# Gaia-X 4 KI
# Whitepaper Datenräume

**Vertrauenswürdige Daten- und Diensteökosysteme**

Eine Publikation aus dem Projekt

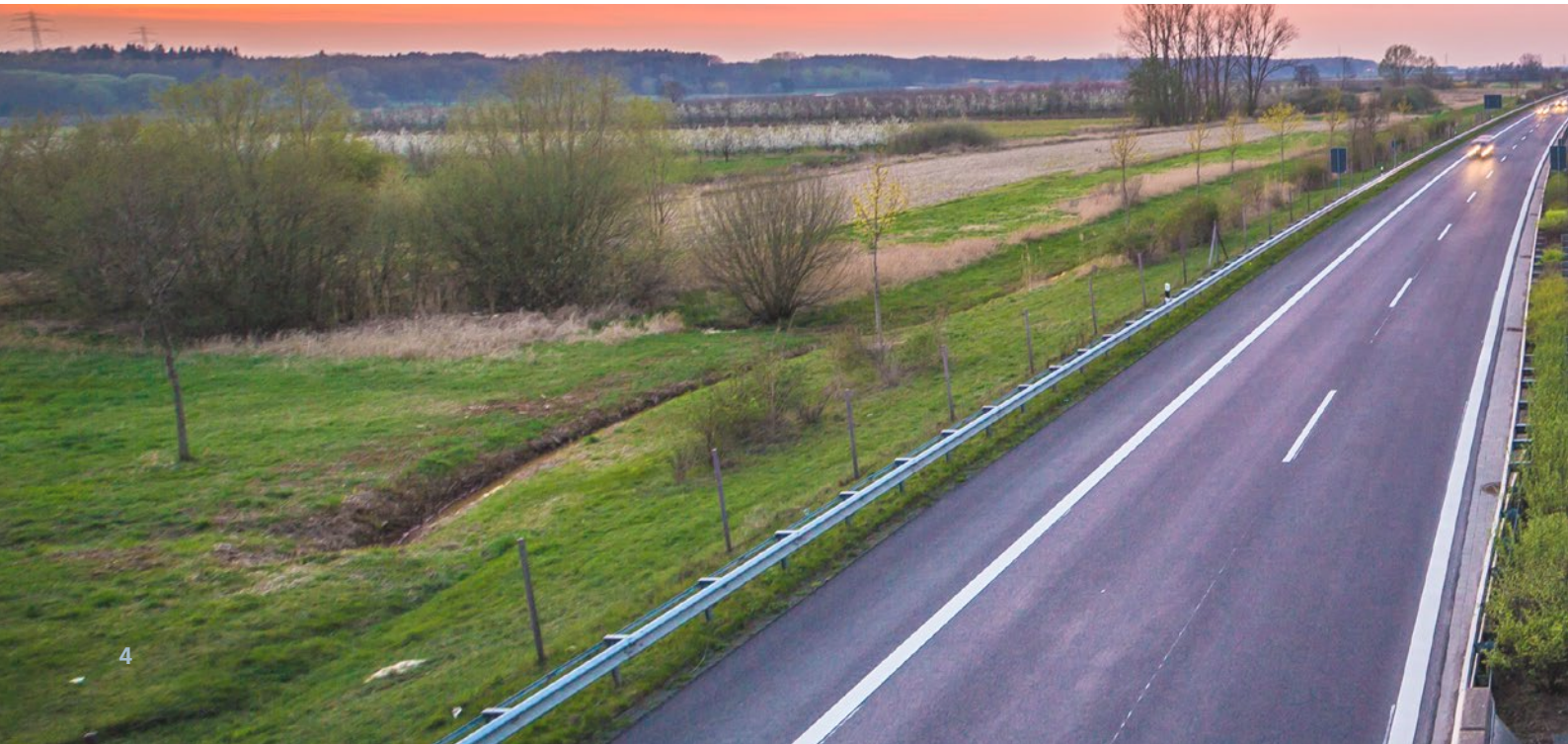
gaia-x 4 KI

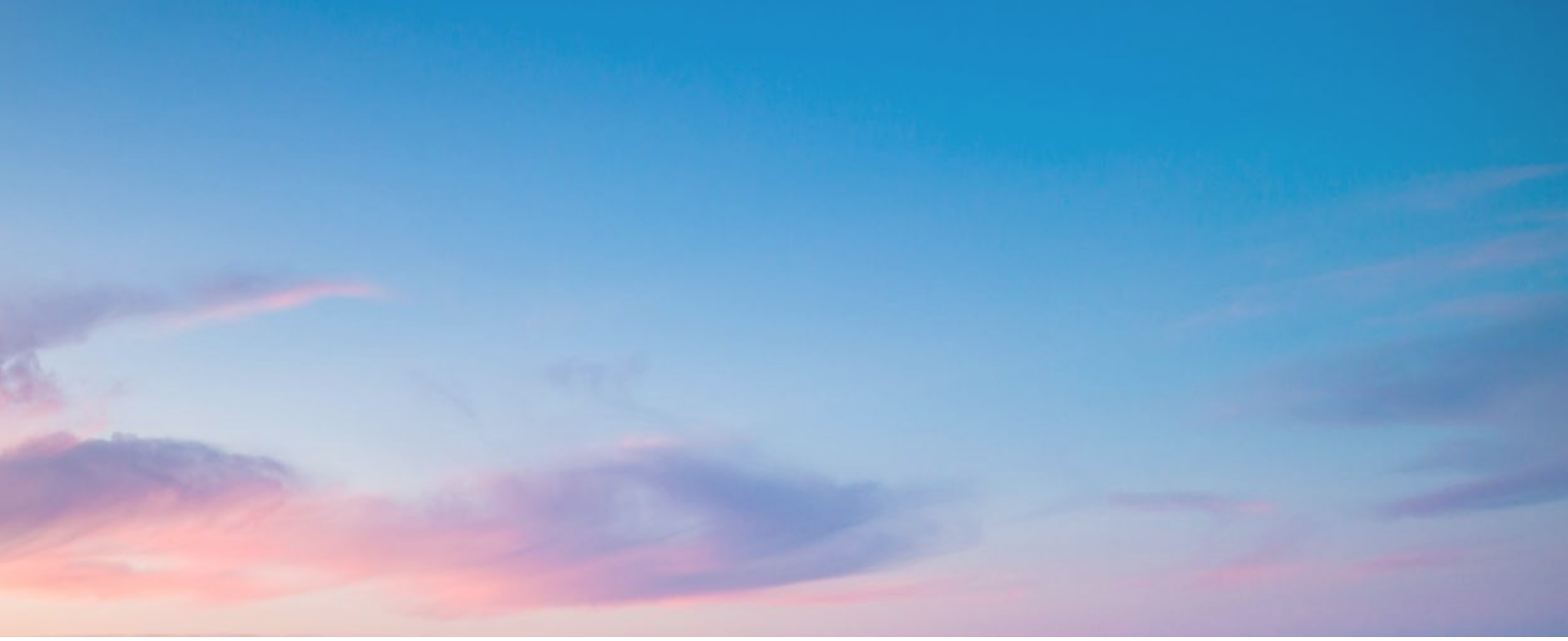
# Inhalt

# Einleitung

## Management Summary

Datenräume schaffen einen Vertrauenskontext um Daten kontrolliert verfügbar zu machen. Damit eignen sie sich insbesondere für die industrielle Produktion und das autonome Fahren, da sensible Daten entlang des gesamten Entwicklungspfades geteilt werden können. Dieses Dokument erklärt das Konzept von Datenräumen in einem domänenspezifischen Kontext und illustriert die Anwendung für innovative KI-basierte Anwendungen im Bereich Mobilität und Automotive. Dazu werden Rollen und durch Datenräume adressierte technische und rechtliche Herausforderungen beleuchtet. Im zweiten Teil des Dokuments wird dann eine Komponenten-basierte Blaupause für einen dezentralen Datenraum zur Verfügung gestellt. Mit einem dezentralen Datenraum wird größtmögliche Souveränität und Eigenständigkeit aller Teilnehmer geeignet. Mit dem beschriebenen Aufbau können verschiedenste Datenräume aufgebaut werden oder miteinander verbunden, so dass es sich für große und komplexe Szenarien eignet.





# Ziele des Dokuments

Dieses Dokument erklärt Datenräume und ihre Bausteine und Komponenten und gibt einen Überblick über die verschiedenen Schritte zum Aufbau und Nutzen von Datenräumen. Mit dem konkreten Anwendungsfall bewegt sich das Dokument hiermit vom Allgemeinen zum Konkreten, in dem Entscheidungen für einen dezentralen, industriellen Datenraum unter Verwendung der Eclipse Dataspace Components (EDC) aufgezeigt werden. Auch werden die Gaia-X Prinzipien und Vertrauensmechanismen genutzt.

Nach einer kurzen Einleitung in das Themenfeld Datenräume und den domänenspezifischen Herausforderungen werden konkrete Architekturentscheidungen, Technologien und open-source Softwarekomponenten sowie Erweiterungen verwendet, um den Aufbau und die Anwendung entlang des gesamten Lebenszyklus eines Datenraums zu illustrieren.

Diese Veröffentlichung richtet sich an Entscheidungsträger mit technischem Hintergrund, wie z. B. CIOs, und dient als Leitfaden für Unternehmen, Institutionen und Einzelpersonen, die beginnen sich über Datenräume zu informieren, um einen eigenen Datenraum einzurichten.

# Motivation und Kontext

Datenräume sind ein Schlüsselfaktor für die Daten- und Digitalwirtschaft, die auf einem effektiven Nutzen von Daten bei gleichzeitiger Datensouveränität, also Selbstbestimmung und Autonomie über die Datenverwendung, aufbaut: Sie ermöglichen Unternehmen einen vertrauenswürdigen und selbstbestimmten Datenaustausch durchzuführen, um gemeinsame Ziele zu erreichen und Daten zwischen verschiedenen Organisationen, Systemen, und sogar Rechtsräumen in einem sicheren und transparenten Weg zu ermöglichen. Damit bilden Datenräume die technische Grundlage für offene Datenökosysteme.

Das autonome Fahren und die moderne Fertigung der Automobilindustrie haben durch die fortschreitende Digitalisierung nicht nur ein hohes Potential zur Optimierung, sondern auch die Möglichkeit Mehrwehrte zu generieren, die über die reine Wirtschaftlichkeitsbetrachtung hinausgehen: Sie umfassen das optimierte Nutzen und Schonung von natürlichen Ressourcen, ein frühzeitiges Erkennen oder gar Vermeiden von Produktionsfehlern, eine optimierte Wartung und die Vermeidung von Unfällen. Durch die Digitalisierung von Abläufen wird ein intelligentes, vernetztes Handeln ermöglicht, das für viele Stakeholder Vorteile bereithält. Insbesondere der Einsatz von künstlicher Intelligenz (KI) kann neue, innovative Dienste ermöglichen und Potenziale eröffnen.

Zwar gibt es bereits umfassende und etablierte Verfahren und Technologien, um Daten zu erfassen und Werkzeuge, um KI einzusetzen. Jedoch stellt die nahtlose Verfügbarkeit der richtigen Daten zum richtigen Zeitpunkt, der Umgang mit sensiblen Daten und den damit verbundenen Rechten, sowie der große und heterogene Datenmengen, das Zusammenführen und die Interoperabilität zwischen einzelnen Systemen wiederkehrende Herausforderungen dar. Dabei spielen insbesondere Vertrauensmechanismen und das maschinelle Verwalten von Bedingungen und Berechtigungen eine Rolle, da ein Datenökosystem viele verschiedene Teilnehmer umfasst, die sich mitunter weder vorab kennen noch unbedingt Vertragspartner sind. Häufig stehen die Teilnehmer sogar in einem Konkurrenzverhältnis zueinander und das Umfeld ist von großem Misstrauen geprägt. Ein fehlerhaftes oder missbräuchliches Verwenden der Daten kann gravierende Konsequenzen für die beteiligten Unternehmen haben.

Die Datenwertschöpfungskette und insbesondere die KI-Wertschöpfungskette umfasst nicht nur verschiedene Daten mitsamt ihren unterschiedlichen Datenspeicherungen und Verarbeitungsweisen, sondern auch eine Vielfalt an unterschiedlichen Diensten, Organisationsstrukturen und historisch gewachsene Systemlandschaften und etablierte Cloud- und Edge Landschaften, die miteinander interagieren müssen. Neben den technischen und organisatorischen Herausforderungen begleiten verschiedene rechtliche Fragestellungen und Vorgaben das gesamte Informationssystem und die verschiedenen Prozessschritte.

Das Ziel von Datenräumen ist es, neue, datengetriebene Dienste zu ermöglichen und ihre Nutzung attraktiver zu machen, indem Vertrauensmechanismen etabliert werden der Zugang zu Diensten ohne das Eingehen von Abhängigkeiten ermöglicht wird. Dabei stehen nicht mehr nur einzelne Stakeholder und individuelle Geschäftsmodelle im Fokus, sondern es soll durch den Aufbau einer neutralen, vertrauenswürdigen technischen Umgebung für ein ganzes Netzwerk an Stakeholdern ein fairer Wettbewerb geschaffen werden. Für die Realisierung eines solchen Netzwerks müssen Herausforderungen adressiert werden, die für einzelne Akteure nicht im Alleingang lösbar sind.

Beim Aufbau und der Verwendung der Lösungen treten verschiedene Aspekte zum Vorschein, die Abstimmungen und Ausarbeitungen in eigenen gemeinnützigen Vereinen und Gremien motivierten, um in großen Communities und breit gefächerten Konsortien gemeinsame Lösungen zu suchen, oder sogar geeignete technische Lösungsansätze gemeinschaftlich bereitzustellen. Diese gemeinsame Gestaltung und Netzwerkbildung verschiedener Organisationen hat sich zu einem prägenden Begleitphänomen von Datenraumtechnologien entwickelt. Parallel zu diesem Phänomen trieb insbesondere die EU-Kommission den Data Governance Act, sowie EU Data Act voran, um die EU-Datenstrategie mit entsprechenden regulatorischen Maßnahmen zu flankieren. Somit gibt es bereits Grundsteine um die aufkeimenden Ökosysteme aus Technologie, Recht, und datengetriebenen Geschäftsmodellen in Zukunft noch weiter wachsen zu lassen. Für den Aufbau und den Betrieb von Datenräumen stehen mittlerweile verschiedene open-source Lösungen und Konzepte bereit, die sich

in Teilen bereits als de-facto Standards etabliert haben und in formale Standards gegossen werden. Damit ist eine langfristige und zukunftsfähige Einsatzfähigkeit gesichert.

Eine besondere Herausforderung in Gaia-X 4 KI ist hierbei, dass die Stakeholder-Situation von verschiedenen Vertrauensstrukturen und Abhängigkeiten charakterisiert ist. Diese umfassen sowohl technische Abhängigkeiten und sogar Einschränkungen, aber auch organisatorische Hürden durch die Vielzahl von miteinander verwobenen Geschäftspartnern. Die umfassen dabei nicht immer nur den Austausch zwischen Unternehmen, sondern innerhalb von Großunternehmen, z.B. über Standorte und Unternehmensbereiche hinweg. Dazu spielt die rechtliche Situation und die häufig damit einhergehenden Unklarheiten eine wesentliche Rolle. Bestehende rechtliche Herausforderungen rund um den Datenschutz und Datensicherheit werden um die Regularien im Datenökosysteme ergänzt.

# Ziele und Charakteristika von Datenräumen

## Kurzeinführung in das Thema Datenräume

Der Begriff Datenraum bezeichnet umgangssprachlich ein Zusammenkommen von verschiedenen Organisationen in einem geschäftlichen Kontext um inter-organisationale Probleme des Datenaustausches, der Datenverarbeitung und der kollaborativen Wertschöpfung zu adressieren. Genauer betrachtet haben Datenräume jedoch eine weitaus spezifischere Bedeutung als Kooperationsformat für den Austausch von Daten unter einer gemeinsamen Governance. Auch die technische Bedeutung ist konkreter:

Ein Datenraum ist ein vertrauenswürdiges und verteiltes Datenökosystem, das es den Teilnehmern ermöglicht, Peer-to-Peer-Daten auszutauschen und gleichzeitig Autonomie und Datensouveränität zu fördern. Ein Datenraum schafft einen auf Attributen und individuellen Bedingungen basierenden Vertrauenskontext, in dem andere Teilnehmer und ihre Datenbestände auffindbar sind. Dabei sind Dataspace-Kommunikationsprotokolle der Eckpfeiler für die Erleichterung von Vertragsabschlüssen durch den Austausch von Metadaten, die Orchestrierung der Vertragsabwicklung zwischen den Teilnehmern und die Nutzung von Datenübertragungstechnologien und Technologien zur Nachverfolgung des Austauschs. Mit diesen Funktionalitäten können Datenraumtechnologien als Werkzeug eingesetzt werden, um komplexe Governance- und Datenhandhabungsrichtlinien einzuhalten.

Von dieser technischen Perspektive heraus betrachtet sind Datenräume Konzept zur Datenintegration und dem Zusammenführen von Daten, das durch eine Reihe wesentlicher Dienste und Vereinbarungen ermöglicht wird. Wesentlich ist dabei die dezentrale Datenhaltung, also das Verbleiben der Daten an ihrem ursprünglichen Speicherort. Die Daten werden erst geteilt, wenn die beteiligten Akteure sich situationsbezogen auf beidseitige Bedingungen geeinigt haben. Auch die Verschachtelung und Überlappung von verschiedenen Bedingungen und Datenangeboten ist möglich, um die Informationsbedarfe und Bedingungen aller Beteiligten zu erfüllen.

**Ein Datenraum ist definiert als ein vertrauenswürdiges und verteiltes Datenökosystem, das es den Teilnehmern ermöglicht, Peer-to-Peer-Daten auszutauschen und gleichzeitig Autonomie und Datensouveränität zu fördern.**

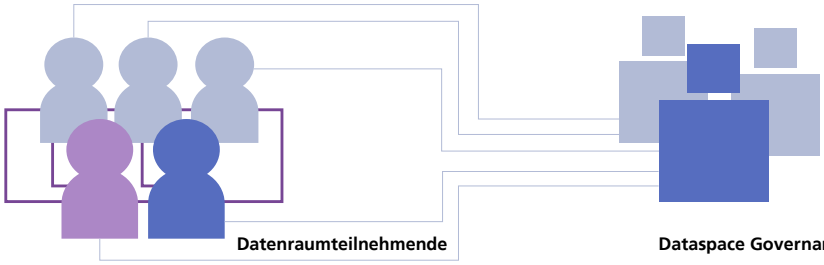
# Rollen und Akteure

Das Konzept des Datenraums impliziert ein generisches Set an konzeptuellen und technischen Rollen, die in jedem Datenraum vorhanden sein müssen und auch in den funktionalen Komponenten eines Datenraums reflektiert sind. In der konkreten Anwendung werden die Rollen dann von ein oder mehreren Personen, Gruppen, oder auch technologischen Agenten oder Diensten eingenommen. Um die verschiedenen Datenraumrollen und ihre Ausführung in einem konkreten Kontext zu erklären, werden im Folgenden zuerst die Grundlagen umrissen und dann mögliche Szenarien aus dem Bereich des autonomen Fahrens und der Produktion hinzugezogen.

## Rollenkonzepte in Datenräumen

**Generische Rollen**

Die Rollen innerhalb eines Datenraums lassen sich grob in die Kategorien Dataspace Governance Authority und die Datenraumteilnehmer aufteilen. Während die Datenraumteilnehmer die aktiven Nutzenden des Datenraums sind und den Datenraum im Falle von Informationsbedarfen hinzuziehen, hat die Rolle der Dataspace Governance Authority mit dem Design und dem fortlaufenden Ausführen der Governance während des gesamten Datenlebenszyklus eine Aufgabe inne. Da in der Praxis Datenräume häufig aus Konsortien und einem kollaborativen Projektrahmen heraus entstehen, sind die initialen Datenraumteilnehmer meist gemeinsam in dem Design beteiligt und nehmen praktisch zumindest initial auch die Rolle der Dataspace Governance Authority ein.



**Datenraumteilnehmende**

- natürliche Personen, rechtliche Personen, Akteure wie z. B. IoT Devices oder Roboter
- aktiv involviert in bereitsteller/abrufen von Daten und anderen Assets
- verfügen über notwendige Technologien, um mit Authorities und anderen Teilnehmenden zu interagieren
- Ziele: Daten aus großem Ökosystem beziehen undkontrolliert teilen
- Prioritäten: Kontrolle über Daten und Vertrauensmechanismen

**Dataspace Governance Authorites**

- technische Rolle: Bereitstellen von Diensten
- Kann z.B. durch eine Organisation, einen Verein, ein Konsortium geführt und verwaltet werden
- Teilnehmende können (und idealerweise sollen) sich mit mehreren Authorities und somit Datenräumen verbinden
- verarbeitet nur Metadaten und vertrauensrelevante Informationen, z. B. zur Identifizierung
- gibt technischen Zugang zum Datenraum unter definierten Bedingungen
- führt Mitgliederliste

*Abbildung 1: Generische Rollen innerhalb eines Datenraums und ihre grundlegenden Aktivitäten*

**Rollen in Gaia-X 4 KI**

Die Datenraumteilnehmer umfassen dabei, am Beispiel des Projektes Gaia-X4KI alle Akteure, die an der Datenwertschöpfungskette beteiligt sind: Von Herstellern, die Daten über ihre Produkte analysieren und weiterverarbeiten möchten, über Dienstleister, die verschiedenste Arten der Daten Verarbeitung anbieten, oder Technologieanbieter, die notwendige Technologien anbieten, um im Datenraum teilzunehmen oder die Daten für das anschließende Teilen bereitzustellen und miteinander zu verlinken. Trotz der Vielzahl an beteiligten Akteuren und mitunter langen Prozessketten über mehrere Datenraumteilnehmer hinweg, besteht im Grunde jede Interaktion zwischen zwei Teilnehmern. Ein Teilnehmer nimmt die Rolle des Datenanbieters ein, und der andere die Rolle des Datenkonsumenten.

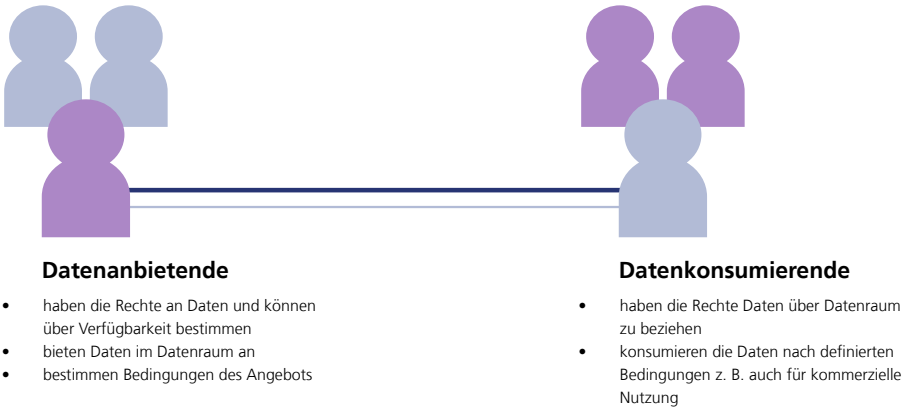



*Abbildung 2: Fokus auf die generischen Rollen der Datenanbietenden und Datenkonsumierenden*

Ein Beispiel für die Verbindung verschiedener Teilnehmer ist in dem unten beschriebenen Szenario zu sehen. Hierbei wird dargestellt, wie der Datenraum die Simulationsergebnisse unterstützt. Dabei werden Dateneigentümer, zumeist Hersteller von Fahrzeugen, mit Dienstleistern für die Datenaufbereitung wie Anonymisierung oder Annotation verbunden und ein Infrastrukturanbieter für die rechenintensive Ausführung der Simulation eingebunden.

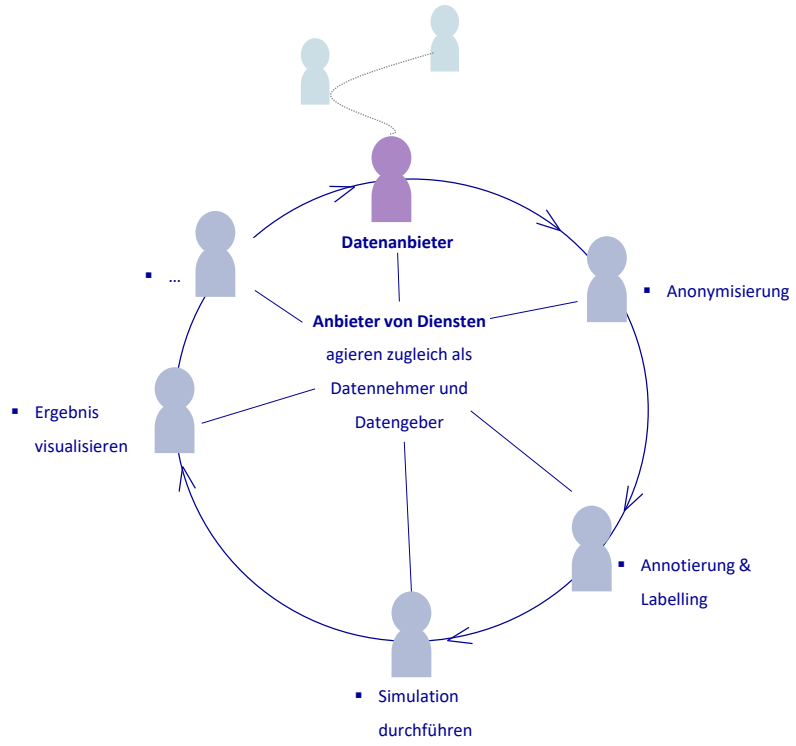



*Abbildung 3: Beispielhafte Verkettung von Datennehmenden und Datennehmenden*

Die Teilnahme an einem Datenraum erfordert gewisse technische Voraussetzungen, zum Beispiel um Authentifizierungen durchzuführen oder Regeln und Bedingungen angeben zu können. Die benötigten technischen Voraussetzungen können Teilnehmende eigenständig bereitstellen, oder sie durch Datenraumanbieter beziehen. Diese Dienstleister und Technologieanbieter können je nach Präferenz der jeweiligen Teilnehmer oder auch Anforderungen der Dataspace Governance Authority ausgewählt werden. Dazu können sie beispielsweise beratend unterstützen, aber auch komplette Lösungen als Dienst bereitstellen. Als intermediär können sie so die Datenaustauschbeziehung zwischen zwei Teilnehmern herstellen, ohne in den direkten Prozess eingebunden zu sein.

Das Gaia-X 4 KI Projekt zeigt die konkrete Bedeutung der Rollen im Datenraum und die Möglichkeit zur Realisierung komplexer Geflechte auf. Als kollaboratives, domänenübergreifendes Ökosystem zielt der Datenraum auf Automobilhersteller und ihre Zulieferer, Forschungseinrichtungen, Technologieunternehmen, Infrastruktur- und Dienstleistungsanbieter, sowie möglicherweise auch öffentliche Stellen als Datengeber ab - wobei einige gleichzeitig mehrere Rollen innehaben. Zum Beispiel werden in einem Anwendungsfall alle Schritte einer KI-Trainingspipeline abgedeckt, also den verschiedenen Teilschritten und Aufgaben um ein Problem mittels KI zu lösen. Dies erfordert unter anderem das Training des KI-Modells, sowie den anschließenden Austausch der Daten. Der Zugang zu einem Datenpool unter definierten Bedingungen erhöht die Qualität der resultierenden KI-Analyse, und die Analyseergebnisse können wiederum über den Datenraum zur Verfügung gestellt werden.

Neben den Datengebern und Datennehmern gibt es eine Vielzahl an Anbietern von Infrastrukturdiensten, die zum einen die erforderlichen Datenraumtechnologien, aber auch den Transfer und die Speicherung der Datenangebote ermöglichen.


*Abbildung 4: Beispiele an Datennehmenden und Datengebenden im Kontext von Gaia-X 4 KI*

# Rechtliche Rollen und ihr Zusammenspiel

Neben dem technischen Rollenkonzept haben Datenraumteilnehmer auch eine juristische Rolle, denn die technischen Funktionen der am Datenraum beteiligten Akteure sind von rechtlicher Bedeutsamkeit. Der Datenmarktplatz im Sinne einer dezentralen Plattform ermöglicht den Handel mit personenbezogenen (DSG-VO) und nicht personenbezogenen (free flow of Data-VO) Daten [1]. Zu differenzieren ist daher zunächst zwischen den Datenarten, die für die Datenraumteilnehmenden (natürliche Person, juristische Person i.S. privater Unternehmen oder öffentlicher Einrichtungen) des Datenökosystem von rechtlicher Relevanz sind.

Im Wesentlichen können folgende Daten unterschieden werden:

(1) Personenbezogene Daten – Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person beziehen.
(2) Anonymisierte Daten – Aufhebung des Personenbezuges, so dass eine Re-Identifizierung der Person nur mit unverhältnismäßigem Aufwand möglich ist
(3) Nicht personenbezogene Daten – Informationen, die sich nicht auf eine identifizierte oder identifizierbare Person beziehen
(4) Technische Daten – Informationen, die sich auf einen Gegenstand/Dienst beziehen und
(5) Offene Daten – Daten, die ohne Einschränkungen genutzt werden dürfen [2].

**Europäische Datenschutzgrundverordnung (DSG-VO)**
Die DSG-VO vereinheitlicht den Datenschutz in Europa und verfolgt primär den Schutz personenbezogener Daten sowie die Sicherstellung des freien Verkehrs personenbezogener Daten. Die DSG-VO legt die datenschutzrechtlichen Anforderungen an die Verarbeitung personenbezogener Daten durch private Unternehmen und öffentliche Stellen fest. [3]. Der Begriff „personenbezogene Daten" bezeichnet alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person beziehen; als identifizierbar wird eine natürliche Person angesehen, die direkt oder indirekt, insbesondere mittels Kennungszuordnung, wie z.B. Name, Standortdaten oder physische Merkmale, identifiziert werden kann. [3]. Der Schutz personenbezogener Daten wird durch Verarbeitungsgrundsätze, z.B. Zweckbindung, Speicherbegrenzung oder Vertraulichkeit ermöglicht.

**ePrivacy-VO**
Die ePrivacy-VO befindet sich derzeit noch im Rechtsetzungsverfahren und wird die DSG-VO im Sinne einer spezielleren Regelung (lex specialis) präzisieren und ergänzen [3]. Der Anwendungsbereich bezieht sich auf elektronische Kommunikationsdaten, die als personenbezogene Daten einzustufen sind.

**EU Data Act**
Europäische Datenräume sollen die sichere Nutzbarkeit von Daten, insbesondere auch personenbezogener oder sensibler Geschäftsdaten, ermöglichen und den Zugang zu Industriedaten, die z.B. aus IoT-Anwendungen stammen, erleichtern. Die Anforderungen an die Datenwertschöpfung sind im Data Act dargelegt, der Rechtsvorschriften hinsichtlich des Datenschutzes als auch der besseren Datennutzung im Rahmen horizontaler Regelungen bezüglich der Rollenverteilung auf Dateninhaber, Nutzer und Dritten etabliert. Der Regelungsbereich des EU Data Act erstreckt sich auf personenbezogene und nicht-personenbezogene Daten. [4]. So haben z.B. in erster Linie Produkt- und Dienstnutzer die Verfügungs- und Nutzungsrechte an den generierten Daten, so dass vernetzte Produkte so zu konzipieren und herzustellen sind, dass Nutzer wie Unternehmen oder Verbraucher einfach und sicher auf die erzeugten Daten zugreifen, diese verwenden und teilen können. Zu den weiteren Regelungen gehören die Erleichterung des Wechsels zwischen Cloud-Anbietern, Garantien für den Schutz nicht-personenbezogener Daten vor unrechtmäßiger Übermittlung an oder staatlichem Zugriff aus Drittstaaten sowie Vorgaben zur Interoperabilität.

**Data Governance Act (DGA)**
Mit dem DGA hat die Europäische Kommission den Grundstein für die Schaffung eines europäischen Modells für die gemeinsame Nutzung von Daten gelegt. Der DGA regelt unter anderem eine verbesserte Interoperabilität sowie einen neutralen Datenzugang und legt Rahmenbedingungen für Datenanbieter und Datennutzer, z.B. für die Weiterverwendung öffentlicher oder geschützter Daten in verschiedenen Sektoren mit der Intention fest, das Vertrauen in Datenvermittlungsdienste und Datenökosysteme zu stärken, um so die Entwicklung und Interoperabilität gemeinsamer europäischer Datenräume in bestimmten strategischen Wirtschaftssektoren und Bereichen von öffentlichem Interesse (z.B. industrielle Fertigung, Mobilität) sicher zu stellen. Etabliert werden u.a. ein Registrierungs- und Aufsichtsrahmen für Datenvermittler und eine freiwillige Registrierung von altruistischen Datenorganisationen für Datenspender (freiwilliges Zurverfügungstellen von Daten), um die Datenquantität als auch -qualität zu fördern. Weiterhin wird die Nutzung von Daten des öffentlichen Sektors durch nicht-öffentliche Stellen fazilitiert.

# Funktionalitäten und Dienste

Um Vertrauenskonzepte und die Verknüpfung von Teilnehmern in Datenökosystemen zu realisieren, stellen Datenräume grundlegende Funktionen bereit. Die verschiedenen Bausteine sind vom Data Spaces Support Center (DSSC) zusammengefasst [8]. Konkrete funktionale Anforderungen werden von der International Data Spaces Association (IDSA) Im folgenden Abschnitt werden die für Gaia-X 4 KI relevanten Funktionalitäten anhand der Gaia-X Kategorien Identität & Vertrauen, Souveräner Datenaustausch, Katalogisierung zur Auffindbarkeit, und Compliance erläutert.

## 1. Identitäten und Vertrauen

Das oberste Ziel von Datenräumen ist das Schaffen von Vertrauen. Dazu werden Technologien und Mechanismen eingesetzt, die Teilnehmenden eine informierte Entscheidung darüber ermöglichen, welchen Akteuren im Ökosystem sie auf welcher Basis vertrauen möchten. Der Begriff Vertrauen bezieht sich dabei zu einem großen Teil auf das Nutzen von digitalen Identitäten zur Authentifizierung und Autorisierung. Vertrauen kann jedoch darüber hinaus auch in weitere Angaben von Teilnehmern bestehen, zum Beispiel in die Angabe des Firmensitzes.

Zur Umsetzung einer vertrauenswürdigen Umgebungen dienen Vertrauensanker und Funktionalitäten zur Prüfung und Bestätigung. Diese können mit dem Alltagsbeispiel eines Führerscheins verglichen werden: Der Führerschein wird vorgezeigt um jemandem die Fähigkeit und Erlaubnis zum Führen eines Fahrzeugs zu bescheinigen. Dieses Prinzip kann auf jegliche Eigenschaften der Teilnehmer in einem Ökosystem angewendet werden.

Die Grundlage für eine vertrauensvolle und selbstbestimmtes digitales Handeln ist die Kontrolle in über die eigene digitale Identität und die freie Entscheidung darüber, welchen Identitäten anderer Akteure Vertrauen geschenkt wird.

Dabei geht es nicht nur um die Entscheidung, ob sich ein Akteur entsprechend ausweisen kann, um einem Datenraum technisch beizutreten, sondern auch um die verschiedenen

Rechte und Pflichten innerhalb eines Datenraums in allen Phasen, von der Datenexploration bis zur Datennutzung und -verarbeitung. Ein Vertrauen in die Identifizierung bildet eine Grundlage für alle weiteren Mechanismen. Dabei muss sich das Identitäts- und Zugriffsmanagement in eine Umgebung einfügen, die eine vielfältige Landschaft von Identitäten aufweist, z. B. nicht nur verschiedene Unternehmen, sondern auch verschiedene Geschäftsbereiche, Abteilungen und technische Benutzer, die alle eine gemeinsame Verwaltung benötigen.

## 2. Souveräner Datenaustausch

Neben Vertrauensmechanismen zur Identifizierung sind Kontroll- und Nachverfolgungsmöglichkeiten über den Datentransfer ein wesentliches Merkmal von Datenräumen. Diese basieren auf der maschinenlesbaren Formulierung, die Bereitstellung und Ausführungsmöglichkeiten von Zugriffs- und Nutzungsbedingungen während des gesamten Datenaustauschs. Insbesondere in der industriellen Anwendung ist aufgrund der erforderlichen Skalierbarkeit die Möglichkeit zur Automatisierung entscheidend.

Ein Missbrauch von Daten könnte zum Beispiel bedeuten, dass ein Datenkonsument sensible Informationen an Konkurrenten weitergibt. Ein solcher Missbrauch kann vorsätzlich oder fahrlässig erfolgen und könnte durch eine Systemumgebung verhindert werden, die eine richtlinienkonforme Datenverarbeitung von vornherein gewährleistet.

Dazu gibt es verschiedene Rollenkonzepte, um ein transitives Vertrauensmodell und attribut-basiertem Vertrauen umzusetzen. Transitives Vertrauen bedeutet, dass eine Vertrauenskette zwischen mehreren Stationen aufgebaut werden kann, die auf ihren Ursprung zurückverfolgt werden kann. Zum Beispiel können so verschiedene Partner an einer KI-Pipeline kollaborativ arbeiten, und es ist nachverfolgbar wer daran beteiligt war. Attribut-basiertes Vertrauen bedeutet, das Vertrauen nicht an fixierten, einzelnen Stellen oder Teilnehmern festgemacht wird, sondern dass relevante Merkmale (Attribute) definiert werden und wer sie belegen kann. Für die Bestätigung von Merkmalen, wie bspw. Einen Unternehmenssitz in Deutschland, der von dem Handelsregister bestätigt ist, kann als ein Vertrauensanker definiert werden.

Aus dem Zusammenspiel mehrerer Attribute können Datenraumteilnehmer ihren individuellen Vertrauensgrad ermitteln, und daraus das damit verbundene Risiko bestimmen. Die beiden Parameter sind für die Freigabe von Daten entscheidend. Der Vorteil des attributbasierten Vertrauens ist es, dass die Möglichkeit und Auswahl der Attribute eine große Flexibilität ermöglichen. So kann jeder Teilnehmer für sich oder für jede Art von Daten bestimmte Attribute verlangen. Unter anderem ist damit auch eine anonyme Teilnahme möglich, da die Authentifizierung nicht immer unter notwendigen Attributen stehen muss. Zur Verifizierung der Attribute werden dezentrale Validierungsmechanismen wie zum Beispiel W3C Verifiable Credentials (VC) genutzt. Für Teilnehmer ist es außerdem möglich zu definieren, welche Validierungsmechanismen für sie akzeptabel sind. Die Komplexität der Implementierung ist dabei stark von Anzahl und Art der Attribute abhängig. Das attributbasierte Vertrauen kann außerdem zusammen mit anderen Vertrauenskonzepten genutzt werden. Nach der Überprüfung und Berechnung des möglichen Risikos werden Beschränkungen eingeführt. Die Beschränkungen sind durch Policies ausgedrückt. Die Policies bestimmen, wie der Zugriff auf bestimmte Daten erfolgen soll, wer den Zugriff beantragen darf und welche Sicherheitsmaßnahmen unternommen werden müssen.

## 3. Auffindbarkeit, Katalogisierung und Selbst-Beschreibung

Ein weiteres Ziel ist das unabhängige Finden von Geschäftspartnern, das Teilen von Datenangeboten und letztlich den Daten zwischen unterschiedlichen Systemen hinweg.

In einem Datenökosystem geht es im Kern darum, flexibel und unkompliziert einen Datenaustausch mit anderen Teilnehmern zu realisieren. Dies bedeutet, dass effektive Mechanismen zum Auffinden von Datenangeboten, sowie zum vertrauenswürdigen Ausführen der Datenaustauschbeziehung vorhanden sind. Zusätzlich dazu nutzen Datenraumteilnehmer meist verschiedene Infrastrukturen, um ihre Daten und Dienste zu sichern und zu verarbeiten. Dies bedeutet, dass ein Datenraumsystem die Kommunikation zwischen verschiedenen Cloud-Systemen, Edge- und On-Premise-Systemen unterstützen muss.

Ein Mechanismus, um die gegenseitige Auffindbarkeit zu erreichen sind sogenannte Selbstbeschreibungen. Diese Selbstbeschreibungen sind technisch so gestaltet, dass sie nachprüfbar und je nach Bedarf des Datenraumes ausgestaltet sind, zum Beispiel durch das Hinzufügen von spezifischen Attributen. Die Datenraumteilnehmenden haben eine Selbstbeschreibung, dazu gibt es eine Selbstbeschreibung für den jeweiligen Datenraum oder auch Beschreibungen für angebotene Daten und Dienste.

## 4. Compliance– auch in komplexen Szenarien

Die gemeinsame Nutzung von Daten ist mit zahlreichen rechtlichen und regulatorischen Herausforderungen verbunden, deren Einhaltung durch Datenräume unterstützt werden kann. Datenraum-Technologien und Standards ermöglichen es, Daten flexibel auszutauschen. Dadurch tragen sie unter anderem Datensicherheit und einen effektiven Datenschutz. Ferner umfassen sie sowohl die sichere technische Infrastruktur als auch die dazugehörigen Governance-Mechanismen. Die meisten Anbieter virtueller Datenräume erfüllen den ISO/IES 27000-Standard. Die ISO/IEC 27000-Reihe regelt Standards zur Informationssicherheit, die von der International Organization for Standardization (ISO) und der International Electrotechnical Commission (IEC) herausgegeben werden. Ihr Fokus liegt auf Best-Practice-Empfehlungen zur Organisation der Informationssicherheit im Kontext eines Information Security Management Systems (ISMS). Erhält ein Anbieter virtueller Datenräume diese Zertifizierung, kann allgemeingültig davon ausgegangen werden, dass das Unternehmen alles Notwendige tut, um Daten zu schützen, da es von der internationalen Organisation von Standardisation verpflichtet ist. Gleichzeitig begegnen auch Datenräume zahlreichen Fragen rechtlicher Natur: Welche Geschäftsmodelle bspw. unterliegen beim Teilen von Daten der europäischen Gesetzgebung wie dem Data Governance Act und welche rechtlichen Implikationen ergeben sich hieraus für die verschiedenen Teilnehmer an Datenräumen?

### I. Data Governance Act

Der Data Governance Act ist die wohl relevanteste Rechtsquelle für die juristische Einordnung von Datenräumen. Er wird unter anderem die Einrichtung und Entwicklung gemeinsamer europäischer Datenräume in strategischen Bereichen unterstützen, an denen sowohl private als auch öffentliche Akteure in Bereichen wie Gesundheit, Umwelt, Energie, Landwirtschaft, Mobilität, Finanzen, Fertigung, öffentliche Verwaltung und Kompetenzen beteiligt sind. Mit dem Data Governance Act vermag die Europäische Kommission Rahmenbedingungen für Anbieter einer gemeinsamen Datennutzung und für Anbieter von Datenspenden einzuführen. Dazu gehören ein Registrierungs- und Aufsichtsrahmen für Datenvermittler und eine freiwillige Registrierung von altruistischen Datenorganisationen für Datenspender. Der Data Governance Act ist somit als Rahmen gedacht, um das Vertrauen in diese Einrichtungen zu fördern. Ebenso will die Kommission die Entwicklung und Interoperabilität gemeinsamer europäischer Datenräume in bestimmten strategischen Wirtschaftssektoren und Bereichen von öffentlichem Interesse fördern (dies sind: Industrielle Fertigung, Gesundheit, Finanzdaten, Green Deal, Mobilität, Energie, Agrarsektor, öffentliche Verwaltung und Bildung). In diesen neun Bereichen sollen Datenräume gebildet werden, die die Nutzung und den Austausch von Sachdaten durch Behörden, Unternehmen und die Wissenschaft erleichtern. Im Gegensatz zur digitalen Selbstbestimmung, die die vertrauenswürdige Nutzung von Sach- und Personendaten fördern will, setzt die Europäische Kommission im Bereich der Datenräume in einem ersten Schritt vor allem auf die Verknüpfung von Sachdaten.

Der Data Governance Act zielt folglich darauf ab, die Verfügbarkeit von Daten zur Nutzung zu fördern, indem das Vertrauen in Datenmittler gestärkt und Mechanismen zur gemeinsamen Nutzung von Daten in der gesamten EU ausgebaut werden. Das Instrument bezieht sich auf die Bereitstellung von Daten des öffentlichen Sektors zur Wiederverwendung in Fällen, in denen diese Daten den Rechten anderer unterliegen, auf die gemeinsame Nutzung von Daten durch Unternehmen

gegen eine Gebühr in jeglicher Form, auf die Ermöglichung der Nutzung personenbezogener Daten mit Hilfe eines »Vermittlers für die gemeinsame Nutzung personenbezogener Daten«, um Einzelpersonen bei der Ausübung ihrer Rechte gemäß der DSGVO zu unterstützen, und auf die Ermöglichung der Nutzung von Daten aus altruistischen Gründen. Mit dem Data Governance Act hat die Europäische Kommission den Grundstein für die Schaffung eines europäischen Modells für die gemeinsame Nutzung von Daten gelegt. Es soll ab dem 24. September 2023 angewendet werden. Gemeinsam mit dem EU Data Act ist der Data Governance Act Teil der „europäischen Datenstrategie" der EU-Kommission vom 19. Februar 2020.

**II. EU Data Act**

Der besagte EU Data Act stellte die zweite relevante Rechtsquelle neben dem Data Governance Act dar: Angesichts der unbestrittenen Bedeutung von Daten als Wirtschaftsgut und Innovationsmotor hat die EU-Kommission bereits im Jahr 2020 ihre Datenstrategie formuliert und das Ziel ausgegeben, einen echten Binnenmarkt für Daten (Datenraum) zu schaffen, um die europäische Datenwirtschaft zu stärken. Der einheitliche europäische Datenraum soll die sichere Nutzbarkeit von Daten, insbesondere auch personenbezogener oder sensibler Geschäftsdaten, ermöglichen und den Zugang zu Industriedaten, die zum Beispiel aus Anwendungen des Internets der Dinge stammen, erleichtern. In diesem Datenraum soll EU-Recht wirksam durchgesetzt und europäische Standards entwickelt werden, wozu die Kommission am 23.02.2022 ihren Entwurf für ein Datengesetz vorgelegt hat. Zu den Regelungen gehören die Erleichterung des Wechsels zwischen Cloud-Anbietern, Garantien für den Schutz nicht-personenbezogener Daten vor unrechtmäßiger Übermittlung an oder staatlichem Zugriff aus Drittstaaten sowie Vorgaben zur Interoperabilität. Außerdem Regelungen, die den Nutzern eines Geräts weitgehende Verfügungsrechte über die mit ihrer Nutzung verbundenen Daten einräumen (Data Sharing). Wichtig ist in diesem Zusammenhang auch, dass Behörden erstmals einen Anspruch gegen Dateneigentümer in Notsituationen haben werden. Das Datengesetz ist als Verordnung für primär nicht-personenbezogene Daten geplant und regelt u.a. die Nutzung von Daten aus vernetzten Geräten (IoT) und Maschinen. Ziel dieser Verordnung ist es, die Entwicklung neuer, innovativer Produkte oder damit verbundener Dienstleistungen zu fördern, Innovationen in nachgelagerten Märkten anzuregen, aber auch die Entwicklung völlig neuer Arten von Dienstleistungen zu stimulieren, die Daten nutzen, einschließlich solcher, die auf Daten aus einer Vielzahl von Produkten oder damit verbundenen Dienstleistungen basieren. Gleichzeitig soll vermieden werden, dass Investitionsanreize für die Art von Produkt, von dem die Daten stammen, untergraben werden, beispielsweise durch die Nutzung der Daten zur Entwicklung eines konkurrierenden Produkts.

Daneben existieren weitere Rechtsquellen, die für die rechtliche Handhabe von Datenräumen bedeutsam werden könnten:

**III. DSGVO**

Die Europäische Datenschutzgrundverordnung zielt darauf ab, den Datenschutz in Europa zu vereinheitlichen und somit gleiche Datenschutzstandards für alle Mitgliedsstaaten zu schaffen. Sie gilt seit dem 25. Mai 2018 und enthält Bestimmungen zur Verarbeitung personenbezogener Daten durch private Unternehmen und öffentliche Stellen. Die DSGVO verfolgt primär den Schutz der Grundrechte und Grundfreiheiten natürlicher Personen und insbesondere deren Recht auf Schutz personenbezogener Daten (Art. 1 Abs. 2 DSGVO) und den freien Verkehr personenbezogener Daten (Art. 1 Abs. 3 DSGVO). Die vorangestellten Ziele sollen durch die in Art. 5 DSGVO festgelegten Grundsätze der Verarbeitung personenbezogener Daten erreicht werden: Rechtmäßigkeit, Treu und Glauben, Transparenz, Zweckbindung, Datenminimierung, Richtigkeit, Speicherbegrenzung, Integrität und Vertraulichkeit, Rechenschaftspflicht.

**IV. Richtlinie über offene Daten**

Die Richtlinie beruht auf dem allgemeinen Grundsatz, dass öffentliche und öffentlich finanzierte Daten für kommerzielle und nichtkommerzielle Zwecke weiterverwendet werden sollten. Die Richtlinie fördert die Verwendung von Open Data. Öffentliche Einrichtungen und öffentliche Unternehmen müssen ihre Dokumente in allen bestehenden Formaten oder Sprachen und, soweit möglich und angemessen, in elektronischer Form zur Verfügung stellen, die offen, maschinenlesbar, zugänglich, auffindbar und wiederverwendbar sind, komplett mit Metadaten.

**V. ePrivacy Verordnung**

Die ePrivacy Verordnung schafft verbindliche Regelungen für den Umgang mit personenbezogenen Daten in Online-Medien auf EU-Ebene. Die Regelungen sollen das Vertrauen der Bürger in elektronische Kommunikationskanäle stärken und die Rahmenbedingungen für digitale Unternehmen in den EU-Ländern vereinheitlichen. Hintergrund der ePrivacy Verordnung ist eine Gesetzesinitiative der Europäischen Union, die den digitalen Sektor betrifft und auch die DSGVO umfasst.

# Interoperabilität und übergreifendes Vertrauen

## Chancen durch Datenräume

Das Datenraum-Konzept ermöglicht es Teilnehmern je nach Bedarf in verschiedenen, voneinander unabhängigen Datenräumen aktiv zu sein. Vor dem Hintergrund von wachsenden Datenökosystemen, ist eine Teilnahme an mehreren, anwendungsfallspezifischen Datenräumen abzusehen. Deshalb muss Interoperabilität unabhängig von einzelnen Datenräumen beachtet werden und in einen größeren Kontext gesetzt werden. Aus der Sicht der funktionalen Anforderungen ist die Interoperabilität in die folgenden vier Stufen zu unterteilen, die im European Interoperability Framework unterschieden werden [9]. Technische Interoperabilität bedeutet, dass mehrere, unterschiedliche Systeme miteinander verbunden werden müssen, um einen reibungslosen Datenfluss zu ermöglichen.

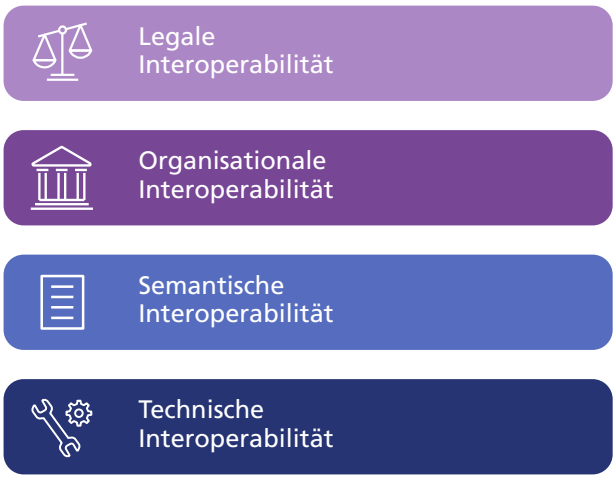



*Abbildung 5: Säulen des European Interoperability Frameworks*

Neben den technischen Aspekten gibt es auch organisatorische Fragen der Interoperabilität. Wie beispielsweise in einem Fußballverein mit Vereins- und Spielregeln gibt es auch für einen Datenraum Regeln. Die Nutzer eines Datenraums müssen einen Vertrag unterzeichnen und ihm zustimmen, in dem die Nutzungsbedingungen festgelegt sind, bevor sie beitreten. Diese Regeln sind unabhängig von der beteiligten Organisationen und ihres Zusammenschlusses, der eine Verbindung aus öffentlich, privat oder öffentlich-privaten Organisationen sein kann. Während die zugrunde liegenden Governance- und Compliance-Regeln sowie insbesondere das Identitätsmanagement die effiziente Interaktion innerhalb eines Datenraums sorgen, können sie erhebliche Hürden für die Interoperabilität zwischen verschiedenen Datenräumen und Ökosystemen darstellen.

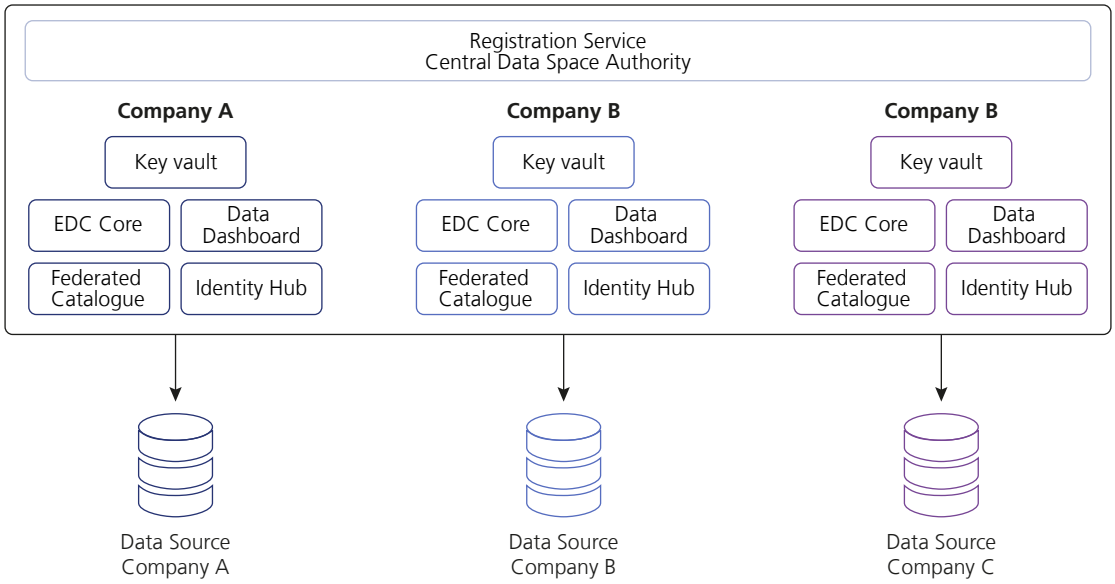



*Abbildung 6: Interoperabilität am Beispiel des MVDaaS by T-Systems*

Gaia-X 4 KI stellt ein Szenario der organisatorischen Interoperabilität dar, das in der Praxis durchaus vorkommen kann. GAIA-X 4 KI, GAIA-X 4 AMS, GAIA-X 4 ROMS, GAIA-X 4 PLC-AAD und GAIA-X 4 moveID sind Teilprojekte von Gaia-X 4 Future Mobility  Projektfamilie. Jedes dieser Projekte würde idealerweise einen Standard-Datenraum inkl. Regeln und Service einrichten, um entsprechende Anwendungsfälle zu ermöglichen. Dabei handelt es sich um allgemeine Regeln, die in jedem Projekt gleich sind und miteinander verknüpft werden, wie z.B. Identitätsmanagement, vertragliche Vereinbarungen oder das erforderliche Maß an Vertrauen, um Teilnehmer des Datenraums zu sein, z.B. spezifische Zertifizierungsanforderungen. Gaia-X 4 KI hat dazu Prozesse und Komponenten ausgearbeitet, die nicht nur im Hinblick auf das Projekt selbst, sondern auch über das Projekt hinaus im Rahmen der Projektfamilie sowie mit Blick auf andere Gaia-X Leuchtturmprojekte nutzbar sind. Damit wird Effizienz und Agilität für die Datenraumteilnehmenden erreicht. Die von T-systems im Data Intelligence Hub (DIH) eingerichtete Datenraum-Benutzerschnittstelle, die den Teilnehmern des GX4KI-Teilprojekts dient, kann somit auch auf andere Projekte und dessen Akteure ausgeweitet werden. Ein Effekt ist, dass einige Konsortialpartner an mehreren Projekten innerhalb der Projektfamilie beteiligt sind und den Onboarding-Prozess jedoch nur einmal durchlaufen müssen. Ihre Identitäten und erforderlichen Software-Komponenten, die in einem Projekt autorisiert sind, können auch in anderen Teilprojekten gleichermaßen anerkannt und genutzt werden.

Auf der semantischen Ebene ist ein gemeinsames semantisches Datenmodell eine ideale Lösung, die jedoch alle Teilnehmer des Datenraums betrifft, erhebliche Investitionen erfordert und sich daher nur sehr schwer durchsetzen wird. Eine greifbare Lösung wird die Harmonisierung der Metadaten-Interoperabilität innerhalb und mit Datenräumen sein. Mögliche Implementierungen werden in Gaia-X 4 KI untersucht, wie z.B. die Konvertierung oder Anpassung von Datenmodellen und semantische Dienste.

# Das Dataspace Protocol

Viele Datenräume können die verwendeten Technologien ihrer Teilnehmer nicht beeinflussen und unterschiedliche Anwendungen bringen unterschiedliche Anforderungen mit sich. Dennoch müssen die verschiedenen Datenraum-Komponenten miteinander interagieren. Um dieses Problem zu beheben, wurde Dataspace Protocol (DSP) definiert. Das DSP ist eine Spezifikation und beschreibt die technische Kommunikation innerhalb eines Datenraums. Besonders wichtig ist, dass sie Unabhängig von konkreten Technologien, Implementierungsweisen, oder Programmiersprachen ist. Komponenten, die dieser Spezifikation folgen, sind interoperabel zueinander. Dazu werden existierende Standards verwendet, sowie einzelne Spezifikationen modular erstellt, um sie ebenfalls modular implementierbar zu halten.

Das DSP strebt drei grundlegende Ziele an:

1. Unterstützung moderner Datentransfer-Anforderungen wie Streaming oder den Umgang mit Big Data durch asynchrone Nachrichtenstrukturen

2. Etabliere Interoperabilität und ein Mix-and-match von verschiedenen Datenraum-Komponenten

3. Etablierung von Standards und Best Practices durch vereinheitlichte Beschreibung und Einrichtung einer Verifikationsstelle für alle normativen Aspekte

Das Dataspace Protokoll bezieht sich auf die technische Interoperabilität. Nach dem ISO-Modell [11] beschreibt diese im Detail die Transport-Ebene und die syntaktische Ebene.

Das DSP fokussiert sich auf die Entdeckung der Datenangebote im Datenraum (Data Discovery) und die maschinenlesbare Aushandlung und Vereinbarung von Bedingungen (Data Contract Negotiation). Für die Auffindbarkeit von Daten schreibt das DSP die Verwendung von dem W3C Standard DCAT vor, einem Vokabular und Schema um Interoperabilität zwischen Datenkatalogen im Web herzustellen [12]. Die Vereinbarung über Bedingungen wird mit der Open Digital Rights Language (ODRL) getroffen. ODRL ist eine Sprache um Regeln (Policies) [13] auszudrücken.

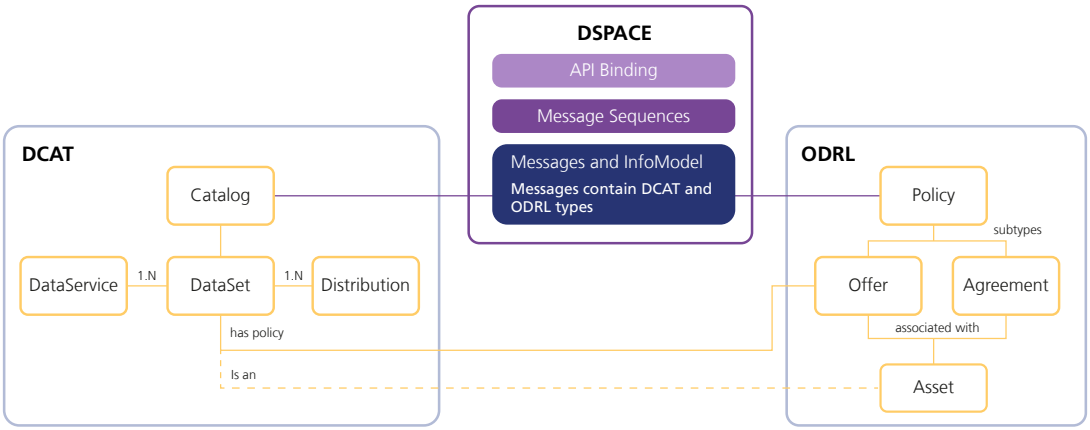



*Abbildung 7: DSP Struktur*

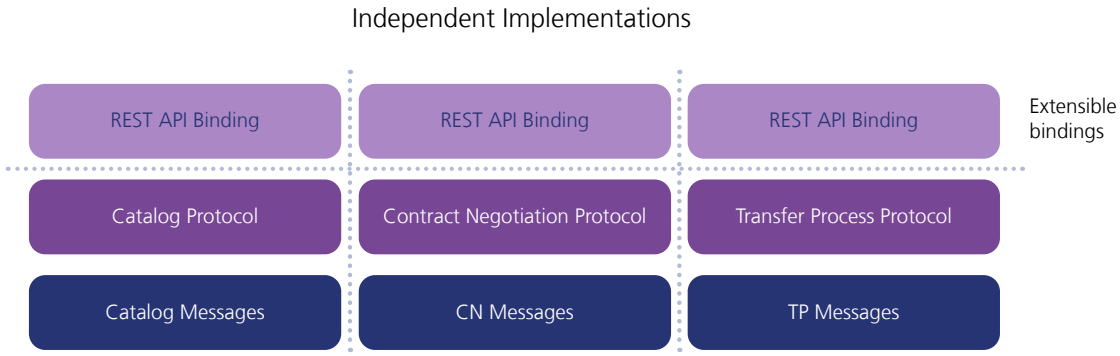
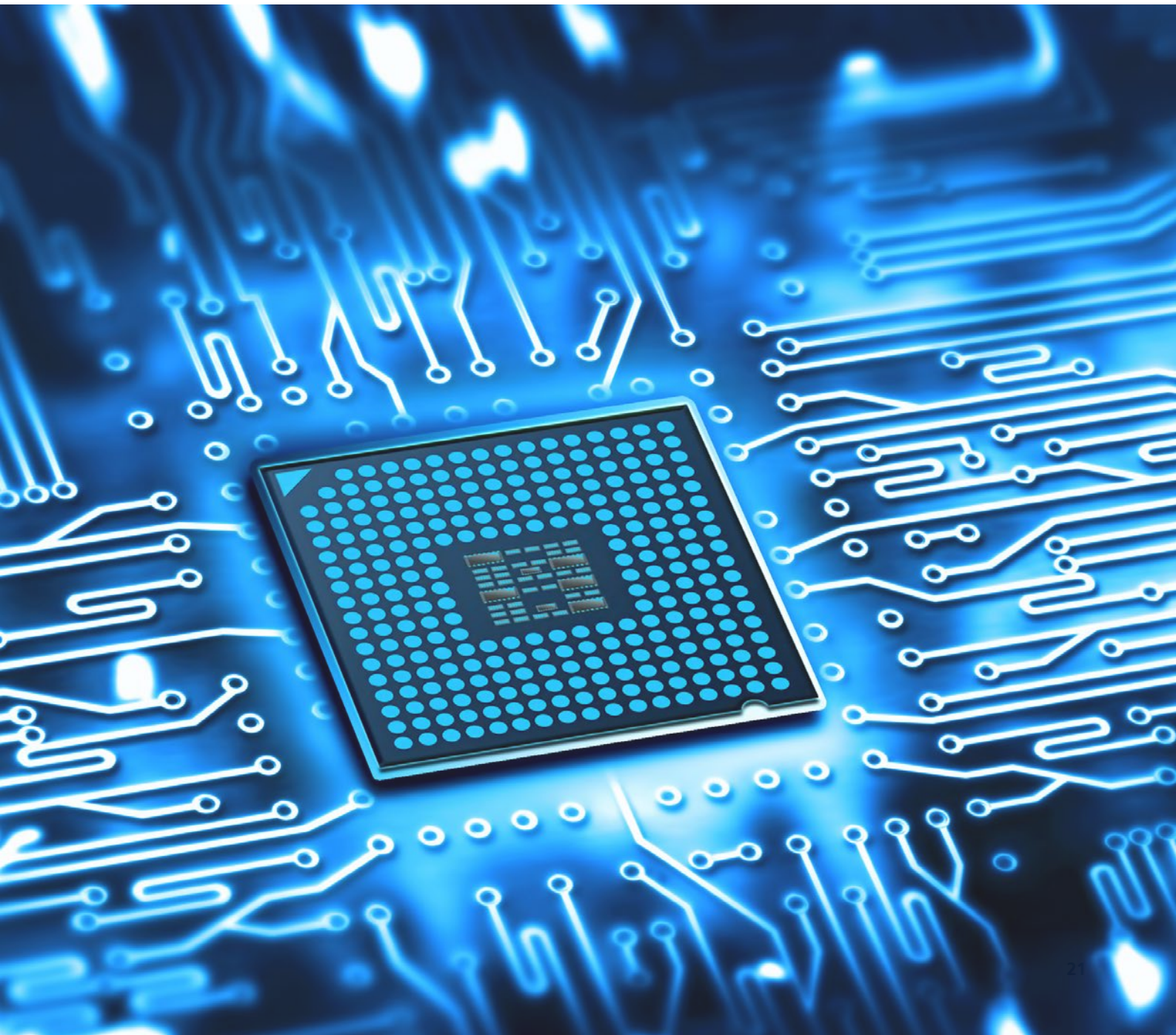



*Abbildung 8: DSP Architektur*



# Wege zum Datenraum

## Aufbau eines Datenraums

**Technische Bereitstellung**
Der Aufbau eines Datenraums kann im Wesentlichen durch zwei Wege erfolgen:

Die erste Option ist die Auswahl von geeigneten open-source Komponenten, die nach Anpassung auf die Bedürfnisse der initialen Datenraum-Gemeinschaft zusammen mit anderen erforderlichen Komponenten verfügbar gemacht werden. Dazu erfolgt nach der Auswahl der benötigten Software-Komponenten und möglicherweise einer Anpassung des Codes die Bereitstellung. Dazu wird ein Anbieter ausgewählt, oder eigenes Hosting durchgeführt. Dies kann vor allem für erste Tests sinnvoll sein und für das Erproben von neuen Datenraumfunktionalitäten.

Die zweite Option ist, out-of-the-Box die notwendigen Dienste und Komponenten von einem oder mehreren vertrauenswürdigen Anbietern zu beziehen, die auf die Bedarfe des Datenraumteilnehmers eingehen und in verschiedenen Liefermodellen notwendige Dienste bereitstellt. Dies ermöglicht den Datenraumteilnehmern, sich rein auf ihr individuelles Kerngeschäft und das Anbieten ihrer eigenen Daten und Dienste zu fokussieren.

Im Gaia-X4KI Datenraum ist der Datenraum umgesetzt die Eclipse Dataspace Components (EDC). Die Komponenten bilden ein Framework für Datenraumtechnologien und sind open-source verfügbar. Sie werden durch einen Projektpartner betrieben und passend auf die Bedürfnisse der jeweiligen Teilnehmer zur Verfügung gestellt. Aber es gibt auch Datenraumteilnehmer, die ihre notwendigen Dienste auf eine eigene Weise bereitstellen und betreiben. Durch die Nutzung eines gemeinsamen de-facto Standards ist es möglich, eine Interoperabilität für die grundlegenden Vertrauensmechanismen sowie auch für den Datentransfer herzustellen.

Das Ziel des Gaia-X4KI Datenraums ist es, die Ziele und Interessen aller Teilnehmer und ihrer individuellen, Anwendungsfall-bedingten Interessen zu berücksichtigen.. Dabei wird eine Balance geschaffen, die zum einen den Datenraumteilnehmern so viel Autonomie und Gestaltungsfreiheit ermöglicht und zum anderen eine Interoperabilität und Berücksichtigung gemeinsam festgelegter Regeln umsetzt.

**Gemeinsame Regeln und Governance**
Eine wesentliche Fragestellung ist, welche Partei der anderen unter welchen Bedingungen vertraut, und somit welche Rahmenbedingungen ein Datenraumbetreiber erfüllen muss um allen Teilnehmern ein vertrauenswürdiges, aber dennoch praktikables und flexibles Umfeld zu ermöglichen.

Ein Grundsatz des Datenraums ist, dass es ein Set an Regeln gibt, die von allen Teilnehmern gleichermaßen eingehalten werden müssen. Diese Regeln können Richtlinien jeder Art umfassen, wie z.B. Zughörigkeit zu einem Verband oder Nachhaltigkeitsanforderungen, aber insbesondere auch technische Anforderungen, um ein Mindestmaß an Kommunikation und Interoperabilität herzustellen. Mit den einmalig definierten Regeln, die für jede Teilnahme und Transaktion automatisiert umgesetzt werden, gibt es keine Konflikte oder willkürlichen Entscheidungen durch eine zentrale Instanz. Durch die Verwendung von gemeinsamen Vertrauensankern und Vertrauensmechanismen wird ein transitives Vertrauen aufgebaut, dass allen Interaktionen zugrunde liegt und rückverfolgbar ist.

# Eintritt in den Datenraum

Teilnehmer können einem Datenraum im Wesentlichen durch zwei Schritte aus einer technischen Perspektive teilnehmen:

1. Eine Mitgliedschaft beantragen bei der Data Space Authority,

2. Auswahl und Bereitstellung der erforderlichen Komponenten, oder auch Auswahl eines entsprechenden Anbieters.

Vor dem ersten Schritt findet meist die Regelung der Mitgliedschaft auf organisatorischer Ebene statt, zum Beispiel in dem initialen Projektkonsortium. Der erste technische Schritt ist anschließend die Bereitstellung eines Identifizierungsmittels, sowie einer Selbstbeschreibung. Diese enthält obligatorische Angaben zum teilnehmenden Unternehmen und der technischen Konnektivität. Basierend auf dem Identifizierungsmittel und den Angaben in der Selbstbeschreibung entscheidet die Dataspace Governance Authority, also eine technische Akteur wie in den vorherigen Kapiteln erläutert, im besten Falle automatisiert über die Zulassung zum Datenraum.

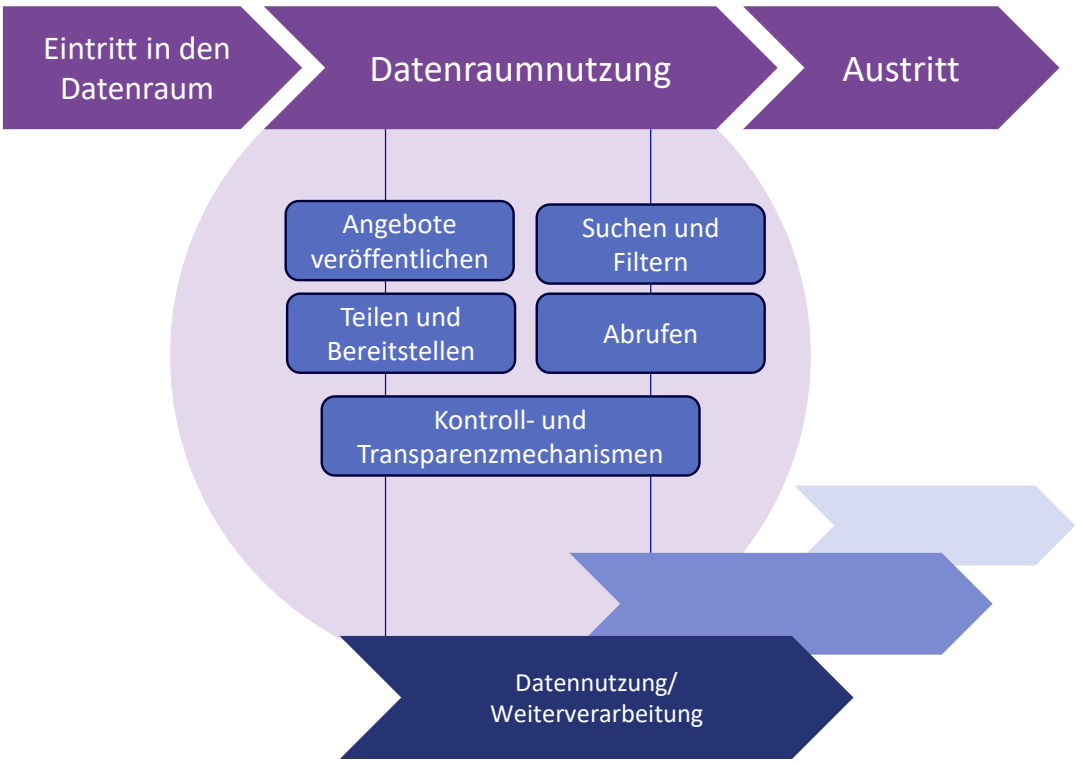



*Abbildung 9: Grundlegende Aktivitäten von Teilnehmenden in Datenräumen*

Nach der Genehmigung enthält die teilnehmende Organisation von der Dataspace Governance Authority einen verifizierbaren Berechtigungsnachweis, typischerweise in Form eines Verifiable Credentials (VC). Damit wird die Mitgliedschaft nachgewiesen. Zudem listet die jeweilige Dataspace Governance Authority die Teilnehmenden in ihrem Mitgliederverzeichnis. Es ist wichtig zu beachten, dass dieser Prozess der Aufnahme in den Datenraum bei mehreren Dataspace Governance Authorities erfolgen kann, so dass bei Bestehen der notwendigen technischen Voraussetzungen eine mehrfache, zeitgleiche Teilnahme möglich ist. Die Zulassungsprozesse werden von der jeweiligen Dataspace Governance Authority festgelegt und können beliebige weitere Schritte umfassen, wie zum Beispiel die Prüfung durch einen externen Partner, manuelle, oder ergänzende vertragliche Schritte.

Abgesehen von dem zuvor beschriebenen Antrag auf die Teilnahme, müssen Teilnehmende die jeweils erforderliche technische Infrastruktur einrichten. Diese können je nach den Anforderungen und Regeln des Datenraums und des ausgewählten technischen Rahmens variieren. Um die Sicherheit und Kontrollmechanismen zu gewährleisten, ist die technische Gestaltung möglicherweise auch bereits Teil des initialen Genehmigungsverfahrens.

**Aktivitäten im Datenraum**
Nachdem der Eintritt in den Datenraum organisatorisch und technisch erfolgt ist, können die Teilnehmenden ihre fortlaufenden Aktivitäten vornehmen: Diese umfassen das Management und die Handhabung von Datenangeboten, den zugehörigen Verträgen über den Datenzugang und der Datennutzung, sowie die Durchführung von Transaktionen und das Beschließen von Vertragsbeziehungen mit anderen Teilnehmenden. Ebenfalls gehört die Instandhaltung der technischen Infrastruktur zu der fortlaufenden Aufgabe, sowie etwaige Rechte und Pflichten der Dataspace Governance Authority. Dies können beispielsweise Vorgaben zur Pflege und Aktualisierung der Selbstbeschreibungen und notwendigen Attribute sein, aber auch Aktivitäten zur Entwicklung der Datenraum-Community. Auch werden Regelverstöße und Ausnahmesituationen von der jeweiligen Dataspace Governance Authority gehändelt.

**Austritt und Verlassen des Datenraums**
Das freiwillige Austreten aus dem Datenraum erfolgt durch einen vom Teilnehmer initiierten Prozess. Der Teilnehmende lässt den Berechtigungsnachweis aufheben und sich aus dem Mitgliedsregister entfernen. Je nach Datenraum gibt es noch weitere Vorgaben zum Verlassen des Datenraums, zum Beispiel um bestehende Datenaustausche nicht abrupt zu unterbrechen oder gegen bestehende Verträge mit anderen Teilnehmenden zu verstoßen.

Neben dem freiwilligen Austritt gibt es noch den Fall des erzwungenen Austretens. Diese Maßnahme erfolgt zumeist, wenn Teilnehmende die Regeln des Datenraums verletzen oder sich nicht an Vereinbarungen halten. Auch hier werden Berechtigungsnachweise entzogen und die betroffenen Teilnehmenden von der Mitgliedsliste entfernt. Auch ein hinzuziehen von weiteren Eskalationsstellen und -prozessen ist je nach Governance denkbar.

# Gestaltung eines dezentralen Datenraums

## Modularität und Dezentrale Service-Strukturen

Die Gestaltung von Datenräumen, insbesondere mit Blick auf die Rolle der Dataspace Governance Authority, kann in verschiedenen Strukturen erfolgen [9]: Die zentrale Gestaltung bedeutet, dass alle Dienste um den Datenraum zu betreiben von einer einzigen Stelle bereitgestellt werden. Um technischen Herausforderungen und Sicherheitsanforderungen gerecht zu werden, werden in einer föderierten Umsetzung die notwendigen Dienste verteilt angeboten. Die geteilte Verantwortung für die Dienste benötigt technische Aufwände für beispielsweise Synchronisierungen. Das höchste Maß an Autonomie und Souveränität wird mit einer dezentralen Struktur erreicht, in der Teilnehmende durch ein dezentrales Identitätssystem befähigt werden, unabhängig von Identitätsanbietern Vertrauen aufzubauen und Transaktionen durchzuführen.

Die dezentrale Gestaltung von Datenräumen ermöglicht Skalierbarkeit und Resilienz, sowie die Anpassungsfähigkeit an verschiedene Systeme. Eine zentrale Gestaltung bringt insbesondere Einfachheit im Aufbau, der Koordination und im Umgang mit den Komponenten mit sich, sowie zentrale Kontrollmechanismen und eine leichte Auffindbarkeit von anderen Teilnehmenden und ihren Datenangeboten. Diese Eigenschaften erweisen sich vor allem bei ersten Prototypen als nützlich. Im Gaia-X 4 KI – Projekt dienten zentrale Komponenten von daher insbesondere zur Realisierung von ersten technischen Durchstichen und Minimalbeispielen. Diese dienten dazu, das Konzept in einem konkreten Anwendungsfall technisch umzusetzen. Im weiteren Verlauf des Projektes und insbesondere im Kontext der übergreifenden Gaia-X 4 Future Mobility Projektfamilie etablierte sich eine dezentrale Struktur, die Autonomie in der Technologieauswahl sowie Skalierfähigkeit mit sich brachte.

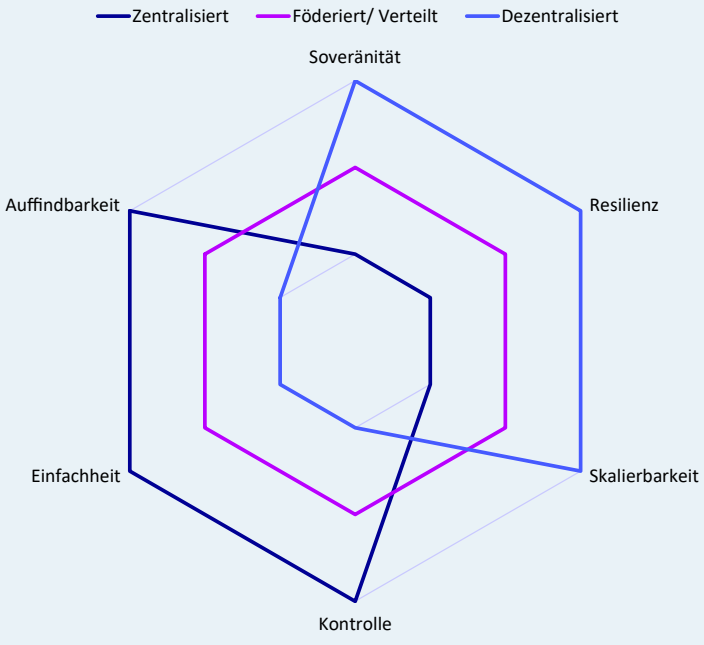



*Abbildung 10: Gegenüberstellung von verschiedenen Datenraumstrukturen*

## Datenraum-Gestaltung in Gaia-X 4 KI

In einem Datenraum wird es unterschiedlichen Teilnehmern ermöglicht, auf Assets anderer Teilnehmenden unter selbstdefinierten Bedingungen zuzugreifen. Dazu ist eine Verbindung zwischen den Teilnehmern nötig, die den besonderen Anforderungen innerhalb eines Datenraumes Rechnung trägt. In Gaia-X4KI übernehmen dies die Eclipse Dataspace Components. Diese beinhalten verschiedene Komponenten, die einen dezentralen Datenraum aufbauen und Herausforderungen von Datenräumen begegnen, während sie gleichzeitig bestehende Technologien und Standards so weit wie möglichen nutzen.

Die Datenraum-Implementierung in Gaia-X4KI entspricht dem EDC Minimum Viable Dataspace, kurz MVD genannt. Das MVD ist ein integriertes Set an konkreten Implementierungen von Dataspace-Komponenten. Der Hauptzweck besteht darin, die Fähigkeiten des EDC zu demonstrieren, Datenraumkonzepte anhand einer spezifischen Implementierung greifbar zu machen und als Ausgangspunkt für die Implementierung eines individuell angepassten Datenraums zu dienen.

Der Aufbau des ursprünglichen EDC MVD-Repository, das im Open-Souce Projekt als integriertes Set-Up zum Testen verfügbar ist, umfasst drei Datenraumteilnehmer, die unter Einhaltung beispielhafter Regeln untereinander kommunizieren und Datensätze austauschen. Um die technischen Komponenten und Schlüsselmechanismen zu beschreiben, konzentrieren wir uns im Folgenden auf die Erklärung der ausgewählten Komponenten, um eine Art high-level Blaupause zu erstellen. Dabei realisiert die MVD-Struktur einen dezentralen Ansatz, der sich in der Gesamtarchitektur sowie in Form eines dezentralen Katalogs, der Verwendung von Web DID zur Identifizierung und der Architektur der Konnektor-Komponente ausdrückt. Darüber hinaus ermöglicht das Dataspaces Communication Protokoll[1] Interoperabilität durch eine standardisierte Interaktion zwischen den Komponenten.

Die Folgenden Komponenten sind enthalten:

- Registration Service
- EDC Connector
- Federated Catalog
- Identity Hub

1 https://docs.internationaldataspaces.org/dataspace-protocol/overview/readme

Zudem sind folgende Dokumente, Komponenten und Dienste eingesetzt, um die Kommunikation in einem Datenökosystem vertrauenswürdig zu gestalten:

- Eine unterstützende DID Infrastruktur, die einen Speicher für DID Dokumente und einen Secrets Storage bereitstellt
- Einen Speicher für eine Liste aller Datenraumteilnehmer
- DID Dokumente und die dazugehörigen privaten und öffentlichen Keys
- Ein Set an verifizierbaren VCs

Diese zusätzlichen Elemente sind auf dem Markt als Cloudangebot in verschiedenen Formen von verschiedenen Anbietern und in verschiedenen Bereitstellungs- und Erwerbsmodellen erhältlich.

### Registration Service

Der Registration Service ist eine Komponente, die sich von den anderen Komponenten unterscheidet, da sie als einzige von ein oder mehreren Dataspace Governance Authorities bereitgestellt wird. Die Registration Service Komponente setzt die Funktionen um, alle Teilnehmer zu listen und ermöglicht es, Connectoren initial die Adressen (URLs) anderer Connectoren für die peer-to-peer Kommunikation mitzuteilen. Zudem wickeln sie den technischen Eintritt in den Datenraum ab und können auch Funktionen wie das Offboarding von Teilnehmern oder Entfernen nach Regelverstößen ermöglichen.

Die weiteren Komponenten können als Agenten eines Teilnehmers gesehen werden, da sie ein System sind um Aktivitäten in einem Datenraum im Auftrag eines Teilnehmers ausführen.

### Connector

Der Connector beschreibt in seiner Gesamtheit das verbindende Element zwischen der Datenquelle und dem Datenraum. Der EDC Connector ist eine Komponente, die im Wesentlichen die Verwaltung und Aushandlung der Bedingungen, sowie die Initiierung des eigentlichen Datentransfers umsetzt. Über eine REST Schnittstelle wird der Endpunkt des Konnektors mit dem jeweiligen System verbunden. Der Connector besteht dabei aus zwei Komponenten, der Control und Data Plane. Im Kontrast zu einer monolithischen Architektur setzt der EDC MVD

Connector damit eine Trennung der Verwaltung (Control Plane) und des eigentlichen Datenaustausch (Data Plane) um. Somit können sie unabhängig voneinander definiert, betrieben, sowie bedarfsweise kombiniert werden. Dabei setzt die Data Plane mithilfe von bestehenden Datenaustauschtechnologien den eigentlich Datenaustauch um, während die Control Plane die für den souveränen Datenaustausch relevanten Kontroll- und Managementmechanismen ausführt. Ein wesentlicher Aspekt ist hierbei, dass die Data Plane für den eigentlichen Datentransfer zwischen den Teilnehmern zuständig ist, ohne dabei eine zentrale Instanz zu passieren oder kontrolliert zu werden. Die Control Plane als Verwaltungsfunktion übernimmt dagegen anhand von Metadaten und Vertragsangaben die im Datenraum notwendigen Vertrauensmechanismen, wozu auch die Kommunikation mit einer (oder mehreren) Dataspace Authorities zählt. Durch Verbindungen zwischen Data und Control Plane können somit zum Beispiel auch Datenflüsse anhand Änderungen der Control Plane gestoppt werden oder ausgewählte Logs und Statusinformationen weiterverarbeitet werden.

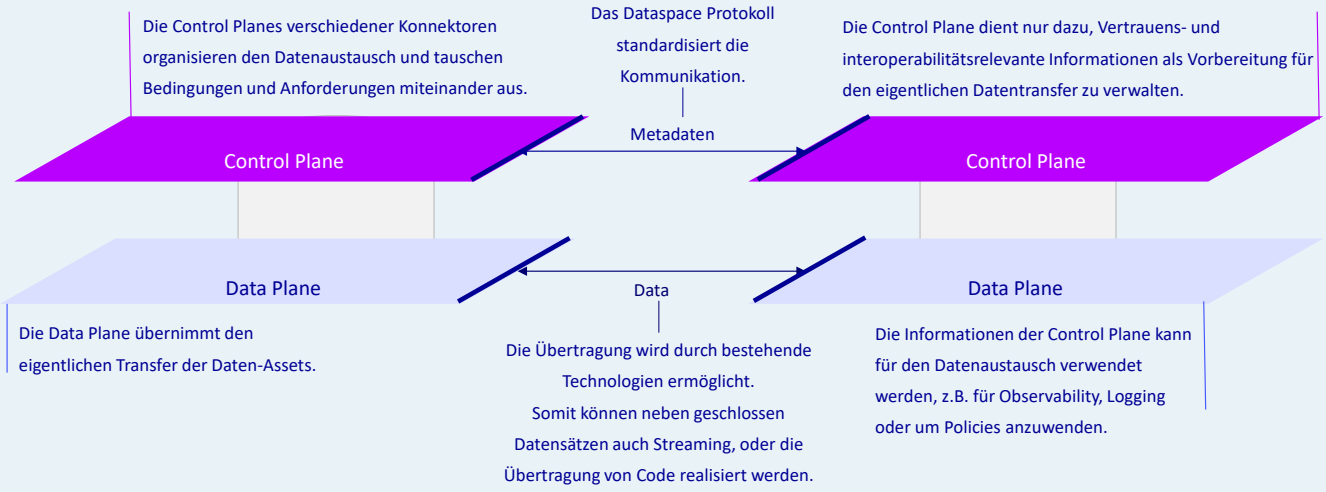



*Abbildung 11: Unterscheidung zwischen Control und Data Plane*

## Katalogisierung

Eine Katalogisierung der verfügbaren Datenangebote wird durch den Federated Catalog, bestehend aus einem Cache und einem Crawler umgesetzt. Der Registration Service unterstützt in unserem Set-Up um den Katalog initial mit Einträgen zu füllen und Connectoren miteinander bekannt zu machen.

Der Node führt die Funktionalität aus, für seinen jeweiligen Connector die gewünschten Katalogeinträge über verfügbare Assets zu publizieren. Diese Einträge können mit Policies versehen werden, die bestimmen welche anderen Teilnehmer den Eintrag sehen dürfen und unter welchen Bedingungen die Assets versehen sind. Der Crawler ist für die Entdeckung bestehender Einträge von anderen zuständig und gibt dem Connector an, welche Assets zur Verfügung stehen.

## Identity Hub

Unter Verwendung einer dezentralen Identifizierung und der Methode Did:Web, wird eine dezentrale Identifizierung der Teilnehmer ermöglicht. Eine Identifizierung ist die Grundlage, um mit anderen in einen abgesicherten Austausch treten zu können.

Die verwendeten DID Mechanismen für die Identifizierung können auch für die Autorisierung im Datenraum verwendet werden. Im Kontext von Datenräumen ist damit die Identifizierung von einzelnen Datenraumteilnehmern gemeint. Dies betrifft insbesondere, wer zum Datenraum zugelassen wird. Weitere Rechte und Autorisierungen werden anhand von Attributen umgesetzt. Die Attribute werden ebenfalls auf eine dezentrale Art und Weise geprüft und attestiert.

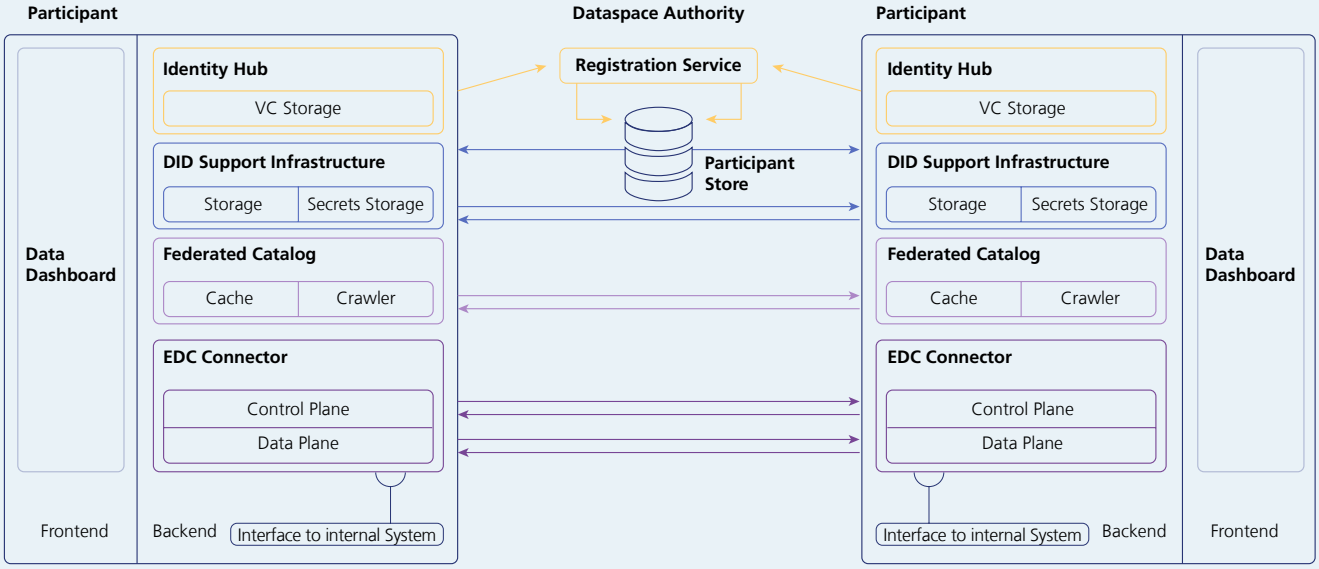



*Abbildung 12: Detailsicht auf die verwendeten Datenraum-Komponenten*

In der Gesamtheit können wird somit Teilnehmer ermöglicht mit einem Set an Komponenten, das sie selbständig kombinieren, kontrollieren und Entscheidungen über sie treffen können (z.B. über den Betriebsmodus) an verschiedenen Datenräumen teilzunehmen. Die Modularität erlaubt es, verschiedene Data Planes, z.B. für Streaming, einzubinden, sowie mit mehreren Dataspace Governance Authorities zu interagieren.

## Gaia-X Vertrauensmechanismen

Eine wesentliche Rolle im Projekt Gaia-X 4 KI spielen die Vertrauensmechanismen, die von der Gaia-X Association definiert sind. Gaia-X bietet mit dem Gaia-X Trust Framework eine gemeinsame Governance für Vertrauensmechanismen und ein grundlegendes Maß an Interoperabilität. Diese Interoperabilität bezieht sich insbesondere auf die Interoperabilität zwischen verschiedenen Ökosystemen und betonen die Autonomie der Teilnehmenden. Der Fokus liegt auf dem Trust Framework, weil dies die Operationalisierung der Anforderungen von Gaia-X ist, zum Beispiel den Policy Rules oder der Architektur [14].

Gaia-X adressiert somit die Trust Plane, also die Vertrauensebene, die verschiedenen Ökosystemen und ihren Technologien zugrunde liegt [15].

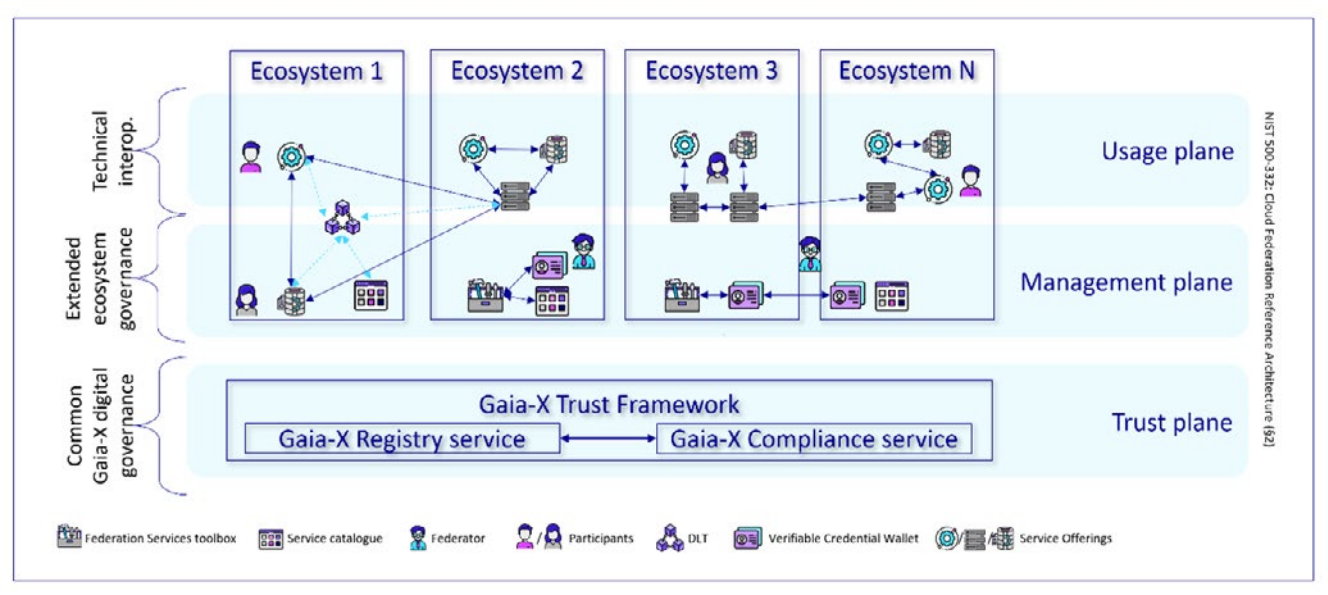



*Abbildung 13: Zusammenhang Gaia-X Vertrauensmechanismen und Individuelle Ökosysteme [15]*

Um dies umzusetzen, wird das Gaia-X Digital Clearing House als Mechanismus zur Validierung der Gaia-X Regeln verwendet. Eine Komponente des Gaia-X Digital Clearing House ist die Gaia-X Registry, ein Service, der die benötigten Vertrauensanker bereithält. Es werden die Vertrauensanker festgehalten [16]. Vertrauensanker untermauern die Angaben („Claims“) der Akteure in Datenräumen und in Zertifikatsketten muss mindestens einer der vom Gaia-X Trust Framework benannten Vertrauensanker zur Signatur verwendet werden.  Die Gaia-X Registry wird von dem Gaia-X Compliance Service adressiert. Der Gaia-X Compliance Service validiert die Inhalte und die Form von relevanten Berechtigungsnachweisen, und signiert diejenigen, die den Gaia-X Regeln entsprechen.

Das Gaia-X Trust Framework baut auf den logischen Prinzipien der W3C Verifiable Credentials auf [17]. Die Datenraumteilnehmenden agieren als Holder, die ihre Berechtigungsnachweise verwalten und kontrollieren. Die Nachweise werden von den jeweiligen Vertrauensankern ausgestellt (Issuer), die den Nachweis digital signieren und zur Nachprüfung in einer Registry ablegen. Bei der Verwendung der Nachweise präsentieren die Teilnehmenden ihre Nachweise anderen Teilnehmenden (Verifier), die wiederum verifizieren, dass die ursprüngliche Ausstellung ihren Anforderungen entspricht.

Im Gaia-X 4 KI Projekt wird die Implementierung des T-Systems Digital.ID verwendet. Somit werden vertrauenswürdige Authentizitätsnachweise erstellt, die auf Vertrauensankern basieren, die von nationalen und europäischen Regulierungen anerkannt sind [18].

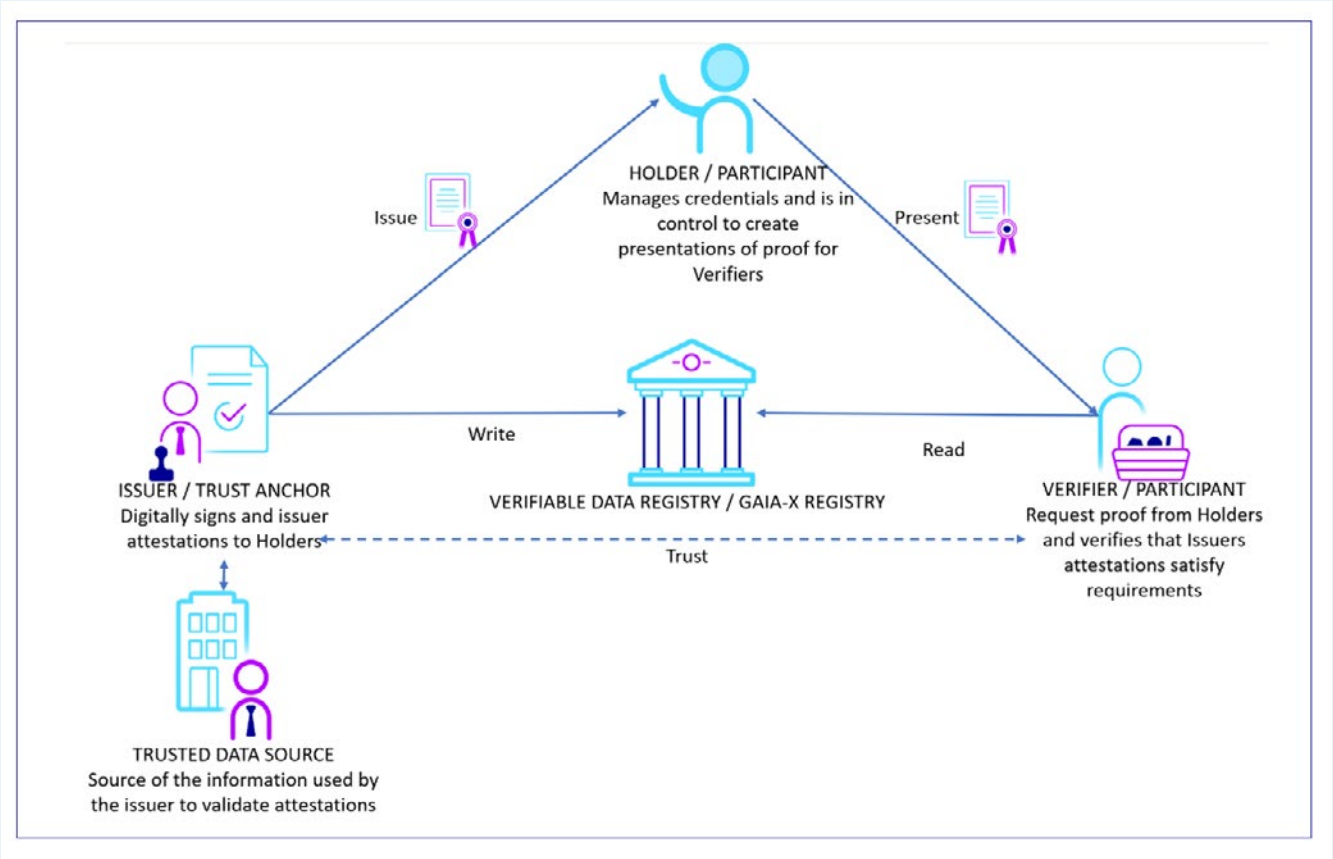



*Abbildung 14: Mechanismen des Gaia-X Trust Frameworks [14]*

# Anwendung in Gaia-X 4 KI

## Big Picture Gaia-X 4 KI

Gaia-X 4 KI das Ziel einen übergreifendes Daten- und Dienste Ökosystem zu gestalten. Dabei werden drei große Anwendungsbereiche abgedeckt und miteinander verbunden. Diese drei Teilbereiche sind die Produktion von Sensoren, die Entwicklung und der Betrieb & Service.

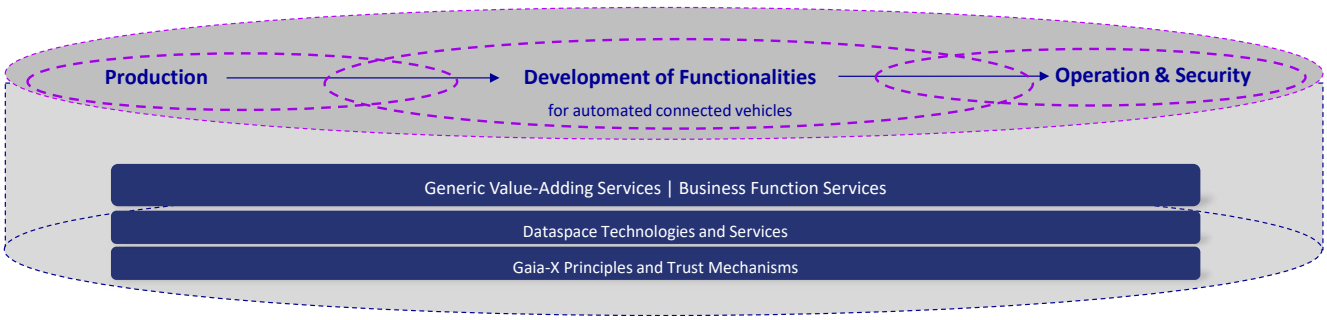



*Abbildung 15: Dienste und Datenquellen des Anwendungszenarios*

Die Ziele des Datenraum-Bereiches Produktion sind die Steigerung der Produktionsqualität und Zugriff auf einen breiten Datenpool, um datenbasierte KI-Dienste nutzen zu können. In dem Entwicklungs-Bereich wird die Erhöhung der Genauigkeit und Qualität verschiedener Algorithmen zum autonomen und vernetzten Fahren ermöglicht. In der Phase des Betrieb und Service werden (Echtzeit-)Daten zur Erhöhung der Sicherheit operativer Dienste genutzt.

Zusammenhängend betrachtet bilden die drei Teilbereiche des Big Pictures eine Einheit, welche Kooperation und inter-organisationale Datenflüsse fördert.

# Einblicke in die Anwendungsfälle

Das Projekt Gaia-X 4 KI umfasste eine Reihe von Anwendungsfällen (Use Cases). Jeder Use Case hat einen unterschiedlichen Schwerpunkt und ein eigenes Ziel, baut dabei jedoch auf denselben Datenraumtechnologien und Vertrauensmechanismen auf.

Auf der Homepage des Projektes sind kurze Demo-Videos zu den Use Cases aus dem Halbzeit-Event verfügbar:

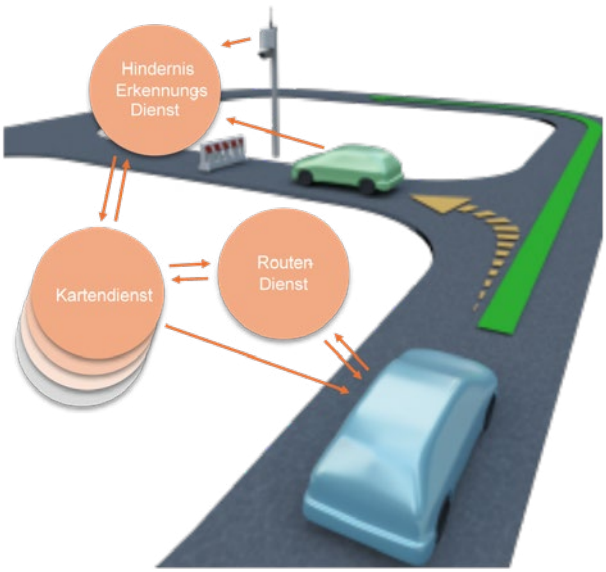
## Zu den Demo-Videos

Die Anwendungsfälle umfassen im Bereich der Fertigung die automatische optische Qualitätskontrolle (AoQ) sowie einen digitalen Zwilling für Prüfmethoden, Datengenerierung und Algorithmen. Im Bereich des automatisierten und vernetzten Fahrens werden die Erfassung und gesicherte Bereitstellung von Daten umgesetzt, sowie ein Security Incident Event Monitoring (SIEM), Funktionale Sicherheit als Service (FSS), online Kartenannotation und Parametrisierung, und die Reklamationsbearbeitung im Transportbetrieb der Automobilproduktion als „Gaia-X compatible claim handling in automotive production transport drives" (CAPTD).

### Einblick in den Anwendungsfall Online Kartenannotation und Parametrisierung von Automatisiertem und Vernetztem Fahren

In Gaia-X 4 KI wurde die Anwendung von Gaia-X im Kontext von automatisiertem und vernetztem Fahrzeugfunktionen erprobt. Die zunehmende Vernetzung von Fahrzeugen vor allen im Hinblick auf zukünftige Mobilität war hierbei der Treiber für das Vorhaben.

Ausgangspunkt ist ein einfaches Szenario bei dem Fahrzeug A in einer Fahraufgabe an ein Hindernis gerät, dass als Blockade auf der Route identifiziert wird. Diese Information wird über verknüpfte Datendienste im Datenraum weiterverteilt sodass ein Fahrzeug B eine Alternativroute geliefert bekommt um die blockierte Straße zu umgehen.



Hauptaugenmerk wurde auf die praktische Durchführung gelegt und darauf ob die Verknüpfung der Dienste durch einen Connector den zeitkritischen Anforderungen vom automatisierten Fahren genügen. Eine Herausforderung die sich schnell zeigte war der Umgang mit dynamischen IPs auf Fahrzeugseite. Die hier angebundenen Dienste brauchten eine feste Adresse für die Antwort auf Anfragen. Ergebnis der Überlegungen dazu war eine Middleware welche mit der Fahrzeugflotte kommuniziert und die Anfragen an die Dienste über die Connector-Technologie kapselt. Die Middleware bietet zur

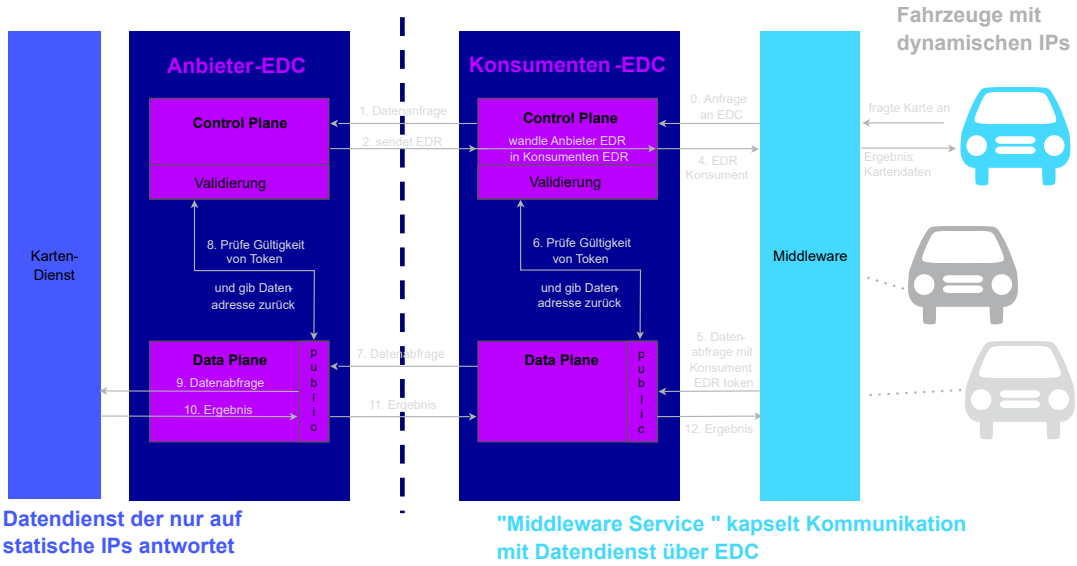



*Abbildung 16: Technische Datenraum-Struktur des Anwendungszenarios*

Fahrzeugseite eine REST-Schnittstelle, in diesem Fall konkret für die Abfrage von Kartendaten, Anfrage einer Reiseroute und die Übertragung von Sensordaten für die Hinderniserkennung. Die Middleware nutzt den Connector zur Aushandlung eines Vertrages mit dem passenden Datendienst. Wird der Vertrag angenommen sendet der Connector vom Datendienst eine EndpointDataReference (EDR) zurück. Damit kann die Middleware die von Fahrzeugseite angefragten Daten unter Angabe einer eigenen EDR anfragen. Die Antwort vom Datendienst, die gewünschten Daten, werden darüber auf das anfragende Fahrzeug zurückgeführt. Diese Architektur eignet sich vor allem für Flottenbetreiber wo eine Form von Flottenmanagementsoftware ohnehin benötigt wird. An den Datendiensten musste für die Anbindung mit Gaia-X nichts geändert werden. Eine Antwort benötigte etwa 8-10 Sekunden. Das zeitliche Verhalten von dieser Implementation war damit ausreichend für das Szenario. Das Aushandeln eines Vertrages dauerte dabei am längsten, ca. 90% der Zeit, was auf eine sehr frühe Version der verwendeten Connector-Technologie (EDC) zurück zu führen ist.

Der Vorteil von Gaia-X erschließt sich vor allem dann, wenn das reduzierte Szenario auf die Größe einer Anwendung im Realbetrieb skaliert wird. Im Projekt wurde sich auf ein Kartendienst, ein Routendienst, usw. beschränkt. Jedoch würde es in einem Datenraum realistisch mehrere Anbieter mit unterschiedlicher Ausprägung geben. Regionale oder preisliche Unterschiede könnten automatisiert in die Auswahl des Dienstes einbezogen werden. Somit könnte immer der für die Situation am besten passende Dienst ausgewählt werden. Das automatisierte Aushandeln von Verträgen führt dazu, dass eine Vertrauensbasis für die Daten existiert und somit die Daten technisch und juristisch sicher verwendet werden können.

## Anwendungsfall Datenanreicherung, semantische Metadaten und Datenverfolgbarkeit auf Basis von Graphtechnologie

„Knowledge Graphs“ werden von Gartner [16] als »Critical Enablers« u.a. für Generative AI, Digital Twins wie auch das Industrial Metaverse eingestuft [16].

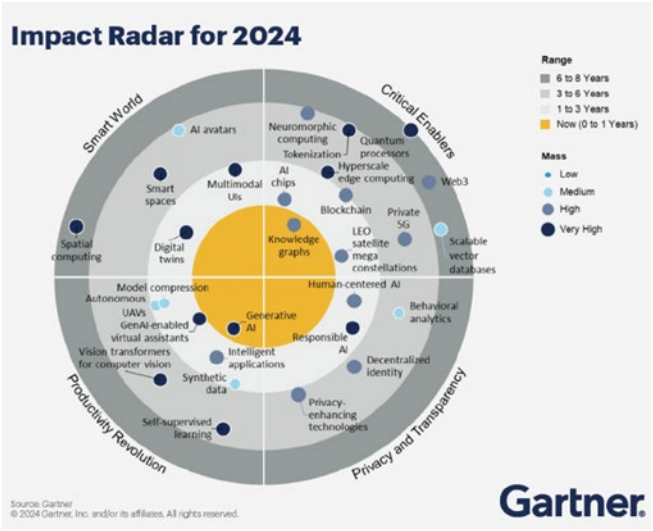



*Abbildung 17: Graphtechnologien im Gartner Emerging Technologies Radar [16]*

Die Technologie ermöglicht z.B. natürlichsprachliche Fragestellungen, die via KI aus dem Graphen heraus unter Berücksichtigung des Kontexts qualitativ hochwertiger beantwortet werden können und greift das zentrale Thema des GAIA-X 4 KI-Projekts auf.

Das Fallbeispiel hat durch die Adressierung elektrifizierter, bedarfsgesteuerter Mobilität auch Bezug zum Thema der gesamten GAIA-X 4 Future Mobility Projektfamilie. Im Fallbeispiel werden vier Aspekte gezeigt:

1. Die Speicherung von Daten in einem Graphen als Basistechnologie für alle Aspekte des Fallbeispiels. Die verwendete Plattform ist nicht im Rahmen des Projekts entwickelt worden, sondern besteht aus einer Reihe von Standardtechnologien.

2. Wie Datenangebote mit semantischen Informationen versehen und verbunden werden können, um automatisierte Interpretation und Verarbeitung zu unterstützen.

3. Wie die Herkunft von Daten nachvollzogen werden kann.

4. Wie Datenquellen aus verschiedenen Datenräumen übergreifend gefunden und verarbeitet werden können, um daraus neue Datenangebote mit Mehrwert bereitzustellen.

Das angenommene Szenario beschreibt die Idee, dass Betriebe oder Organisationen, die über Flächen im öffentlichen Bereich verfügen, eine Vorhersage über den möglichen PV-Ertrag solcher Flächen als Datenservice einkaufen können, um zu prüfen, ob durch die Nutzung ein wirtschaftlicher Beitrag zur Ladeinfrastruktur des Mobilitätsnetzwerks entstehen kann. Durch die Kombination von Daten, die aus verschiedenen Ökosystemen stammen, werden Daten angereichert und ein Mehrwert erzeugt, der in Form eines neuen Datenangebots nutzbar gemacht werden kann.

Im konkreten Fall wird die Lage und Größe von Haltestellen im ÖPNV verwendet und durch einen Qualifizierungsdienst ergänzt, der über eine mobile Workforce (sog. Crowd-Sourcing) die Eigenschaften von Haltestellen vor Ort überprüft und erfasst.

Die Daten zum Grad der Abschattung durch angrenzende Bebauung oder Vegetation, verfügbare Infrastruktur (Elektrifizierung, Dächer, Masten, Verteilerkästen) etc. werden durch einen externen Dienstleister erhoben und als Datenangebot in einem Gaia-X konformen Datenraum angeboten. Solche ergänzenden Dienste sind typischerweise durch vertragliche Vereinbarungen, die im Rahmen der Gaia-X Infrastruktur automatisch verhandelt werden können, kommerziell nutzbar. Die Haltestellendaten sind oft ohne besondere Einschränkungen öffentlich und kostenlos erhältlich. Im Beispiel wurden Daten genutzt, die für das Haltestellenmanagement im Rahmen des Gaia-X 4 ROMS-Teilprojekts entstanden sind. Im Fallbeispiel werden Namen und Geolokation einer Haltestelle, sowie die Abmessungen (Länge und Breite) als Ausgangsbasis benötigt.

Aus diesen Daten wird eine Grobkalkulation des möglichen Ertrags je Haltestelle durchgeführt, die auf der Annahme basiert, dass die Fläche der Haltestelle mit PV-Elementen überdacht werden kann. Anhand dieser PV-Kategorie können Haltestellen selektiert werden, die zur Qualifizierung geeignet erscheinen. Name und Geolokation, sowie die Arbeitsanweisung werden an die Crowd-Sourcing Community übergeben.

Über das Abonnement eines kommerzielles Datenangebots des Vermarkters dieser Daten, werden diese Informationen verarbeitet und mithilfe geeigneter Methoden und eigenem Know-How zu einem realistischen Ertragspotenzial je Haltestelle verrechnet.

Dieses Ergebnis wird als neues Angebot in einem geeigneten Datenraum registriert, mit Policies zur Nutzung versehen und über eine GraphQL API-Schnittstelle zum Abruf bereitgestellt.

Der Betreiber von Haltestellen kann nun für seine als Haltestellen genutzten Flächen ein Konzept zur Photovoltaik (PV) Nutzung erarbeiten, indem er realistische Ertragsprognosen im Datenkatalog auffinden und einkaufen kann, um sie zu verwenden.

Außerdem wurde im Fallbeispiel gezeigt, wie Datenangebote mit semantischen Informationen versehen werden können, die die automatische Interpretation und Verarbeitung bestimmter Datenangebote unterstützen können. Das Datenangebot zum Abruf der Haltestellendaten wird im Rahmen eines Serviceangebots, das aus mehreren Datenangeboten besteht, bereitgestellt. Über den Knowledge-Graphen kann die Verknüpfung des Datenendpunkts mit einem Metadaten-Endpunkt sichtbar gemacht, und die beschreibende Datei (in diesem Falle im Turtle/ttf Format) zur Bewertung des Angebots bezüglich des Umfangs und der Eignung einbezogen werden.

Durch die Einbindung von formalen Datenbeschreibungen wie z.B. Ontologien, können Datenangebote in einen Kontext gesetzt werden, der die Auffindbarkeit geeigneter Daten mit semantischen Suchen unterstützt.

Verschiedene Ansätze zur Realisierung von semantischen Methoden in Gaia-X wurden im Rahmen von BASE-X, insbesondere in der Semantic Core Workgroup diskutiert und erarbeitet.

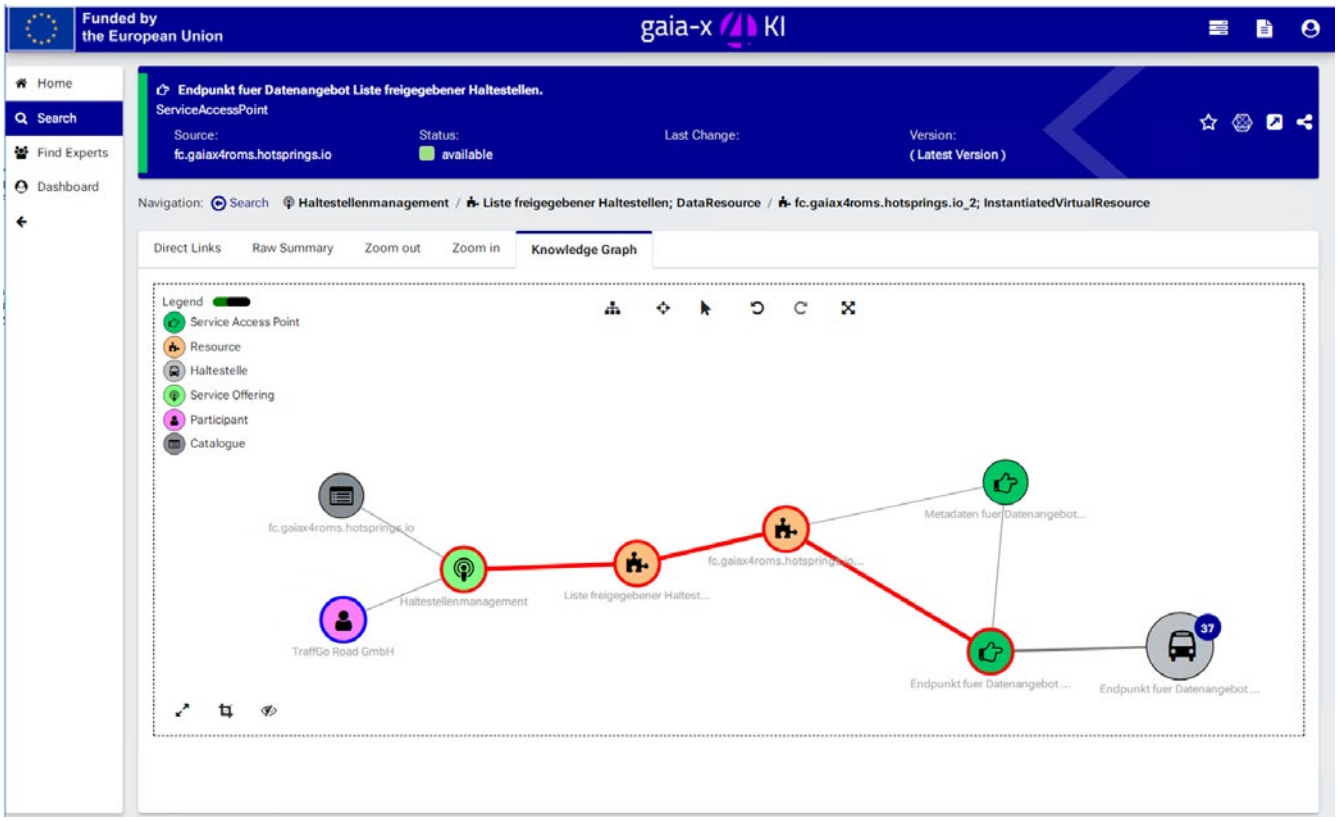



*Abbildung 18: Wissensgraph mit Pfad vom Serviceangebot zu den verarbeiteten Daten*

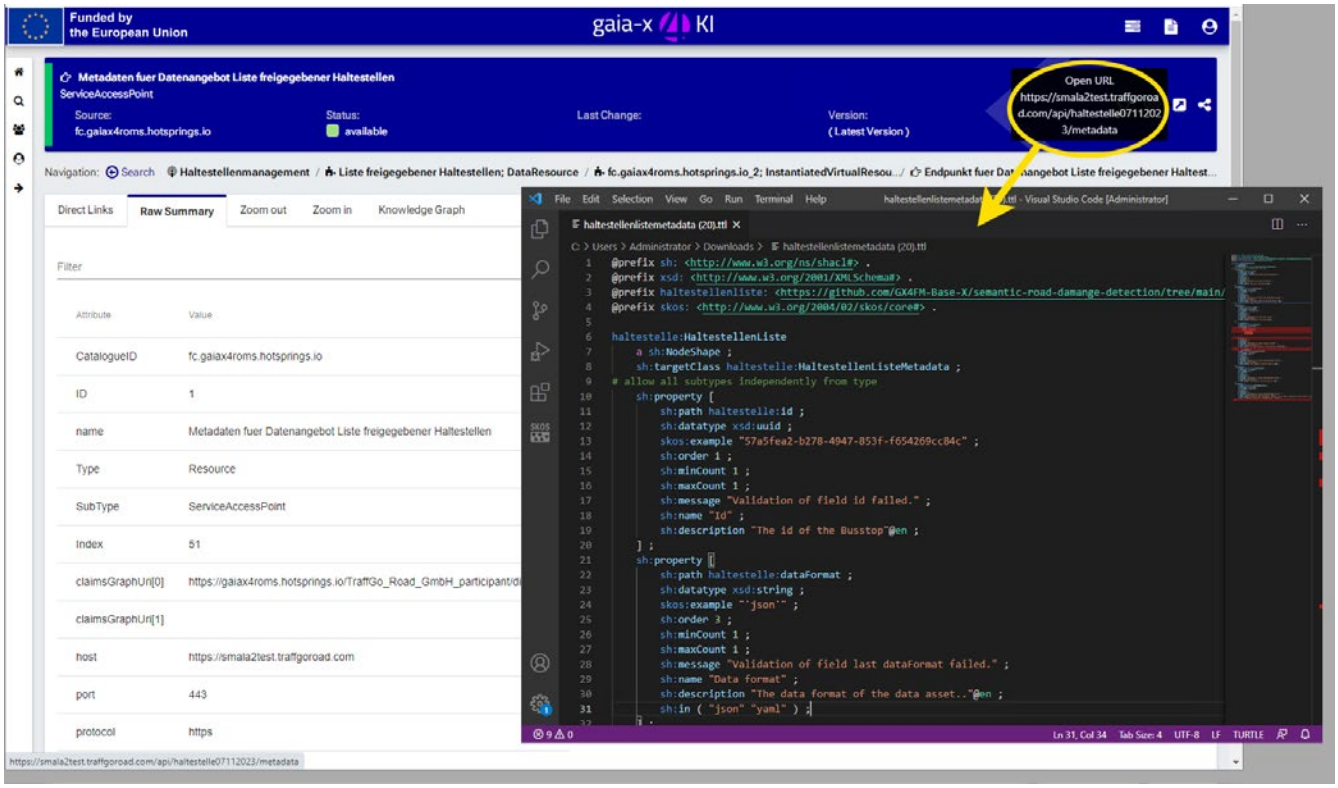



*Abbildung 19: Abruf der Metadaten zu einem Datenangebot*

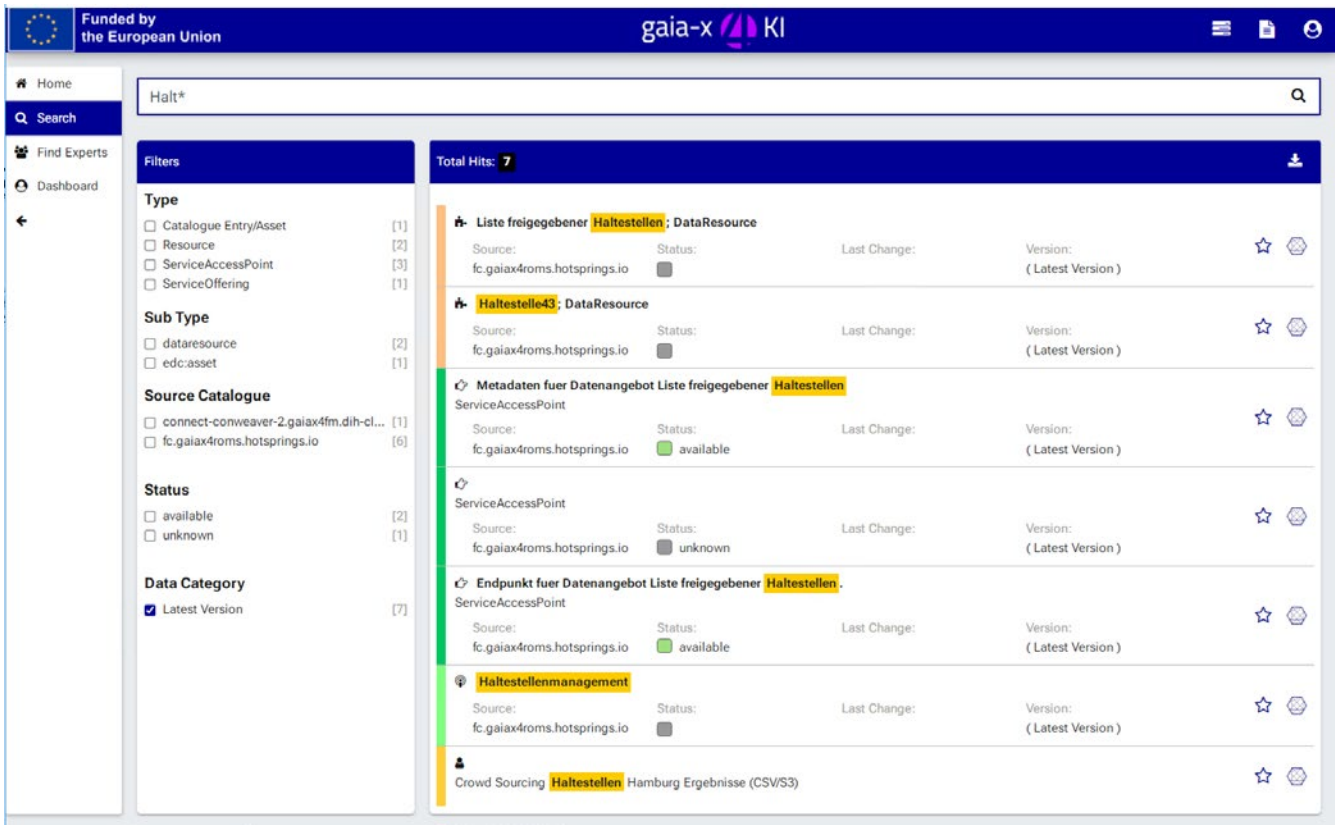



*Abbildung 20: Volltextsuche über mehrere Kataloge hinweg*

# Referenzen


1. Bußmann-Welsch TM, Tholey F Der Handel mit personenbezogenen Daten auf dezentralen Datenmarktplätzen Zeitschrift zum Innovations- und Technikrecht (InTer), Heft 4, S 225–233

2. Caldarola MC, Schrey J (2018) Big Data und Recht. Einführung für die Praxis. C.H. Beck

3. Europäisches Parlament und Rat (2017) Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates über die Achtung des Privatlebens und den Schutz personenbezogener Daten in der elektronischen Kommunikation und zur Aufhebung der Richtlinie 2002/58/EG. Verordnung über Privatsphäre und elektronische Kommunikation

4. Staudenmayer D (2011) Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates über ein gemeinsames europäisches Kaufrecht. München: Beck:15f.

5. Data Spaces Support Centre (2023) Data Spaces Blueprint. Verfügbar unter https://dssc.eu/

6. International Data Spaces Association e.V. (2023) IDSA Rulebook V2. Functional requirements for a data space. Verfügbar online in der IDSA Knowledge Base.

7. European Commission (2017) The New European Interoperability Framework. ISA² - Interoperability solutions for public administrations, businesses and citizens.

8. International Organization for Standardization Information technology. Cloud computing (ISO/IEC 19941:2017).

9. W3C Data Catalog Vocabulary (DCAT). Version 3.

10. W3C ODRL Information Model 2.2.

11. Gaia-X AISBL (2023) Gaia-X Trust Framework. Gaia-X Trust Framework - 22.10 Release.

12. Gaia-X AISBL (2023) Architecture Document 22.10. Ecosystems.

13. Gaia-X AISBL (2023) List of defined Trust Anchors.

14. W3C (2022) Verifiable Credentials. Data Model - Ecosystem Overview.

15. T-Systems (2024) Digital.ID.

16. Gartner (2024) 30 Emerging Technologies That Will Guide Your Business Decisions.

# Impressum

Finanziert von der Europäischen Union
NextGenerationEU

Gefördert durch:
Bundesministerium für Wirtschaft und Klimaschutz
aufgrund eines Beschlusses des Deutschen Bundestages

Diese Arbeit wurde vom Bundesministerium für Wirtschaft und Klimaschutz im Rahmen des Gaia-X 4 KI-Projekts gefördert (19A21011E).

## Kontakt

www.gaia-x4ki.eu

Projektkoordinator: Dr.-Ing. Sascha Knake-Langhorst
E-Mail: info@gaia-x4ki.de

1. Auflage, April 2024

**Herausgeber**

Fraunhofer-Institut für Software- und Systemtechnik ISST
Speicherstraße 6
44147 Dortmund

**Autoren**

Bendiek, Katrin
Koretskaia, Daria
Lobig, Thomas
Schleimer, Anna Maria
Theissen, Natalia
Wang, Dandan
Dörr, Sebastian
Pospischil, Frank

**Satz und Layout**

Elisa Kadelka