AVATAR

Teilprojekt Privacy Enhancing Technologies und Dataspace-Konnektoren

Schlussbericht Teil II: Eingehende Darstellung

Zuwendungsempfänger: Fraunhofer-Institut für Digitale Medientechnologie (IDMT) und Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung - Institutsteil Angewandte Systemtechnik (IOSB-AST)

Förderkennzeichen: 16KISA026

Teilprojekt: AVATAR - Privacy Enhancing Technologies und Dataspace-Konnektoren

Projektleitung Fraunhofer IDMT: Patrick Aichroth, patrick.aichroth@idmt.fraunhofer.de

Projektleitung Fraunhofer IOSB-AST: Dr. Daniel Rohrbach, daniel.rohrbach@iosb-ast.fraunhofer.de

Projektlaufzeit: 15.11.2022 - 31.12.2025

Inhaltsverzeichnis

1 Ausgangssituation / Motivation 1

2 Zielstellung 1

3 Durchführung und erzielte Ergebnisse 2

3.1 TP1 – Übergreifende Koordination und Architektur 2

3.2 TP2 – Anonymisierung von EEG- und MRI-Daten 4

3.3 TP4 – Datenpool und Dataspace-Infrastruktur 5

3.4 TP5 – Erzeugung der Avatare 8

3.5 TP6 – Reallabor und Proof of Concept 8

3.6 Abweichungen vom Arbeitsplan 9

4 Wichtigsten Positionen des zahlenmäßigen Nachweises 10

5 Notwendigkeit und Angemessenheit der Projektarbeiten 10

6 Nutzen und Verwertbarkeit der Projektergebnisse 11

7 Abgrenzung der Projektergebnisse zum Stand der Technik 12

8 Erfolgte oder geplante Veröffentlichungen des Ergebnisses 12

1 Ausgangssituation / Motivation

Die zunehmende Digitalisierung im Gesundheitswesen erzeugt große Mengen sensibler Gesundheitsdaten mit erheblichem Potenzial für Forschung und Versorgung. Gleichzeitig stellt deren Nutzung aufgrund regulatorischer Anforderungen eine Herausforderung dar. Vor diesem Hintergrund adressierte das Projekt AVATAR die Frage, wie Datensouveränität und Datennutzung technisch miteinander vereinbart werden können.

Im Fokus standen die Entwicklung und Integration von Privacy Enhancing Technologies (PET) sowie deren Einbindung in Dataspace-Infrastrukturen. Die im Projekt betrachteten Avatar-Konzepte dienten dabei als Anwendungskontext für die Aufbereitung und Nutzung anonymisierter Datenrepräsentationen, ohne selbst Gegenstand eigenständiger Technologieentwicklung zu sein.

Im Teilvorhaben der Fraunhofer-Institute IOSB-AST und IDMT bestand die Zielsetzung darin, technische Grundlagen für einen souveränen und datenschutzkonformen Datenaustausch zu schaffen. Das Fraunhofer IOSB-AST brachte hierzu seine Expertise im Bereich sicherer Datenräume, insbesondere im Kontext von GAIA-X und International Data Spaces, ein und konzipierte eine Infrastruktur für den standardisierten Datenaustausch mit integrierter Anonymisierung.

Das Fraunhofer IDMT ergänzte diese Arbeiten durch seine Expertise im Bereich technischer Datenschutzverfahren. Im Mittelpunkt stand die anwendungsspezifische Entwicklung und Integration modularer PET-Komponenten, insbesondere Anonymisierungs- und Evaluationsverfahren, sowie die Untersuchung weiterer Ansätze wie Differential Privacy und Federated Learning, sowie deren Integration in die Plattform. Ziel war es, einen technischen Rahmen zu schaffen, der eine datenschutzkonforme Bereitstellung sensibler Daten und deren Nutzung in exemplarischen Anwendungsszenarien ermöglicht.

2 Zielstellung

Das Teilvorhaben der Fraunhofer-Institute IOSB-AST und IDMT verfolgte das übergeordnete Ziel, eine technische Infrastruktur für den datenschutzkonformen und souveränen Austausch von Gesundheitsdaten im Kontext der Avatar-Erzeugung zu entwickeln. Dabei sollten Privacy Enhancing Technologies (PET) unmittelbar in die Datenaustauschprozesse integriert werden, um die Anonymisierung personenbezogener Daten bereits an der Quelle sicherzustellen.

Ein zentrales Ziel bestand in der Untersuchung und Anpassung eines GAIA-X-konformen Dataspace-Konnektors, der die Integration von PET-Verfahren in den Datenaustausch ermöglicht. Der Aufbau einer GAIA-X-Föderation gemeinsam mit den Projektpartnern sollte eine Interoperabilität mit übergeordneten Datenrauminitiativen, insbesondere im Gesundheitsbereich, unterstützen.

Neben klassischen Anonymisierungsverfahren wurden ergänzend Differential Privacy und Secure Federated Learning betrachtet. Die Arbeiten zielten dabei primär auf eine Einordnung der Verfahren im Kontext der vorgesehenen Anwendungsszenarien. Während Differential Privacy konzeptionell analysiert und prototypisch implementiert wurde, lag der Fokus der weiteren Arbeiten auf Verfahren, die sich im Kontext der Use Cases unmittelbar integrieren und evaluieren ließen.

Ein weiterer Fokus lag auf der Entwicklung resilienter Systemarchitekturen für Dataspace-Konnektoren. Hierzu wurden self2x-Ansätze, etwa self-integration und self-healing, sowie Methoden zur Lastverteilung betrachtet, um eine stabile und skalierbare Systemumgebung zu gewährleisten. Ergänzend war die Umsetzung eines GAIA-X-Demonstrators im Reallabor als Proof-of-Concept vorgesehen, der als Referenzarchitektur für die Integration weiterer Projektpartner dienen sollte.

Zur strukturierten Umsetzung der Projektziele wurden die Arbeiten in mehrere Teilprojekte und Arbeitspakete gegliedert, deren Aufwände im Teilvorhaben der Fraunhofer-Institute wie folgt verteilt waren:

Teilprojekt	Arbeitspaket	IOSB-AST	IDMT
TP1 – Koordination	AP 1.2, 1.3, 1.4	1 PM	3,5 PM
TP2 – Anonymisierung	AP 2.4, 2.5	–	3 PM
TP4 – Datenpool	AP 4.1, 4.2, 4.3, 4.4	45 PM	25 PM
TP5 – Avatar-Erzeugung	AP 5.1, 5.3	–	17 PM
TP6 – Reallabor	AP 6.1, 6.2	10 PM	–

3 Durchführung und erzielte Ergebnisse

3.1 TP1 – Übergreifende Koordination und Architektur

Im Rahmen von TP1 erfolgte die konzeptionelle Abstimmung der Projektziele sowie die Ausarbeitung einer gemeinsamen Systemarchitektur für das AVATAR-Gesamtsystem. Ausgangspunkt bildeten konsortialübergreifend entwickelte Personas und User Stories, aus denen funktionale und nicht-funktionale Anforderungen für die technische Umsetzung abgeleitet wurden. Diese Anforderungen dienten insbesondere dazu, die Integration der beteiligten Komponenten in eine konsistente Gesamtarchitektur zu unterstützen und zentrale Anwendungsfälle für die datenschutzkonforme Nutzung von Gesundheitsdaten zu konkretisieren.

Auf dieser Grundlage wurde eine initiale Architektur des AVATAR-Systems entworfen, welche die Interaktion der beteiligten Akteure, Datenquellen sowie der geplanten Plattformkomponenten abbildet. Der ursprüngliche Architekturentwurf ist in Abbildung 1 dargestellt. Er veranschaulicht insbesondere die geplante Einbindung der Dataspace-Konnektoren, der Anonymisierungskomponenten sowie der Dienste zur Avatar-Erzeugung innerhalb eines föderierten Datenökosystems.

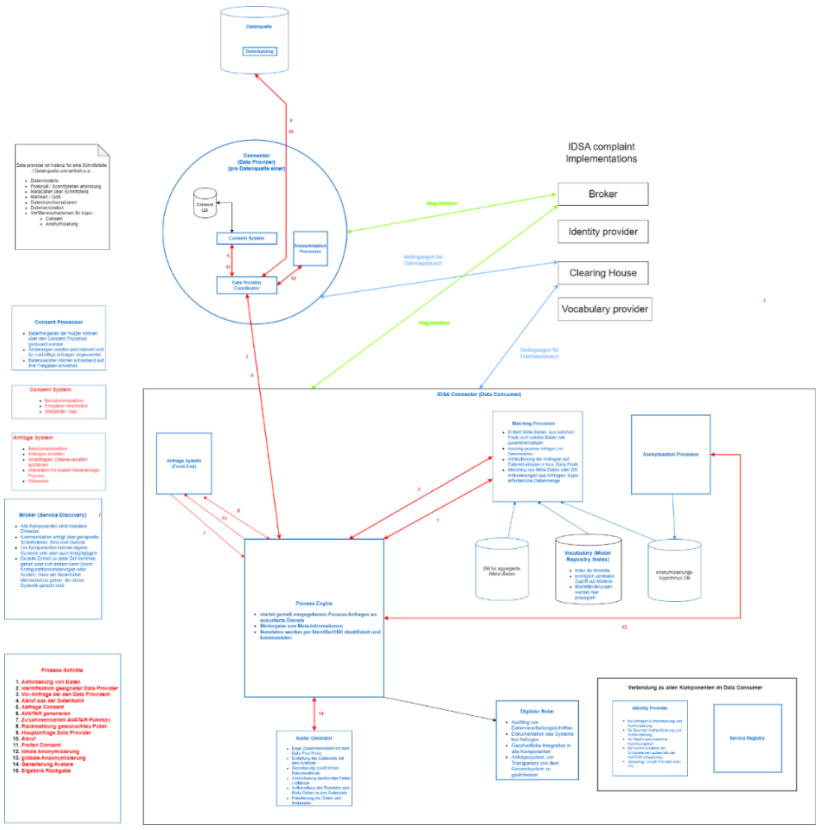

Abbildung 1: Skizze zum ursprünglichen Architekturentwurf des AVATAR-Systems

Im weiteren Projektverlauf wurde diese Architektur iterativ vom Fraunhofer IOSB-AST weiterentwickelt und an die Erkenntnisse aus den technischen Arbeitspaketen angepasst. Die überarbeitete Architektur berücksichtigt insbesondere die konkrete Umsetzung der Dataspace-Infrastruktur, die Integration der PET-Komponenten sowie die Anforderungen an resiliente Datenflüsse. Eine Übersicht der weiterentwickelten Architektur ist in Abbildung 2 dargestellt. Die iterative Verfeinerung der Architektur stellte sicher, dass sowohl konzeptionelle Anforderungen als auch praktische Implementierungserfahrungen in die Systemgestaltung einfließen konnten.

Aus Sicht des Fraunhofer IDMT bestand der Beitrag in TP1 vor allem in der Überführung der aus den Personas und User Stories abgeleiteten Anforderungen in konkrete technische Anforderungen an Privacy Enhancing Technologies. Hierzu wurden Anforderungen an die Integration von Anonymisierungsverfahren, Evaluationsmechanismen sowie an die Schnittstellen zwischen PET-Komponenten und Dataspace-Infrastruktur spezifiziert. Darüber hinaus erfolgten Architekturbeiträge aus PET-Perspektive, insbesondere zur Einbindung von Anonymisierungsdiensten in Datenflüsse sowie zur Berücksichtigung von Evaluationsmechanismen zur Bewertung von Privacy-Eigenschaften.

Das Fraunhofer IOSB-AST brachte seine Expertise im Bereich Dataspace-Architekturen ein und verantwortete die konzeptionelle Ausgestaltung der Infrastrukturkomponenten für den souveränen Datenaustausch. Die Arbeiten umfassten unter anderem die Ausarbeitung von Schnittstellenspezifikationen, Ablaufmodellen sowie die Ableitung technischer Anforderungen für die Integration der AVATAR-Komponenten in föderierte Datenräume.

Insgesamt stellte TP1 die konzeptionelle Grundlage für die nachfolgenden technischen Arbeiten dar. Die iterativ entwickelte Architektur ermöglichte eine konsistente Integration der in den weiteren Teilprojekten entwickelten Komponenten und bildete damit die Basis für die Umsetzung des Reallabors sowie die Validierung der entwickelten Konzepte im Projektverlauf.

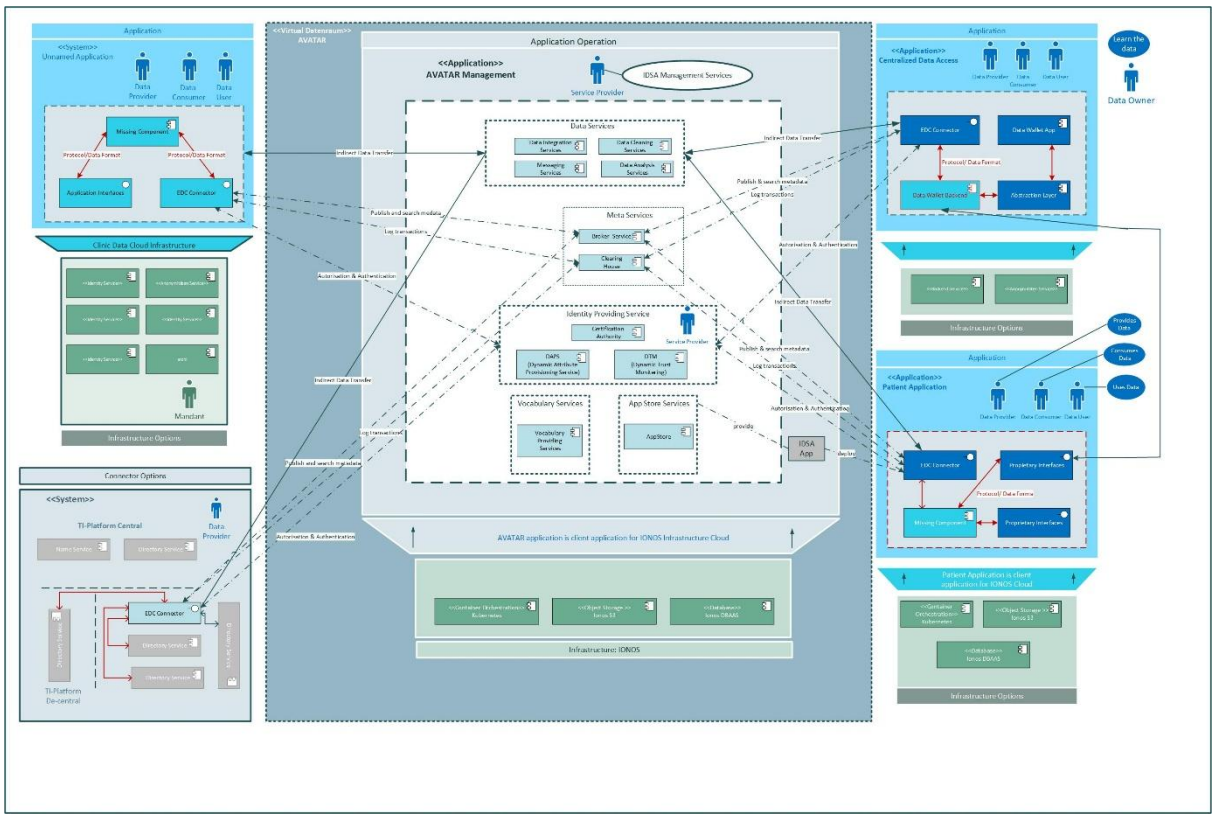

Abbildung 2: Weiterentwickelter Entwurf für die AVATAR-Komponenten

3.2 TP2 – Anonymisierung von EEG- und MRI-Daten

Im Rahmen von TP2 wurden exemplarische Privacy-Enhancing-Technologien für komplexe medizinische Datentypen untersucht und in die AVATAR-Plattform integriert. Ziel war es, die Funktionsfähigkeit der Plattform für Biosignal- und Bilddaten zu validieren und die Einbindung spezialisierter Anonymisierungsverfahren in die Gesamtarchitektur zu demonstrieren. Der Beitrag des Fraunhofer IDMT konzentrierte sich dabei auf die Bereitstellung und Integration bestehender PET-Komponenten für EEG- und MRI-Daten.

EEG-Anonymisierung und Re-Identifizierung: Für den Bereich der EEG-Daten wurde eine im assoziierten Projekt NEMO entwickelte Anonymisierungskomponente in den Kontext des AVATAR-Projekts übertragen und als wiederverwendbarer Baustein bereitgestellt. Die Integration erfolgte dabei in Form bereitgestellter Komponenten und exemplarischer Evaluierungen und bildete eine Grundlage für weiterführende Arbeiten an komplexeren Workflow-Szenarien.

Zur Bewertung der Privacy-Eigenschaften wurde ergänzend die ebenfalls im Projekt NEMO entwickelte Re-Identifizierungsanalyse genutzt. Diese diente dazu, die Wirksamkeit der Anonymisierung exemplarisch zu überprüfen und die Eignung der Verfahren für die Verarbeitung von Biosignaldaten im AVATAR-Kontext zu bewerten. Die Kombination aus Anonymisierung und Re-Identifizierungsanalyse ermöglichte damit eine praxisnahe Validierung der PET-Integration und zeigte die Übertragbarkeit bestehender Forschungsergebnisse auf neue Anwendungsszenarien.

Die Arbeiten führten zu einer erfolgreichen funktionalen Integration der EEG-Anonymisierung in die Plattform und demonstrierten die Möglichkeit, spezialisierte PET-Komponenten als modulare Bausteine in eine Dataspace-basierte Infrastruktur einzubinden.

Anonymisierung von MRI-Bilddaten: Ergänzend wurden Verfahren zur Anonymisierung von MRI-Bilddaten untersucht und exemplarisch integriert. Ziel dieser Arbeiten war die Erfassung des Stands der Technik im Bereich Defacing- und Anonymisierungsverfahren sowie deren Einbindung in die AVATAR-Plattform, um die Unterstützung weiterer medizinischer Datentypen zu demonstrieren.

Die Untersuchung umfasste eine qualitative Bewertung verschiedener verfügbarer Ansätze und Tools zur Entfernung identifizierender Bildinformationen. Im Ergebnis wurden geeignete Verfahren ausgewählt und als Komponenten bereitgestellt, die analog zur EEG-Anonymisierung als Services in die Plattform integriert werden konnten. Die Integration diente insbesondere dazu, die Plattformfunktionalitäten für Bilddaten zu validieren und die generelle Übertragbarkeit des PET-Ansatzes auf unterschiedliche Datentypen zu demonstrieren.

Einordnung der Arbeiten in TP2: Die Arbeiten in TP2 verdeutlichten, dass die Integration spezialisierter PET-Komponenten für komplexe medizinische Datentypen grundsätzlich möglich ist und dass bestehende Forschungsergebnisse aus anderen Projekten erfolgreich in die AVATAR-Architektur übertragen werden können. Gleichzeitig zeigte sich, dass die Nutzung exemplarischer Datensätze – sowohl aus öffentlichen Quellen als auch synthetischer Natur – eine praktikable Grundlage für die funktionale Validierung darstellt, insbesondere vor dem Hintergrund eingeschränkter Datenverfügbarkeit im medizinischen Umfeld.

Insgesamt leistete TP2 damit einen wichtigen Beitrag zur Demonstration der Modularität der AVATAR-Plattform und legte die Grundlage für die Integration weiterer datentypspezifischer Anonymisierungsverfahren in zukünftigen Projekten.

3.3 TP4 – Datenpool und Dataspace-Infrastruktur

Die Arbeiten in TP4 bildeten den Schwerpunkt des Teilvorhabens des Fraunhofer IOSB-AST und umfassten die Konzeption sowie die prototypische Umsetzung zentraler Komponenten für den souveränen Datenaustausch im AVATAR-System. Ziel war die Entwicklung einer Dataspace-basierten Infrastruktur, die eine datenschutzkonforme Bereitstellung und Nutzung sensibler Gesundheitsdaten ermöglicht und die Integration von Privacy Enhancing Technologies (PET) in die Datenflüsse unterstützt. Das Fraunhofer IDMT brachte in diesem Kontext PET-Komponenten als modulare Bausteine ein, die in die entwickelte Infrastruktur integriert wurden.

Zu Beginn der Arbeiten erfolgte eine systematische Analyse bestehender Konzepte und Referenzarchitekturen im Bereich föderierter Datenräume. Hierzu wurden insbesondere die GAIA-X-Kernkonzepte, das IDS Reference Architecture Model sowie verfügbare Minimum-Viable-Products untersucht. Die Ergebnisse dieser Analyse bildeten die Grundlage für die Ableitung der Systemarchitektur sowie der funktionalen Anforderungen an die im Projekt zu entwickelnden Komponenten. Ein besonderer Fokus lag auf der Frage, wie sensible Gesundheitsdaten bereits beim Zugriff anonymisiert werden können, um Datensouveränität und regulatorische Anforderungen zu gewährleisten.

Auf dieser Basis wurden Schnittstellen und Integrationsmechanismen für heterogene Gesundheitsdaten analysiert und spezifiziert. Als zentraler Standard für strukturierte Daten wurde FHIR identifiziert und für die semantische Beschreibung sowie den interoperablen Datenaustausch herangezogen. Parallel dazu wurden Anforderungen an Verschlüsselung, Zugriffskontrolle und sichere Datenübertragung definiert sowie Integrationspunkte für die Anonymisierungskomponenten festgelegt.

Als technische Basis für die Infrastruktur wurden die Eclipse Dataspace Components (EDC) eingesetzt und projektspezifisch erweitert. Ein zentrales Konzept stellte dabei das Provider-Push-Verfahren dar, bei dem Datenanbieter die Kontrolle über ihre Daten bis zur Initiierung des Transfers behalten. Dadurch konnte die Anonymisierung vor der Datenübertragung durchgeführt werden, was den Prinzipien der

Datensouveränität im GAIA-X-Kontext entspricht. Der resultierende Datenfluss und die Positionierung der Anonymisierung innerhalb der Transferpipeline sind in Abbildung 3 dargestellt.

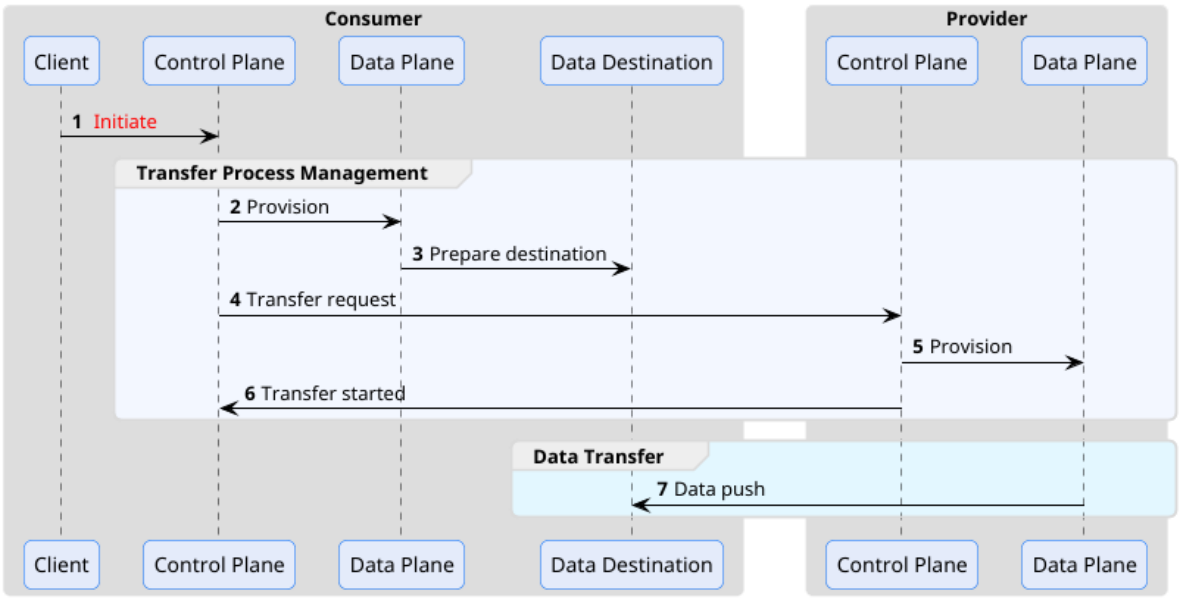

Abbildung 3: Darstellung des Prozesses beim Provider Push-Verfahren der EDC[1]

Im Rahmen der Umsetzung wurden Erweiterungen der Dataspace-Konnektoren entwickelt, die REST-basierte Schnittstellen, Konfigurationsmechanismen sowie die Integration eines Anonymisierungsdienstes umfassen. Ergänzend wurde ein Dispatcher-Service implementiert, der die Vertragsverhandlung und den Datentransfer orchestriert und damit die Interaktion zwischen Workflow-Orchestrierung und Dataspace-Konnektoren unterstützt. Die Infrastruktur wurde containerbasiert umgesetzt und ermöglichte eine modulare Bereitstellung der einzelnen Komponenten.

Ein wesentlicher Beitrag des Fraunhofer IDMT bestand in der Bereitstellung von PET-Komponenten, die in die Dataspace-Infrastruktur integriert wurden. Dazu gehörten insbesondere Verfahren zur Anonymisierung strukturierter Daten, die unterschiedliche Techniken wie Generalisierung, Suppression, Mikroaggregation und Perturbation kombinieren. Ergänzend wurden Evaluationskomponenten bereitgestellt, die auf Metriken wie k-Anonymität, l-Diversität und t-Closeness basieren und eine exemplarische Bewertung der Privacy-Eigenschaften anonymisierter Daten ermöglichen. Die grundsätzliche Einbindung der PET-Komponenten in den Datenfluss ist schematisch in Abbildung 4 dargestellt.

[1] https://eclipse-edc.github.io/documentation/for-adopters/control-plane/#provider-push

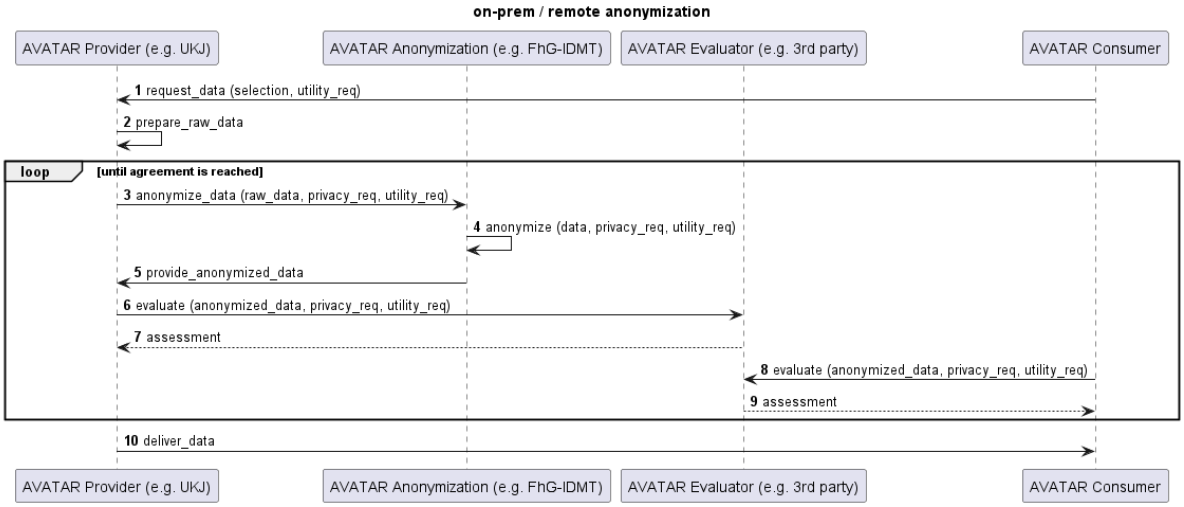

Abbildung 4: Vereinfachter Content Provision Workflow mit Komponenten für Anonymisierung und Evaluation

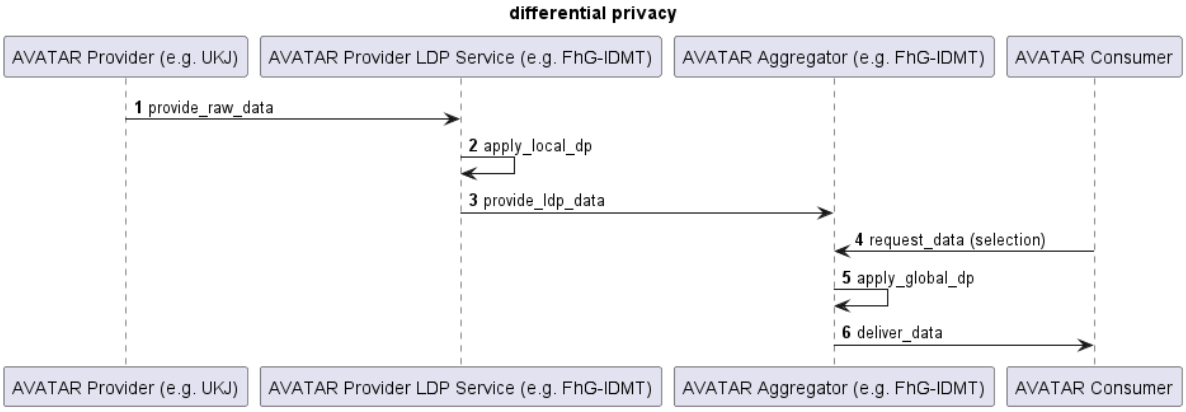

Abbildung 5: Vereinfachter Content Provision Workflow mit Differential Privacy Komponenten

Differential Privacy wurde ergänzend als PET-Ansatz untersucht und prototypisch vorbereitet, vgl. Abbildung 5. Im Verlauf der Use-Case-orientierten Arbeiten zeigte sich jedoch, dass die betrachteten Szenarien überwiegend durch klassische Anonymisierungsverfahren adressiert werden konnten. Die im Projekt gewonnenen Erkenntnisse flossen daher vor allem in eine konzeptionelle Einordnung des Verfahrens und dienen als Referenz für zukünftige Anwendungskontexte mit anderen Daten- und Analyseanforderungen.

Darüber hinaus wurden Konzepte zur Resilienz und Skalierbarkeit der Dataspace-Infrastruktur erarbeitet. Diese umfassten unter anderem Ansätze zur dezentralen Architektur, Mechanismen zur Lastverteilung sowie Self2x-Konzepte zur Unterstützung von Selbstintegration und Fehlertoleranz. Ergänzend wurde die Einbettung der Infrastruktur in ein GAIA-X-Ökosystem untersucht und prototypisch validiert.

Die Funktionsfähigkeit der Infrastruktur wurde anhand exemplarischer Testdaten demonstriert. Dabei wurde der vollständige Datenfluss von der Bereitstellung eines Datenangebots über Vertragsverhandlung und Datentransfer bis zur Anonymisierung und Bereitstellung der anonymisierten Daten nachvollzogen. Aufgrund eingeschränkter Datenverfügbarkeit im medizinischen Kontext erfolgte die Validierung mit exemplarischen und synthetischen Datensätzen.

Ergänzend zu den in die Infrastruktur integrierten PET-Komponenten wurde am Fraunhofer IDMT ein Demonstrator für Secure Federated Learning als eigenständiger Proof-of-Concept umgesetzt. Dieser Demonstrator war nicht Teil des Dataspace-Workflows, zeigte jedoch exemplarisch die Möglichkeiten

verteilter Lernverfahren zur datenschutzfreundlichen Analyse sensibler Daten und lieferte Impulse für zukünftige Erweiterungen der Plattform.

Insgesamt demonstrierten die Arbeiten in TP4 die grundsätzliche Machbarkeit eines souveränen Datenaustauschs mit integrierter Anonymisierung in einer Dataspace-basierten Infrastruktur. Gleichzeitig wurde deutlich, dass die Integration neuer Datenraumkonzepte sowie die eingeschränkte Verfügbarkeit realer Gesundheitsdaten eine iterative Weiterentwicklung der Ansätze erfordert.

3.4 TP5 – Erzeugung der Avatare

Im Rahmen von TP5 wurden konzeptionelle Grundlagen für die Nutzung anonymisierter Daten im Kontext der Avatar-Erzeugung betrachtet. Der Beitrag des Fraunhofer IDMT bestand hierbei in der Bereitstellung von PET-Komponenten, die als Bausteine in entsprechende Datenaufbereitungsprozesse eingebracht werden konnten. Dazu gehörten Verfahren zur Anonymisierung strukturierter Daten sowie spezialisierte Verfahren für EEG- und MRI-Daten und Evaluationskomponenten zur Bewertung der Privacy-Eigenschaften.

Die Einbindung dieser Komponenten erfolgte über definierte Schnittstellen in prototypische Workflows, die durch andere Teilvorhaben orchestriert wurden. Ergänzend wurden erste Arbeiten zur strukturierten Beschreibung und Kombination von PET-Verfahren durchgeführt. Ziel war es, die Auswahl und Parametrisierung von Privacy-Mechanismen stärker modellbasiert zu unterstützen und damit die flexible Anpassung an unterschiedliche Datentypen und Anwendungsszenarien vorzubereiten. Die Arbeiten hatten explorativen Charakter und lieferten konzeptionelle Bausteine für eine mögliche Weiterentwicklung entsprechender Modellierungsansätze.

Die Validierung der bereitgestellten Komponenten erfolgte im Rahmen exemplarischer Use Cases mit unterschiedlichen Datentypen. Dabei kamen sowohl öffentlich verfügbare Datensätze als auch synthetische Daten zum Einsatz, um funktionale Aspekte der Avatar-Erzeugung zu demonstrieren. Die Arbeiten zeigten, dass die Kombination aus Anonymisierung und Privacy-Evaluation eine praktikable Grundlage für die Generierung datenschutzkonformer Avatare darstellt, zugleich jedoch eine weitergehende Integration von Utility-Bewertungen und modellgetriebenen Privacy-Konzepten für zukünftige Anwendungen erforderlich ist.

3.5 TP6 – Reallabor und Proof of Concept

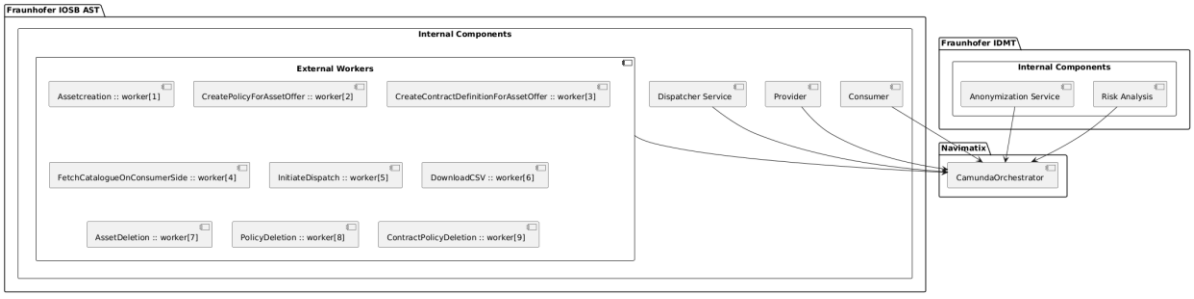

Abbildung 6: Komponentendiagramm für den Demonstrator

Im Rahmen von TP6 wurden die in den vorherigen Teilprojekten entwickelten Konzepte und Komponenten in einem Reallabor zusammengeführt und als Proof of Concept umgesetzt, vgl. Abbildung 6. Ziel war es, die grundsätzliche Funktionsfähigkeit der AVATAR-Architektur sowie die Interaktion der beteiligten Komponenten in einem integrierten Demonstrator zu veranschaulichen.

Ausgangspunkt bildete die Analyse von Stakeholder-Anforderungen und Use Cases, auf deren Basis ein Konzept für den GAIA-X-basierten Demonstrator entwickelt wurde. Die betrachteten Anwendungsszenarien umfassten insbesondere die Bereitstellung von Datenangeboten durch

Datenanbieter, die Katalogabfrage durch Datennutzer, die automatisierte Vertragsverhandlung sowie den anonymisierten Datentransfer. Ergänzend wurde die Protokollierung der Datenzugriffe berücksichtigt, um Transparenz und Nachvollziehbarkeit der Datenverarbeitung zu gewährleisten.

Die Umsetzung des Demonstrators erfolgte durch das Fraunhofer IOSB-AST und umfasste die Integration der entwickelten Dataspace-Komponenten in eine containerbasierte Infrastruktur. Hierzu gehörten insbesondere Provider- und Consumer-Konnektoren, Orchestrierungskomponenten sowie die notwendigen Workflow-Mechanismen zur Abbildung der Datenflüsse. Der Demonstrator ermöglichte damit die exemplarische Durchführung eines vollständigen Datenverarbeitungsprozesses von der Asset-Registrierung über Vertragsverhandlung und Datentransfer bis hin zur Bereitstellung anonymisierter Daten.

Der Beitrag des Fraunhofer IDMT in TP6 bestand in der Bereitstellung der PET-Komponenten, die im Rahmen des Demonstrators genutzt werden konnten. Dazu gehörten insbesondere die Anonymisierung strukturierter Daten sowie die Evaluationskomponenten zur Bewertung der Privacy-Eigenschaften. Diese Komponenten wurden über die definierten Schnittstellen in die Demonstratorumgebung eingebunden und dienten als funktionale Bausteine innerhalb der prototypischen Workflows, ohne dass eine direkte Verantwortung für die Orchestrierung bestand.

Die Validierung des Demonstrators erfolgte im Rahmen interner Demonstrationen und Tests mit exemplarischen Datensätzen. Aufgrund eingeschränkter Verfügbarkeit realer Gesundheitsdaten kamen überwiegend synthetische und öffentlich verfügbare Daten zum Einsatz. Dennoch konnte die prinzipielle Funktionsfähigkeit der Architektur sowie die Integration der PET-Komponenten in die Datenflüsse nachvollzogen werden. Der Demonstrator wurde projektintern vorgestellt und diente als Referenz für die weitere Entwicklung und Integration in zukünftigen Vorhaben.

Insgesamt zeigte TP6, dass die im Projekt entwickelten Komponenten in einer gemeinsamen Infrastruktur zusammengeführt und als integrierter Demonstrator betrieben werden können. Gleichzeitig verdeutlichte das Reallabor die Herausforderungen bei der Integration heterogener Komponenten sowie die Bedeutung standardisierter Schnittstellen und verfügbarer Testdaten für die Weiterentwicklung der Plattform.

3.6 Abweichungen vom Arbeitsplan

Im Projektverlauf traten keine grundlegenden Abweichungen vom geplanten Arbeitsprogramm auf. Die in den Arbeitspaketen vorgesehenen Themenfelder konnten insgesamt bearbeitet und die angestrebten Beiträge des Fraunhofer IDMT sowie des Fraunhofer IOSB-AST realisiert werden. Dennoch ergaben sich im Detail Anpassungen und Herausforderungen, die im Folgenden zusammengefasst werden.

Eine wesentliche Herausforderung stellte die eingeschränkte Verfügbarkeit realer Gesundheitsdaten dar. Diese wirkte sich insbesondere auf die Validierung der entwickelten Komponenten und Demonstratoren aus. Um dennoch eine funktionale Bewertung zu ermöglichen, wurde verstärkt auf öffentlich verfügbare Datensätze sowie synthetische Daten zurückgegriffen. Diese Anpassung hatte keinen Einfluss auf die grundsätzliche Zielerreichung, schränkte jedoch die Möglichkeit ein, bestimmte Anwendungsszenarien mit realen Daten vollständig zu demonstrieren.

Im Bereich der Privacy Enhancing Technologies zeigte sich, dass die im Projekt vorgesehenen Verfahren unterschiedliche Reifegrade aufwiesen und je nach Anwendungsfall variierend geeignet waren. Differential Privacy wurde im Projekt als Komponente bereitgestellt und konzeptionell analysiert. Für die betrachteten Anwendungsszenarien erwiesen sich jedoch klassische Anonymisierungsverfahren als geeigneter, sodass Differential Privacy nicht im Fokus der prototypischen Integration stand. Diese Einordnung ist als use-case-spezifische Entscheidung zu verstehen und schränkt die Relevanz des Ansatzes für zukünftige Anwendungen nicht ein.

Die prototypische Umsetzung von Secure Federated Learning am Fraunhofer IDMT erfolgte planmäßig als eigenständiger Demonstrator. Die Integration in die Dataspace-Infrastruktur war im Projekt nicht vorgesehen und stellte daher keine Abweichung dar. Der Demonstrator lieferte jedoch wertvolle Erkenntnisse zur Machbarkeit verteilter Lernverfahren im Kontext datenschutzfreundlicher Datenanalyse.

Darüber hinaus zeigte sich im Projektverlauf, dass die Integration heterogener Komponenten innerhalb einer Dataspace-basierten Architektur mit erhöhtem Abstimmungsaufwand verbunden ist. Dies führte zu iterativen Anpassungen der Architektur und der Schnittstellen, insbesondere im Zusammenspiel der verschiedenen Teilvorhaben. Diese Anpassungen sind als typischer Bestandteil prototypischer Entwicklungsarbeiten zu verstehen und trugen zur Stabilisierung der Gesamtarchitektur bei.

Insgesamt führten die genannten Aspekte zu keiner wesentlichen Gefährdung der Projektziele. Vielmehr ermöglichten die iterativen Anpassungen eine realistische Validierung der entwickelten Konzepte und lieferten wichtige Erkenntnisse für die Weiterentwicklung der Ansätze in zukünftigen Projekten.

4 Wichtigsten Positionen des zahlenmäßigen Nachweises

Die im Projekt angefallenen Ausgaben entsprechen im Wesentlichen der ursprünglichen Planung und spiegeln die im Arbeitsplan vorgesehenen Tätigkeiten wider. Den größten Anteil der Aufwendungen bildeten Personalkosten für wissenschaftliches und technisches Personal, die insbesondere für die Entwicklung und Integration von Privacy Enhancing Technologies sowie die Konzeption und prototypische Umsetzung der Dataspace-Infrastruktur eingesetzt wurden.

Darüber hinaus entstanden Sachkosten im Zusammenhang mit der prototypischen Umsetzung der Komponenten, der Bereitstellung von Entwicklungs- und Testumgebungen sowie der Durchführung von Integrations- und Validierungsarbeiten. Hierzu zählten insbesondere Aufwendungen für Software, IT-Infrastruktur und projektbezogene Dienstleistungen.

Investive Ausgaben spielten im Projekt eine untergeordnete Rolle, da die Arbeiten überwiegend auf bestehenden Entwicklungsumgebungen und Infrastrukturressourcen aufbauten. Reise- und Disseminationskosten fielen in begrenztem Umfang an und standen im Zusammenhang mit projektinternen Abstimmungen sowie der Präsentation der Projektergebnisse im Konsortium.

Geringfügige Abweichungen zwischen Planung und tatsächlich entstandenen Kosten ergaben sich aus projektbedingten Anpassungen im Verlauf der prototypischen Entwicklungsarbeiten und hatten keinen Einfluss auf Zielerreichung oder Projektverlauf. Insgesamt ist festzustellen, dass die Mittel wirtschaftlich und zweckentsprechend eingesetzt wurden und eine direkte Zuordnung der Aufwendungen zu den im Projekt durchgeführten Arbeiten gegeben ist.

5 Notwendigkeit und Angemessenheit der Projektarbeiten

Die im Projekt durchgeführten Arbeiten waren erforderlich, um den steigenden Anforderungen an eine datenschutzkonforme Nutzung sensibler Gesundheitsdaten im Kontext föderierter Datenräume zu begegnen. Das AVATAR-Projekt adressierte dabei insbesondere die technische Integration von Privacy Enhancing Technologies in Dataspace-Infrastrukturen und leistete damit einen Beitrag zur Vereinbarkeit von Datensouveränität und Datennutzung.

Die Arbeiten des Fraunhofer IDMT zur Entwicklung und Integration von Anonymisierungs- und Evaluationskomponenten waren essenziell, um Privacy-Anforderungen im Projekt technisch umzusetzen und exemplarisch zu validieren. Ebenso waren die Arbeiten des Fraunhofer IOSB-AST zur Konzeption und prototypischen Umsetzung der Dataspace-Infrastruktur notwendig, um die Grundlagen für einen

souveränen Datenaustausch zu schaffen und die Anschlussfähigkeit an Initiativen wie GAIA-X sicherzustellen.

Die prototypische Umsetzung im Reallabor stellte ein angemessenes methodisches Vorgehen dar, da sie eine realitätsnahe Bewertung der entwickelten Konzepte ermöglichte. Die Nutzung exemplarischer und synthetischer Datensätze war dabei ein geeigneter Ansatz, um funktionale Aspekte der Architektur zu demonstrieren und gleichzeitig datenschutzrechtliche Anforderungen zu berücksichtigen.

Die Projektarbeiten waren hinsichtlich Umfang und Vorgehensweise angemessen, um die Projektziele zu erreichen und eine tragfähige Grundlage für weiterführende Entwicklungen im Bereich datenschutzkonformer Datenräume zu schaffen.

6 Nutzen und Verwertbarkeit der Projektergebnisse

Die im Projekt AVATAR erzielten Ergebnisse leisten einen Beitrag zur Weiterentwicklung datenschutzkonformer Datenräume im Gesundheitswesen und zeigen die grundsätzliche Machbarkeit eines souveränen Datenaustauschs mit integrierten Privacy Enhancing Technologies. Insbesondere die prototypische Umsetzung einer Dataspace-Infrastruktur sowie die Integration modularer Anonymisierungs- und Evaluationskomponenten bilden eine Grundlage für weiterführende Forschungs- und Entwicklungsarbeiten.

Ein zentraler Nutzen besteht in der gewonnenen Erfahrung bei der Integration heterogener Komponenten innerhalb einer Dataspace-basierten Architektur. Die Arbeiten lieferten Erkenntnisse zu Schnittstellen, Integrationsaufwand und Architekturentscheidungen, die unmittelbar in Folgeprojekte und zukünftige Plattformentwicklungen einfließen können. Darüber hinaus wurde gezeigt, dass die Einbindung von Privacy Enhancing Technologies als modulare Dienste grundsätzlich geeignet ist, um Datenschutzanforderungen bereits im Datenzugriffsprozess zu berücksichtigen.

Die vom Fraunhofer IDMT bereitgestellten PET-Komponenten, insbesondere zur Anonymisierung strukturierter und unstrukturierter Daten sowie zur Privacy-Evaluation, stellen wiederverwendbare Bausteine dar, die in zukünftigen Projekten weiterentwickelt und in unterschiedliche Anwendungskontexte übertragen werden können. Gleiches gilt für die im Projekt erprobten Ansätze zur Integration spezialisierter Anonymisierungsverfahren für Biosignal- und Bilddaten. Die im Projekt erarbeiteten erste konzeptionelle Arbeiten zu einer PET-Ontologie, die perspektivisch eine modellgetriebene Auswahl und Parametrisierung von Privacy-Verfahren unterstützen kann.

Auch die Arbeiten zur Dataspace-Infrastruktur und die prototypische Umsetzung im Reallabor besitzen Verwertungspotenzial, insbesondere im Hinblick auf die Einbindung in europäische Datenrauminitiativen wie GAIA-X und den European Health Data Space. Die gewonnenen Erkenntnisse können in zukünftige Implementierungen souveräner Datenräume einfließen und zur Weiterentwicklung interoperabler und datenschutzkonformer Plattformlösungen beitragen.

Darüber hinaus liefert der im Projekt entwickelte Demonstrator für Secure Federated Learning Impulse für zukünftige Forschungsarbeiten zur datenschutzfreundlichen Datenanalyse. Auch wenn der Demonstrator nicht Bestandteil des integrierten Dataspace-Workflows war, zeigt er mögliche Erweiterungsperspektiven für verteilte Lernverfahren im Kontext sensibler Gesundheitsdaten auf.

Insgesamt liegt der Mehrwert der Arbeiten sowohl in den entwickelten Komponenten als auch in den Integrations- und Architekturkonzepten, die im Projekt erprobt wurden. Die prototypische Zusammenführung von Dataspace-Infrastruktur und modularen PET-Komponenten lieferte belastbare Erfahrungen zu Schnittstellen, Orchestrierung und Evaluationsmechanismen, die unmittelbar in Folgeprojekte und weiterführende Plattformkonzepte eingebracht werden können.

7 Abgrenzung der Projektergebnisse zum Stand der Technik

Bestehende Referenzarchitekturen für föderierte Datenräume, etwa im Kontext von GAIA-X oder dem International Data Space, adressieren primär Aspekte wie Datensouveränität, Identitätsmanagement und sichere Datenübertragung. Privacy Enhancing Technologies werden dabei häufig nur konzeptionell betrachtet oder als optionale Erweiterung eingeordnet.

Das Projekt AVATAR geht darüber hinaus, indem Anonymisierung und Privacy-Evaluation prototypisch in die Datenflüsse einer Dataspace-basierten Infrastruktur integriert und im Rahmen eines Demonstrators validiert wurden. Darüber hinaus wurden unterschiedliche PET-Ansätze für strukturierte Daten sowie für Biosignal- und Bilddaten kombiniert und exemplarisch eingesetzt.

Ein weiterer Unterschied zum Stand der Technik liegt in der integrierten Betrachtung von Dataspace-Infrastruktur und Privacy-Technologien im Reallabor, wodurch praktische Integrationsaspekte und Herausforderungen sichtbar wurden. Ergänzend liefern die im Projekt angestoßenen Arbeiten zur strukturierten Beschreibung und Kombination von Privacy-Verfahren Ansatzpunkte für eine Weiterentwicklung modellgetriebener Privacy-Konzepte.

8 Erfolgte oder geplante Veröffentlichungen des Ergebnisses

Im Rahmen des Projekts AVATAR erfolgte die Dissemination der Ergebnisse überwiegend in Form projektinterner Präsentationen sowie im Austausch mit den Projektpartnern. Die entwickelten Konzepte und prototypischen Umsetzungen wurden regelmäßig im Konsortium vorgestellt und diskutiert und dienten als Grundlage für die gemeinsame Weiterentwicklung der Architektur und der technischen Komponenten.

Darüber hinaus wurden ausgewählte Projektergebnisse in projektbezogenen Workshops und internen Demonstrationen präsentiert, insbesondere im Zusammenhang mit der Validierung des Reallabors und der Integration der Privacy Enhancing Technologies in die Dataspace-Infrastruktur.

Die Ergebnisse bilden eine gute Grundlage für zukünftige Veröffentlichungen in Folgeprojekten.