integrierten

wir eine exemplarische Ergebnis-Auditierung in einer Art digitaler Notar, um die Verarbeitungsschritte veränderungssicher zu dokumentieren und den in Teilprojekt 4 umgesetzten Model-Driven-Privacy Auditierungsprozess datenschutzqualitätstechnisch abzusichern.

Im Rahmen von **Arbeitspaket 5.4** koordinierten wir die Arbeiten zur Schaffung eines funktionalen Gesamtsystems. Gemeinsam mit der Navimatix GmbH spezifizierten wir die Laufzeitumgebung zur Avatargenerierung (Meilenstein 5.4.1) und definierten den Prozess zur Bereitstellung der notwendigen Daten. Aufgrund der anfallenden Datenmengen konzipierten wir mit der Navimatix eine Lösung zur temporären Ablage auf separaten Storages, wobei die Verlinkung der Daten über Metadaten des Kamunda-Orchestrators erfolgte und durch die Navimatix umgesetzt wurde.

Ein weiterer Schwerpunkt lag auf der Definition des Validierungsprozesses und der Fehlerbehandlung. Im Teilprojekt wurden Use-Cases für Rückinformationen an den Nutzer bei unvollständigen Daten sowie an den Orchestrator bei invaliden Datenbeständen spezifiziert. Abschließend wurde der Prozess zur Nutzendenbenachrichtigung nach erfolgreicher Fertigstellung eines Avatar-Datensets definiert. Dies umfasste eine Erstellungsroutine für temporäre One-Time-Links, die den Anwendenden einen sicheren und persistenten Download der generierten Datenpakete ermöglichten. Durch die Orchestrierung dieser Teilkomponenten wurde die Resilienz und Funktionstüchtigkeit des Gesamtsystems sichergestellt.