



Fraunhofer
ISST

Fraunhofer-Institut für Software- und
Systemtechnik ISST



Data Mesh und Data Spaces

Gegenüberstellung im Kontext Datenökosystem Schweiz

STUDIE IM AUFTRAG VON



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Bundeskanzlei BK
Chancellerie fédérale ChF
Cancelleria federale CaF
Federal Chancellery FCh

Data Mesh und Data Spaces

Gegenüberstellung im Kontext Datenökosystem Schweiz

AUTOREN

Nils Jahnke		Fraunhofer ISST
Katrin Bendiek		Fraunhofer ISST
Malina Klueß		Fraunhofer ISST

Fraunhofer-Institut für Software- und Systemtechnik ISST

Datum: April 2025

EINE STUDIE IM AUFTRAG VON



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Bundeskanzlei BK
Chancellerie fédérale ChF
Cancelleria federale CaF
Federal Chancellery FCh

BILDQUELLEN

Cover: ©CtrlN - AdobeStock - 1295989766

Management Summary

In den letzten Jahren haben sich die Konzepte Data Mesh und Datenräume als vielversprechende Ansätze zur Förderung der Mehrfachnutzung von Daten etabliert. Während Data Mesh die bereichsübergreifende Nutzung von analytischen Daten innerhalb komplexer Organisationen in den Fokus rückt, zielen Datenräume auf die organisationsübergreifende Zusammenarbeit und das Teilen zwischen Organisationen von Daten ab. Beide Konzepte basieren auf soziotechnischen Prinzipien und betonen die Autonomie von Datenbereitstellenden und Datennutzenden. Trotz der Überschneidungen in Architektur, Mechanismen und Technologien fehlt es jedoch an einer Harmonisierung und Integration dieser Ansätze.

Diese Studie untersucht, wie Elemente aus dem Bereich Data Mesh auf Datenräume übertragen werden können und validiert die Erkenntnisse anhand eines praktischen Anwendungsbeispiels im Kontext des Schweizer Gesundheitsdatenraums. Die Ergebnisse bieten verschiedenen Anspruchsgruppen wertvolle Einblicke: Datenraumträgerschaften erhalten eine Übersicht über anwendbare Data Mesh-Ansätze zur Steigerung der Skalierbarkeit und Fairness, während Entscheiderinnen und Entscheider in beteiligten Organisationen ein vertieftes Verständnis für Gestaltungsmöglichkeiten von Datenräumen und die Eigenschaften von Data Mesh gewinnen.

Die Studie identifiziert eine Anwendung der Data Mesh-Konzepte »Federated Computational Governance«, »Data as a Product« und »Self-Serve Data Platform« als Befähiger von Mehrwerten in der Datenbereitstellung und Datennutzung in Datenräumen:

- Die Einführung einer föderierten Gouvernanz impliziert eine Entscheidungsstruktur, die die Selbstsouveränität der einzelnen Akteure fördert und einen Ausgleich der individuellen Interessen bei der Festlegung von Regeln, Richtlinien, gemeinsamen Infrastrukturdiensten und Standards gewährleistet. Eine föderierte Gouvernanz kann sowohl innerhalb eines Datenraums als auch zwischen parallel entstehenden Datenräumen umgesetzt werden, um Interoperabilität und Anschlussfähigkeit zu sichern.
- Durch die Umsetzung des Datenproduktmanagements in Datenräumen entsteht eine strukturierte und nutzendenzentrierte Datenbereitstellung, die die Qualität, Auffindbarkeit und Wiederverwendbarkeit der Datenprodukte fördert. Durch die Fokussierung auf die Bedürfnisse der Datennutzenden können Daten effizienter verwaltet und in verschiedenen Anwendungen genutzt werden. Für Datenbereitstellende ergeben sich zudem neuartige Feedbackmechanismen hinsichtlich der Datennutzung und des Datenwertes, die bei einer Weiterentwicklung des Datenmanagements unterstützen.
- Einen zentralen Beitrag dazu leistet die technische Dateninfrastruktur, da sie Interoperabilität, Sicherheit, Kontrolle und Vertrauen bei der Mehrfachnutzung von Daten fördert. Die Bereitstellung der Dateninfrastruktur für Datenbereitstellende und Datennutzende durch dedizierte Betriebsorganisationen ermöglicht eine Standardisierung der Datenprodukte, reduziert die Komplexität der Teilnahme und fördert die Konzentration der Teilnehmenden auf ihre Kernaufgaben. Ein oder mehrere Betriebsorganisationen können dazu von der

Datenraumträgerschaft beauftragt werden. Sie sollten kein Interesse an der Umsetzung eigener datenbasierter Anwendungsfälle im Datenraum besitzen.

Die Applikation der Konzepte auf eine konzeptionelle Anwendung des Datenraums Gesundheit zeigt, dass diese Konzepte die Autonomie der Teilnehmenden fördern, eine schnellere Datenbereitstellung und ein nutzenorientiertes Datenmanagement ermöglichen und insgesamt die Zusammenarbeit und den Datenaustausch verbessern können. Demgegenüber stehen jedoch auch Herausforderungen wie die Komplexität der Implementierung, mögliche kulturelle Widerstände und eine geringere Dateninteroperabilität, die schlussendlich eine sorgfältige Planung und die Berücksichtigung der spezifischen Anforderungen der beteiligten Akteure erfordern.

Diese Studie wurde im Auftrag der Bundeskanzlei erstellt.

Inhalt

Management Summary	3
Inhalt	5
1. Einleitung	6
2. Data Mesh-Grundlagen	7
2.1. Data Mesh als Konzept verstehen	7
2.2. Data Mesh als Antwort auf die Limitierungen zentralisierter Datenmanagement-Ansätze	8
2.3. Die Kernprinzipien von Data Mesh	9
2.4. Data Mesh als soziotechnisches System	11
2.5. Data Mesh im Vergleich mit weiteren Datenmanagement-Konzepten	12
2.6. Data Mesh-Topologien	18
2.7. Data Mesh-Architekturmuster	21
3. Datenräume in der Schweiz – Befähiger zur Mehrfachnutzung von Daten	25
3.1. Das Datenökosystem Schweiz – Herausforderungen und Ziele	25
3.2. Datenräume als Befähiger der Datenmehrfachnutzung	26
4. Von Data Mesh lernen, um Datenräume zu fördern	29
4.1. Behandlung von Daten »as a Product« zur Ermöglichung von Generativität und Mehrfachnutzung	29
4.2. Institutionalisierte Bereitstellung der dezentralisierten technischen Dateninfrastruktur	32
4.3. Föderierte Gouvernanz in Datenräumen	33
4.4. Herausforderung: Erfüllung der Anforderungen unterschiedlicher Anspruchsgruppen in der Umsetzung	35
4.5. Praxisbeispiel: Data Mesh-Ansätze im Schweizer Datenraum Gesundheit	37
5. Diskussion und Fazit	40
6. Anhang	43
6.1. Glossar	43
7. Literaturverzeichnis	44

1. Einleitung

Mit den Konzepten Data Mesh und Data Spaces (dt. Datenräume) sind in den letzten Jahren vielversprechende Ansätze zur Ermöglichung der Mehrfachnutzung von Daten entstanden. Data Mesh fokussiert vor allem die bereichsübergreifende Nutzung von Daten innerhalb von komplexen Organisationen, während Datenräume auf eine organisationsübergreifende Nutzung von Daten – das sogenannte Data Sharing – abzielen. Beide Ansätze basieren auf einer sozio-technischen Betrachtungsweise des Datenmanagements und stellen die Autonomie von Datenbereitstellenden und Datennutzenden in den Vordergrund. Basierend auf ähnlichen Prinzipien und Paradigmen entstehen unter dem Schirm beider Konzepte Architekturen, Mechanismen, Technologien und Standards, die in ihrer Zielsetzung, ihrem Aufbau und ihrer Umsetzung starke Überschneidungen aufweisen. Allerdings fehlt es derzeit an einer Zusammenführung und Harmonisierung der Arbeiten.

Diese Studie führt beide Welten zusammen, indem sie untersucht, wie sich Ansätze aus dem Bereich Data Mesh auf Datenräume übertragen lassen. Sie greift dabei die bisherigen Arbeiten im Bereich Datenökosystem Schweiz auf und validiert die Erkenntnisse anhand eines praktischen Anwendungsbeispiels. Die Inhalte dieses Dokuments stellen Mehrwerte für verschiedene Anspruchsgruppen bereit:

- *Datenraumträgerschaften* erhalten eine Übersicht zu Ansätzen aus dem Bereich Data Mesh, die sich auf Datenräume übertragen lassen, um diese skalierbarer, generativer und fairer zu gestalten.
- Die *Anlaufstelle Datenökosystem Schweiz* erhält ein verbessertes Verständnis über die Bestandteile und Zusammenhänge von Datenräumen und Data Mesh und kann auf dieser Basis zielgerichtete Unterstützungsleistungen entwickeln.
- *Entscheiderinnen und Entscheider aus an Datenräumen beteiligten Organisationen und Unternehmen* erhalten ein erweitertes Verständnis über konkrete Gestaltungsmöglichkeiten von Datenräumen durch die Applikation bekannter Konzepte aus dem Bereich des organisationsinternen Datenmanagements.

Der Aufbau des Dokuments gestaltet sich wie folgt: Abschnitt 2 erläutert die Grundlagen von Data Mesh als soziotechnischer Ansatz zum Datenmanagement. Ausgehend von einer Darstellung der Entstehung von Data Mesh werden die fundamentalen Prinzipien von Data Mesh erläutert, Data Mesh im Vergleich zu weiteren Datenmanagement-Architekturen eingeordnet und Architekturmuster von Data Mesh beschrieben. Abschnitt 3 erläutert das Konzept Datenräume aus der Perspektive des Datenökosystems Schweiz. Basierend auf einer Zusammenfassung der Beweggründe für Datenräume werden die verschiedenen Sichtweisen auf das Konzept »Datenraum« erläutert und die Bestandteile von Datenräumen zusammengefasst. In Abschnitt 4 erfolgt die konzeptuelle Zusammenführung von Data Mesh und Datenräumen. Konkret wird hier die Anwendung von Strukturen aus den Bereichen Datenproduktmanagement, föderierte Gouvernanz und Self-Service Dateninfrastruktur im Kontext von Datenräumen fokussiert. Anhand des Beispiels Gesundheitsdatenraum Schweiz erfolgen abschließend die praxisnahe Veranschaulichung der Erkenntnisse und die Evaluation der Konzepte durch die Analyse der Stärken, Schwächen, Chancen und Risiken des Ansatzes (Abschnitt 5).

Die Erstellung dieser Studie erfolgte im Auftrag der Bundeskanzlei.

2. Data Mesh-Grundlagen

Der Begriff **Data Mesh** wurde erstmals im Jahr 2019 durch den Blogbeitrag von Zhamak Dehghani »How to Move Beyond a Monolithic Data Lake to a Distributed Data Mesh« populär¹. Dehghanis Vision des Data Mesh basiert auf der Idee, die allgemeinen Hindernisse zentralisierter Datenarchitekturen durch die Adaption von Konzepten verteilter Softwarearchitekturen zu überwinden. Zudem rücken mit Data Mesh soziale und kulturelle Aspekte des Datenmanagements stärker in den Vordergrund. Einige der unter der Data Mesh-Vision zusammengefassten Ansätze werden zu diesem Zeitpunkt bereits von Organisationen erprobt, ohne diese konkret als Data Mesh zu bezeichnen.

Inzwischen haben sich viele Organisationen sowohl auf Anwender- als auch auf Technologieanbieterseite von dieser Vision inspirieren lassen und Anstrengungen unternommen, das Konzept Data Mesh in die Praxis umzusetzen. Entstanden ist eine Vielzahl von Interpretationen, konkreten Umsetzungen und Technologien, die sich kontinuierlich weiterentwickeln. Dementsprechend handelt es sich bei Data Mesh weiterhin um ein aufkommendes Thema, dem es an Klarheit, Konsistenz und Konsolidierung mangelt. Dieser Abschnitt führt in das Thema Data Mesh ein, indem es die relevanten Kernkonzepte erläutert und den aktuellen Stand der Technik analysiert.

2.1. Data Mesh als Konzept verstehen

Für das Verständnis von Data Mesh ist es wichtig, zunächst die grundlegenden Merkmale des Ansatzes einordnen zu können. Als Startpunkt dient dazu die Data Mesh-Definition von Dehghani.

[EN] Data mesh is a decentralized sociotechnical approach to share, access, and manage a nalytical data in complex and large-scale environments—within or across organizations.

*[DE] Data Mesh ist ein **dezentraler soziotechnischer Ansatz** für die **gemeinsame Nutzung, den Zugriff und die Verwaltung von analytischen Daten in umfassenden und komplexen Umgebungen - organisationsintern oder organisationsübergreifend.** (Dehghani, 2022, S. 3)*

Ein Kernaspekt von Data Mesh ist die Dezentralisierung des Datenmanagements. Data Mesh zielt darauf ab, analytische Daten nicht mehr durch zentralisierte Datenteams verwalten zu lassen, sondern die taktische und operative Verantwortung für Daten an diejenigen Bereiche zu geben, in denen die Daten erzeugt und genutzt werden. Damit geht die Verantwortung für Daten an diejenigen Organisationsbereiche über, die üblicherweise ebenso die Hoheit über die operativen Systeme und das fachliche Wissen über die erzeugten Daten besitzen. Für die einzelnen Bereiche ergibt sich damit einerseits eine höhere Autonomie hinsichtlich der Gestaltung ihrer Datenmanagement-Prozesse, andererseits jedoch auch eine höhere Verantwortung hinsichtlich der Verfügbar- und Nutzbarmachung von Daten.

¹ <https://martinfowler.com/articles/data-monolith-to-mesh.html>

Mit dem dedizierten Verständnis als *soziotechnischer Ansatz* versteht Data Mesh das Datenmanagement als Zusammenspiel der genutzten Technologien, der Individuen mit ihren individuellen Fähigkeiten und Bedürfnissen sowie des organisatorischen Rahmens. Damit steht Data Mesh im Kontrast zu den ursprünglich eher durch neuartige Technologien getriebenen Innovationen im Bereich des Datenmanagements². Die Umsetzung von Data Mesh-Ansätzen stellt den organisatorischen und individuellen Wandel, beispielsweise durch den Aufbau von Datenkompetenz, in den Vordergrund. Eine detaillierte Betrachtung von Data Mesh als soziotechnisches System findet sich in Abschnitt 2.4.

Data Mesh fokussiert auf die Verwaltung und Nutzung von analytischen Daten. Analytische Daten bezeichnen Daten, die genutzt werden, um Muster, Trends und Zusammenhänge zu identifizieren und darauf basierend Entscheidungen zu treffen. Herkömmliche analytische Daten sind aggregierte, gefilterte und gereinigte Transaktionsdaten. Dies können beispielsweise Lieferungen oder Bestellungen aus ERP-Systemen sein, die über einen gewissen Zeitraum zusammengefasst werden. In Organisationen entstehen zudem vermehrt neuartige Analysedaten, die oftmals als Big Data bezeichnet werden. Diese Daten entstehen in großen Mengen, in verschiedener Ausprägung (strukturiert oder unstrukturiert) und mit enormer Frequenz und Geschwindigkeit. Ein Beispiel hierfür sind Daten, die aus der Nutzung von intelligenten Geräten stammen. Für die Speicherung, Verwaltung und Analyse dieser Daten ergeben sich zusätzliche technische und organisatorische Anforderungen.

Mit Data Mesh wird insbesondere die Verwaltung von Daten in *umfassenden und komplexen Umgebungen* adressiert. Konkret sind hiermit Organisationsstrukturen gemeint, in denen die Sammlung, Verwaltung und Analyse von Daten aufgrund von technischer oder organisatorischer Komplexität behindert wird. Komplexitäten ergeben sich beispielsweise durch eine Vielzahl von isolierten Datensystemen, konkurrierende Bedürfnisse im Umgang mit Daten durch verschiedene Organisationseinheiten oder inkonsistente Datenprozesse (Sutherland & Cook, 2017). Diese Komplexitäten sind üblicherweise in Organisationen mit einer hohen Anzahl an Beschäftigten sowie Unternehmen mit einer Vielzahl verschiedener Sparten und Produkte zu finden.

Auch im Bereich des organisationsübergreifenden Teilens von Daten treten ähnliche Komplexitätsfaktoren auf. Per Definition zielt Data Mesh daher nicht nur auf das *organisationsinterne* Teilen von Daten ab, sondern schließt auch den *organisationsübergreifenden* Austausch von Daten ein. Dies steht vor allem unter der Prämisse, dass die Anwendung von Data Mesh-Ansätzen innerhalb der Organisation den Datenmanagement-Reifegrad erhöht und somit auch das Teilen von Daten zwischen Organisationen vereinfacht.

2.2. Data Mesh als Antwort auf die Limitierungen zentralisierter Datenmanagement-Ansätze

In den letzten Jahren haben sich die Investitionen in Datentechnologien, Analyseanwendungen und Künstliche Intelligenz durch Organisationen stetig erhöht. Allerdings zeigen Studien, dass sich der erhöhte Einsatz finanzieller Mittel nicht zwangsläufig in einer verbesserten Datenwertschöpfung, Datenkultur oder Innovationsfähigkeit einer Organisation widerspiegeln (Bean, 2021). Als Ursache werden unter anderem Schwierigkeiten bei der Nutzung zentraler Datenmanagement-Architekturen wie Data Warehouses, Data Lakes oder Data Lakehouses identifiziert.

² Ein Beispiel hierfür ist das Konzept von Data Lakes, die fundamental auf dem Hadoop-Framework zur verteilten Speicherung und Verarbeitung von Daten aufsetzen.

Zu den Herausforderungen gehören aus technischer Sicht eine mangelnde Skalierbarkeit von Datenspeicherung und Datenverarbeitung an zentraler Stelle, ein hoher Aufwand zur Pflege von Datenpipelines, um Daten zentral zusammenzuführen, und eine geringe Flexibilität hinsichtlich der zum Datenmanagement zu nutzenden Technologien.

Zentralisierte Architekturen bedingen oftmals ebenso zentrale Datenteams, um deren personelle Ressourcen die Fachbereiche zur Erstellung von Datenlösungen konkurrieren müssen (Machado et al., 2022). Dies macht langfristige Planungen notwendig und verringert die Flexibilität bei der Erstellung von Datenlösungen. Zusätzlich wird in diesem Zusammenhang oftmals die Verantwortung über Daten an diese zentralen Bereiche übertragen. Somit sind letztendlich Personen für Datenqualität, Datengouvernanz und Datenverwaltung verantwortlich, die nur ein geringes Fachwissen über die Daten besitzen und keinen Einfluss auf die Datenerzeugung nehmen können (Bitkom e. V., 2022). Eine von vielen Organisationen anvisierte Datendemokratisierung³ wird zudem durch das isolierte Wissen über Daten erschwert. Beispielsweise fehlt einzelnen Fachbereichen die Kenntnis, welche Daten überhaupt verfügbar und abrufbar sind und wie diese genutzt werden können (Lefebvre & Legner, 2022). Zudem entstehen Risiken, wenn relevantes Wissen über den Personenbezug oder die Sensitivität von Daten nicht verfügbar ist und diese Daten in der Konsequenz unsachgemäß verarbeitet, gespeichert oder verfügbar gemacht werden. Viele dieser Probleme werden durch die ständig wachsenden Datenmengen und die Erwartungen von Wettbewerb und Gesellschaft, datengestützte Entscheidungen zu treffen und datengestützte Lösungen anzubieten, weiter verschärft.

2.3. Die Kernprinzipien von Data Mesh

Data Mesh setzt an diesen Herausforderungen an und zielt auf eine agile und skalierbare Datenmanagement-Organisation ab, die es letztendlich ermöglicht, den Wert von Datenmanagement-Investitionen zu heben. Die im Rahmen von Data Mesh stattfindende Transformation von Strukturen und Prozessen wird üblicherweise anhand von vier Kernprinzipien charakterisiert (Dehghani, 2022). Diese sind in Abbildung 1 dargestellt.

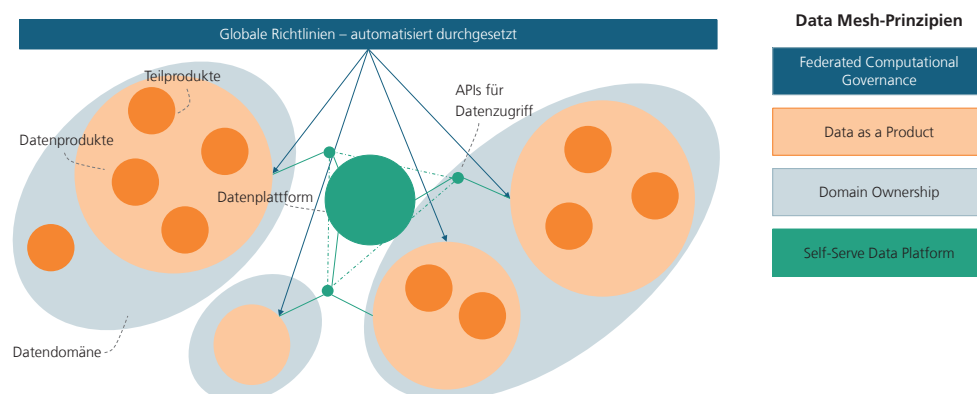


Abbildung 1: Data Mesh-Prinzipien

³ Unter dem Konzept Datendemokratisierung wird die Befähigung eines breiten Spektrums von Beschäftigten zur Nutzung von Daten verstanden.

Basierend auf dem Prinzip des *Domain Ownership* tritt ein dezentralisiertes Verantwortungsmodell anstelle des zuvor beschriebenen zentralisierten Dateneigentums. Ziel ist es, Flexibilität und Agilität im Datenmanagement zu steigern. Dazu wird die Datenorganisation auf Basis der fachlichen Organisationsstruktur in Geschäftsbereiche, sogenannte (Daten-) Domänen, aufgeteilt. Die Verantwortung für Daten, beispielsweise hinsichtlich Datenqualität, Verständlichkeit und Verfügbarkeit, wird unter den Domänen anhand ihrer fachlichen Expertise aufgeteilt. Entsprechend benötigen die Domänen verstärkte Datenkompetenzen und Experten, die ein adäquates Management der Daten aus technischer Sicht sicherstellen können. Andersherum erhalten die Domänen eine erweiterte Autonomie hinsichtlich der zu verwendeten technischen Toollandschaft, solange die von der Domäne geteilten Daten, die auf globaler Ebene definierten Standards erfüllen.

Auf Basis des Prinzips *Data as a Product* werden aus dem Produktmanagement bekannte Konzepte von den Domänenteams entlang des Datenlebenszyklus auf die Daten appliziert. Dies beinhaltet, dass Daten mit derselben Sorgfalt wie herkömmliche Produkte bereitgestellt werden. Die daraus resultierenden »Datenprodukte« sind darauf ausgelegt, mit den Datennutzenden anderer Domänen geteilt und von diesen verwendet zu werden. Entsprechend müssen Daten auffindbar, verständlich, abrufbar und nutzbar sein. Gleichermaßen sind Datenschutz und Datensicherheit zu gewährleisten. Dazu ist eine umfassende Beschreibung der Datenprodukte mittels Metadaten notwendig. Die einzelnen Datenprodukte können jeweils aus zugrundeliegenden Teilprodukten bestehen, die aus verschiedenen Quellsystemen stammen. Datenprodukte werden gemäß dem Zielbild als sogenanntes Data Quantum bereitgestellt. Ein Data Quantum enthält neben den eigentlichen Daten alle relevanten Metadaten, Richtlinien, Codes und die zugrundeliegende Infrastruktur, um eine skalierbare Wiederverwendung zu gewährleisten. Oftmals wird im Zusammenhang mit Datenprodukten auch das Konzept von Datenverträgen erwähnt. Bei Datenverträgen handelt es sich um eine Vereinbarung von Datenbereitstellenden und Datennutzenden über die mit den Daten verbundenen Erwartungen – ähnlich eines Service-Level-Agreement. Sie beschreiben einerseits die Bedeutung und Qualitätseigenschaften von Daten und andererseits, welche Voraussetzungen durch Datennutzende für den Abruf der Daten geschaffen werden müssen. Datenverträge werden üblicherweise in maschinenlesbarer Sprache formuliert und können somit eine automatisierte Umsetzung von Datenzugriffs- und Datennutzungsregeln befähigen.

Aus dem Zusammenspiel der Prinzipien *Data as a Product* und *Domain Ownership* resultieren einige Implikationen: Zum einen Bedarf es in jeder Domäne verschiedener Rollen, die zur Verwaltung von Daten als Produkt kollaborieren. Dazu gehören beispielsweise Datenproduktmanager, die eine anforderungsorientierte Entwicklung von Datenprodukten gewährleisten und Qualität, Nutzbarkeit und Gouvernanz von Datenprodukten sicherstellen. Zusätzlich sorgen Datenproduktentwickler für die technische Bereitstellung, Integration, Dokumentation und kontinuierliche Weiterentwicklung der Daten. Datendomänen lassen sich entlang ihrer Position im Datenlebenszyklus in drei Typen unterteilen. Quellenorientierte Domänen machen Daten aus (operativen) Systemen für andere Domänen verfügbar. Aggregierende Domänen kombinieren Daten aus verschiedenen quellenorientierten Domänen und machen diese nutzbar. Nutzungsorientierte Domänen verarbeiten Daten im Sinne von spezifischen Bedürfnissen und Anwendungsfällen der Endnutzenden und stellen die dort genutzten Daten zur Verfügung.

Das Prinzip der *Self-serve Data Platform* zielt darauf ab, Infrastrukturkomponenten bereitzustellen, mit denen sich übergreifende Dienste wie Datenentdeckung, Datenherkunft, Datenbeobachtung oder Wissensgraphen einfach umsetzen lassen. Diese Komponenten werden dazu durch ein dediziertes Infrastrukturteam entwickelt und verfügbar gemacht, sodass diese durch die Domänenteams einfach und eigenständig genutzt werden können. In diesem Sinne erfolgt einerseits eine Abstraktion der technischen Umsetzung und eine damit verbundene Komplexitätsreduktion für die Domänenteams. Andererseits findet ebenso eine technische

Standardisierung durch die Festlegung von APIs und Schnittstellen für den Datenzugriff statt, wodurch die Zusammenarbeit zwischen den einzelnen Domänen erleichtert wird.

Mit dem Prinzip der *Federated Computational Governance* wird eine dezentrale Datengouvernanz gefördert, in welcher jedes Domänenteam die Autorität über seine eigenen Datenprodukte und Prozesse wahrt. Um die organisatorische Interoperabilität der Daten zu gewährleisten, definieren die Datendomänen gemeinsam einen Satz globaler Richtlinien, die auf die einzelnen Datenprodukte appliziert werden. Dazu gehören insbesondere Belange wie die Gewährleistung von Datenschutz, Sicherheit und Compliance. Die Durchsetzung dieser Richtlinien erfolgt programmatisch. Dazu müssen die entsprechenden Richtlinien als automatisiert durchsetzbarer Code bereitgestellt werden, der entsprechend versioniert und getestet ist.

2.4. Data Mesh als soziotechnisches System

Kerncharakteristik und besonderes Abgrenzungsmerkmal zu vorangegangenen Datenmanagementansätzen stellt die explizite Betrachtung von Data Mesh als soziotechnisches System dar. Mit der Einnahme einer soziotechnischen Perspektive auf das Arbeitssystem Datenmanagement rückt insbesondere das soziale Teilsystem in Form der beteiligten Individuen verstärkt in den Fokus. Insgesamt umfasst das soziotechnische System Data Mesh das Zusammenspiel der Individuen, den organisatorischen Rahmen und die verwendeten Technologien.

Die konkreten Implikationen können anhand der Darstellung von Data Mesh im MTO-Modell (vgl. Ulich (2015)) verdeutlicht werden (s. Abbildung 2). Das MTO-Konzept basiert auf einem integrativen Ansatz zur Arbeitsgestaltung. Dabei werden Mensch, Technik und Organisation (MTO) als miteinander verbundene Teilsysteme zur Erfüllung einer Arbeitsaufgabe verstanden.

Im Kern von Data Mesh als soziotechnisches System steht die Betrachtung der *Arbeitsaufgabe*. Diese besteht aus der möglichst gewinnbringenden Bereitstellung von Datenprodukten zur Ermöglichung von Datenwertschöpfung. Im Hinblick auf die Erfüllung der Aufgabe gilt es, die Nutzung und Qualifizierung der Individuen (Mensch), den Einsatz von Datenmanagement-Technologien (Technik) und die organisatorische Gestaltung des Datenmanagements (Organisation) gemeinsam und im Einklang miteinander zu optimieren.

Das *soziale System* bezieht sich auf die sozialen und menschlichen Faktoren innerhalb eines Unternehmens oder einer Organisation wie zum Beispiel Kommunikation, Motivation und Kompetenzen der Individuen. Wird die Datenmanagementorganisation nach Data Mesh-Prinzipien aufgebaut, so entstehen mit den Domänen einzelne soziale Systeme, die unterschiedliche Eigenschaften hervorbringen können. Eine wichtige Konsequenz von Data Mesh-Umsetzungen ist die stärkere Durchdringung datengetriebener Kultur, Kollaboration und Transparenz, die von allen Beteiligten in den Domänen gelebt werden muss. Auf individueller Ebene wird eine kontinuierliche Verbesserung der Datenkompetenzen und die verstärkte Übernahme der Verantwortung für Daten (insbesondere von Fachexperten) gefordert. In diesem Zusammenhang gilt es, Produktmanagement-Kompetenzen aufzubauen und einen kulturellen Wandel hin zu einem Produkt- und Servicedenken im Umgang mit Daten zu realisieren.

Der Bereich *Organisation* bezieht sich auf die strukturellen und prozessualen Elemente eines soziotechnischen Systems. Aus organisatorischer Sicht müssen die einzelnen Domänen durch die Erarbeitung eines gemeinsamen Leitbilds aufeinander abgestimmt werden. Datenführung

und Datenanwaltschaft spielen eine entscheidende Rolle bei der Förderung der Datennutzung innerhalb der Organisation. So müssen Führungskräfte die strategische Nutzung von Daten fördern und den gewinnbringenden Nutzen von Daten hervorheben. Ein weiterer wichtiger organisationaler Faktor ist die Schaffung effizienter föderierter Gouvernanzstrukturen. Dazu gehört die Einteilung der Datenmanagement-Organisation in Domänen und die Bildung beschlussfähiger Komitees. Weiterhin müssen innerhalb der autonom agierenden Domänen sinnvolle Aufgabenteilungen und Datenmanagementprozesse geschaffen werden. Zusätzlich sind die gemeinsamen Richtlinien für den Umgang mit Daten festzulegen und entsprechende Mechanismen zu deren Umsetzung zu definieren.

Im Zentrum des *technischen Subsystems* steht die globale Bereitstellung der Self-Service Dateninfrastruktur. Zusätzlich erhalten die Domänen die Möglichkeit, individuelle Technologien basierend auf den individuellen Zielsetzungen und Kompetenzen einzusetzen. Wichtige Fähigkeiten stellen dabei die Bereitstellung eines Datenkatalogs zur globalen Datensichtbarkeit, die technische Erzeugung von Data Lineage zur Übersicht der Datenherkunft sowie die technische Umsetzung von Datenverträgen zwischen Datenbereitstellenden und Datennutzenden dar.

Die einzelnen Teilsysteme stehen in bei der Realisierung des Gesamtsystems in starker Abhängigkeit zueinander. So müssen die organisatorischen Richtlinien mittels technischer Mechanismen durchgesetzt werden. Andersherum ist für die passende Umsetzung der technischen Systeme eine auf die Ziele der Datenmanagement-Organisation ausgerichtete strategische Planung notwendig. Einzelne Individuen können durch effizientes Leadership und Kommunikationsmaßnahmen in deren Umgang mit Daten bestärkt werden. Andersherum geben Individuen wertvolle Rückmeldung bezüglich der Effektivität und Umsetzbarkeit der organisatorischen Maßnahmen.

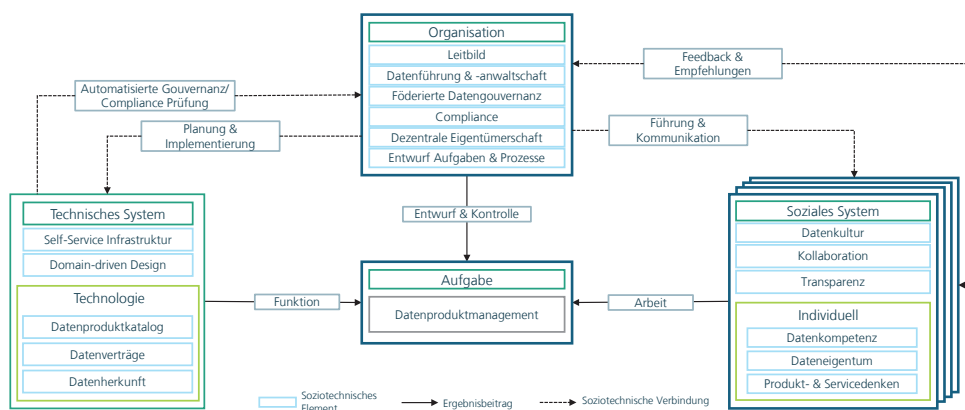


Abbildung 2: Soziotechnisches System Data Mesh

2.5. Data Mesh im Vergleich mit weiteren Datenmanagement-Konzepten

Aufgrund des oben beschriebenen soziotechnischen Ansatzes von Data Mesh ist es schwierig, Data Mesh in seiner Gesamtheit mit anderen, in der Regel technisch orientierten Ansätzen des Datenmanagements zu vergleichen. Dennoch ist es wichtig, Data Mesh von diesen abzugrenzen

und zu verstehen, wie diese Architekturansätze in das Data Mesh-Konzept integriert werden können. Die technischen Teilaspekte der existierenden Konzepte können genutzt werden, um spezifische Bereiche des Data Mesh-Ansatzes zu realisieren. Die Einzellösungen können sich gegenseitig ergänzen. Im ersten Schritt wird Data Mesh dazu mit Ansätzen, die sich auf die Speicherung und die Verwaltung von Daten fokussieren, verglichen. Anschließend erfolgt eine Gegenüberstellung mit etablierten Konzepten zur Verfügbarmachung von Daten.

Data Lake

Ein Data Lake ist eine skalierbare Umgebung zur Verwaltung von strukturierten, semistrukturierten und unstrukturierten Daten in beliebigem Umfang. Data Lakes ermöglichen zudem die Verwaltung von Daten-Streams. Auf der Basis von Cloud Computing bieten Data Lakes eine flexible und kosteneffiziente Lösung für die Speicherung großer Datenmengen. Ein Data Lake speichert Daten ohne explizites Datenschema. Daher können Daten mit geringen Aufwänden aus ihren Ursprungssystemen in einen Data Lake überführt werden. Das Datenschema wird üblicherweise erst beim Datenabruf erzeugt. Entsprechend eignen sich Data Lakes insbesondere zur Speicherung von Daten ohne zuvor konkret definierte Anwendungsfälle. Innerhalb des Data Lakes können verschiedene Zonen definiert werden, die Daten in unterschiedlichen Qualitätszuständen enthalten können. Kommerzielle Cloud-Lösungen kommen zusätzlich mit integrierten Tools zur Datenintegration und Transformation. Die Nutzung von Data Lakes erfolgt aufgrund ihrer Komplexität üblicherweise durch Datenanalytistinnen und Datenanalysten sowie Datenwissenschaftlerinnen und Datenwissenschaftler für Anwendungen aus den Bereichen der Datenanalyse und der Künstlichen Intelligenz. Als stark technologiegetriebener Ansatz rücken Elemente von Datengouvernanz und Metadatenmanagement zunächst in den Hintergrund. Entsprechend ergeben sich viele gescheiterte Data Lake-Projekte, in denen Daten zwar abgelegt, aber aufgrund mangelnder Beschreibung und Nutzungsrichtlinien nicht abrufbar waren. In diesen Fällen wird eine Data Lake-Umsetzung auch als »Data Swamp« bezeichnet.

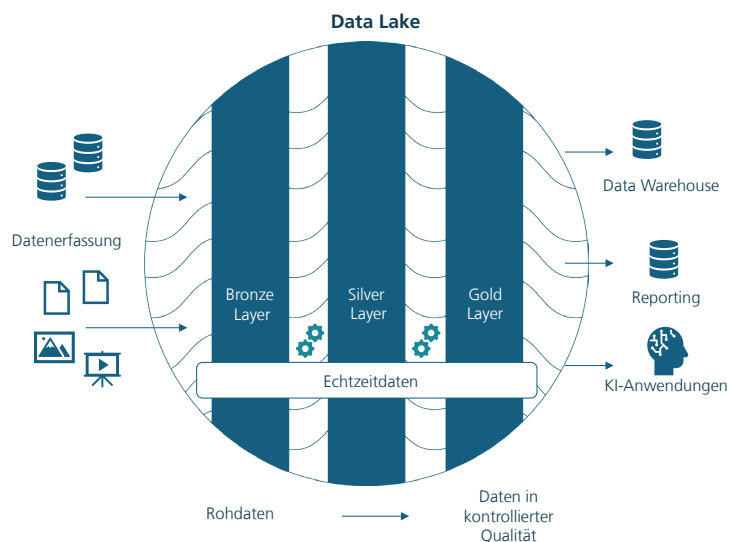


Abbildung 3: Data Lake

Während Data Lake und Data Mesh grundsätzlich unterschiedliche Ansätze verfolgen, spielen derzeit aufgrund des Fehlens dedizierter Data Mesh-Technologien Data Lakes auch bei der technischen Realisierung von Data Mesh eine entscheidende Rolle. Anstelle eines zentralen Data Lake werden nun kleinere Data Lakes innerhalb der einzelnen Domänen genutzt, um Daten zu integrieren, aufzubereiten und zur Verfügung zu stellen. Mechanismen wie »Serverless

Computing« und »Infrastructure as Code« ermöglichen dabei eine einfache Administration und Wiederverwendbarkeit erzeugter Bausteine. Finale Datenprodukte können entsprechend mit weiteren Domänen geteilt werden. Interne Datenbearbeitungsschritte bleiben für diese jedoch teilweise verborgen.

Data Lakehouse

Ein Data Lakehouse ist eine moderne Datenarchitektur, die die Vorteile von Data Lakes und Data Warehouses kombiniert. Es ermöglicht die Speicherung von strukturierten, semistrukturierten und unstrukturierten Daten in einem einheitlichen Repository. Wie ein Data Lake speichert ein Data Lakehouse Daten in ihrem nativen Format. Durch die spezielle Formatierung von Daten im Data Lake (beispielsweise durch Apache Iceberg⁴) werden Datenbearbeitungen mittels Metadaten aufgezeichnet. Dies ermöglicht die Erstellung von virtuellen Data Warehouses (konsolidierten Übersichten über relationale Daten), ohne Daten aus einem Data Lake replizieren zu müssen. Das Datenschema wird dabei auf Abruf erzeugt. Dank ihrer Architektur und der Verwendung von speicheroptimierten Formaten ermöglichen Data Lakehouses eine kosteneffiziente und skalierbare Lösung zur Speicherung großer Datenmengen, während sie gleichzeitig eine hohe Datenqualität gewährleisten. Dies reduziert das Risiko von »Data Swamps«, da Metadaten generiert und Nutzungsvorgaben implementiert werden. Data Lakehouses eignen sich in besonderem Maße für Datenanalytikerinnen und Datenanalysten sowie Datenwissenschaftlerinnen und Datenwissenschaftler, die eine effiziente und nutzungsfreundliche Plattform für Datenanalysen und Künstliche Intelligenz benötigen. Die strukturierten Daten aus den virtuellen Data Warehouses können zudem durch weniger qualifizierte Nutzende beispielsweise für Anwendungen aus dem Bereich der Business Intelligence verwendet werden.

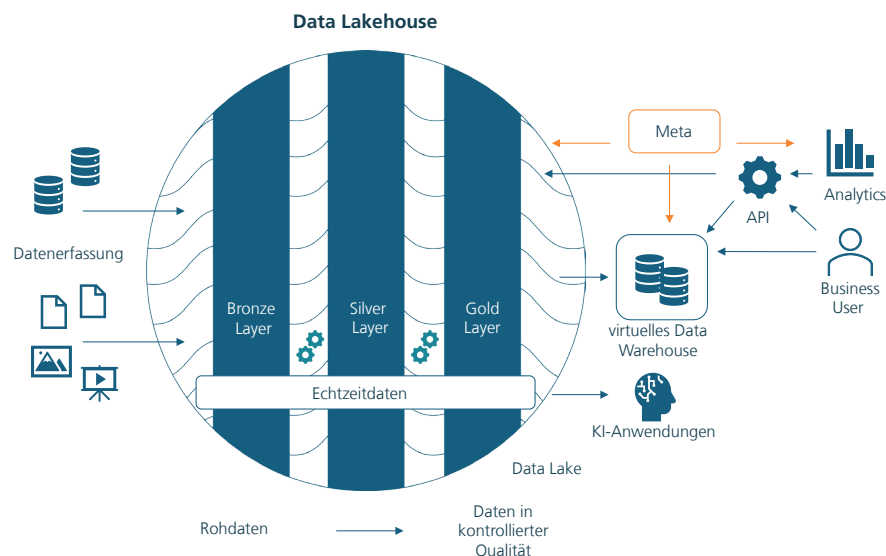


Abbildung 4: Data Lakehouse

Data Lakehouses adressieren ähnliche Beschränkungen konventioneller Data Lakes wie Data Mesh. Allerdings basieren auch diese auf einer zentralisierten Architektur und bedingen damit ähnliche Problemstellungen wie beispielsweise Engpässe bei der Datenintegration. Data Lakehouses eignen sich im Vergleich zu Data Mesh insbesondere für kleinere Organisationen

⁴ <https://iceberg.apache.org/>

mit geringer Datenkomplexität, geringerem Datenreifegrad und geringeren finanziellen Möglichkeiten, in denen beispielsweise eine Aufteilung in Datendomänen schwierig umsetzbar ist. Gleichzeitig können Data Lakehouses analog zu Data Lakes auch von einzelnen Datendomänen im Data Mesh eingesetzt werden. Mithilfe von Data Lakehouses lassen sich dabei verschiedenste Anwendungsfälle umsetzen. Zudem werden Datennutzbarkeit und Datenqualität gefördert.

Data Fabric

Data Fabric ist ein architektonischer Ansatz, welcher eine einheitliche und konsistente Erfahrung zur Verwaltung von Daten aus zuvor isolierten Umgebungen (bspw. Cloud, on-premises, hybride Umgebungen) anvisiert. Im Zentrum von Data Fabric steht die nahtlose technische Bereitstellung von Datenmanagement-Fähigkeiten. Dazu werden verschiedene datenhaltende Systeme sowie Datenintegrations-, Datenverwaltungs- und Datengouvernanz-Tools miteinander kombiniert, sodass wiederverwendbare Datenmanagement-Bausteine entstehen. Data Fabric basiert zudem auf dem Konzept der Datenvirtualisierung. Datenvirtualisierung ist eine Technik, die es ermöglicht, Daten aus verschiedenen Quellen an einem Ort zu nutzen, ohne dass die Daten physisch kopiert oder verschoben werden müssen. Das Data Fabric-Konzept wird vorrangig von der Technologieberatung Gartner propagiert (Priebe et al., 2021). Zur Gewährleistung von Produktivität durch die Integration, Automatisierung und Wiederverwendbarkeit einzelner Bausteine der Data Fabric sind umfassende Metadaten notwendig. Metadaten werden in Form von Wissensgraphen modelliert, welche die Kernkonzepte eines Systems und deren Beziehungen zueinander in einer von Menschen und Computern interpretierbaren Form speichern. Die Metadaten können entweder manuell durch den Menschen, durch die Extraktion von Daten aus bestehenden Systemen oder mittels Künstlicher Intelligenz erzeugt werden.

Capabilities	Tools	
Data Discovery	Data Catalog / Data Marketplaces / Data Governance Tools	
Data Access and Policy enforcement		
Data Preparation and Delivery	Data Pipelines	APIs
Real-Time Processing	Master Data Management Software	
Master Data Management		
Code and Service Sharing	Reusable Data Pipelines / Notebooks / Scripts / Job Definitions	

Abbildung 5: Beispielhafte Umsetzung einer Data Fabric

Sowohl für die Realisierung von Data Mesh als auch zur Realisierung von Data Fabric sind hochqualitative und umfassende Metadaten notwendig (Blohm et al., 2024). Beide Konzepte setzen die Bereitstellung eines Datenkatalogs voraus, um die Automatisierung des Datenmanagements entlang des Datenlebenszyklus voranzutreiben. Ansätze aus dem Bereich der Data Fabric können zudem im Data Mesh zur Bereitstellung und Weiterentwicklung der Self-Service Infrastruktur verwendet werden, indem eine fähigkeitsbasierte Perspektive eingenommen wird.

Tabelle 1 zeigt die präsentierten Datenmanagement-Architekturen im Vergleich. Zusätzlich existieren weitere Ansätze, die insbesondere auf eine Mehrfachnutzung von Daten in organisationsinternen und organisationsübergreifenden Anwendungsfällen abzielen. Hierzu gehören die Ansätze Daten-Hub und Datenmarktplatz, die im Folgenden erläutert werden.

Daten-Hub

Ein Daten-Hub ist eine zentrale Plattform, die den Zugriff auf kuratierte Daten zur weiteren Verarbeitung, beispielsweise im Rahmen von Reporting-Anwendungen ermöglicht. Kernaufgabe eines Daten-Hubs ist die Echtzeit-Integration und Verteilung von Daten. Daten-Hubs können sowohl strukturierte als auch unstrukturierte Daten verwalten und bieten Organisationen eine flexible Lösung zur Datenorganisation. Dazu stellen Daten-Hubs eine zentrale Schnittstelle für den Datenaustausch und die Integration bereit. Durch den daraus resultierenden einheitlichen Zugriff auf aktuelle Daten steigert ein Daten-Hub die Effizienz und Genauigkeit von Geschäftsprozessen. Ein Daten-Hub kann zudem, je nach Auslegung, Daten aus den Quellsystemen replizieren und in diesem Sinne die Datenqualitäts- und Datensicherheitsaspekte in einem zentralen Punkt durchsetzen.

Tabelle 1: Vergleich von Datenmanagement-Architekturen nach Serra (2024)

	Data Lake	Data Lakehouse	Data Fabric	Data Mesh
Vorteile	Skalierbarkeit, Flexibilität, Kosteneffizienz	Kosteneffizienz, Skalierbarkeit, Datenqualität, Transaktionssicherheit	Einheitlicher und effizienter Datenzugriff, Verbesserte Gouvernanz, Produktivität	Skalierbarkeit, Verantwortung & Fachwissen, Agilität
Anwendungsbereich	Organisationen mit großen Mengen unterschiedlicher Datentypen und Bedarfen nach Skalierbarkeit.	Organisationen mit großen Mengen unterschiedlicher Datentypen und Bedarfen nach Skalierbarkeit sowie Anwendungen im Bereich Business Intelligence und Analytics.	Organisationen, die eine einheitliche Datenmanagement-Erfahrung für Daten in verteilten Umgebungen sicherstellen möchten.	Große und komplexe Organisationen mit Domänen, die unabhängig arbeiten und dennoch Bedarfe nach der gemeinsamen Bewirtschaftung analytischer Daten besitzen.
Architektur	Ein zentralisierter Speicherort für große Mengen von Daten in ihrem ursprünglichen Format.	Wie Data Lake mit zusätzlicher Möglichkeit zur schnellen Abfrage strukturierter Daten in vordefiniertem Schema.	Zentralisierte Datenmanagement-Ebene, die einen einheitlichen Datenzugriff und Verwaltung von Daten in verschiedenen Umgebungen ermöglicht.	Dezentralisierter Ansatz mit Fokus auf Autonomie einzelner Domänen und Förderierung durch gemeinsame Gouvernanz und Dateninfrastrukturkomponenten.
Gouvernanz	Geringes Maß an Kontrolle, da Daten oftmals ohne vorher definierten Verwendungszweck verwaltet werden.	Höhere Kontrolle und Nachvollziehbarkeit insbesondere für explizit verwaltete Daten.	Zentralisierte Verwaltung, die einheitliche Richtlinien und deren Einhaltung gewährleistet.	Föderierte Gouvernanz, in der geteilte Datenprodukte einen hohen Grad der Kontrolle durch die Datenbereitstellenden aufweisen.

Die Vorteile eines Daten-Hub resultieren insbesondere in der Schaffung eines zentralen Zugriffspunkts, der den Abruf von Daten beschleunigt, Datenmodelle konsolidiert und Kollaboration zwischen Datenbereitstellenden und Datennutzenden fördert. Allerdings sind vorab größere Aufwände zur Datenintegration notwendig, teilweise ohne den genauen Nutzen der Daten bewerten zu können. Aufwände sind insbesondere zur Überführung in die vordefinierten Schemata und die Gewährleistung der Datenqualitätskennzahlen notwendig. Zusätzlich erfordern Aufbau und Wartung eines Daten-Hubs umfassende Investitionen in Technologie und den Aufbau von Fachwissen (Bühlmann, 2024). Während Data Mesh auf die Autonomie von einzelnen Datenbereitstellenden und geringe Vorabaufwände zur Datenintegration abzielt, setzt das Daten-Hub-Konzept auf eine Zusammenführung der Datenbereitstellung und die zentrale Kontrolle über Daten.

Datenmarktplatz

Eine Struktur ähnlich zu Daten-Hubs stellen Datenmarktplätze dar. Der Begriff »Datenmarktplatz« bezeichnet eine Plattform, die es ermöglicht Datennachfrage und Datenangebot zusammenzuführen und Daten abrufbar zu machen. Ursprünglich resultierend aus dem organisationsübergreifenden Kontext, werden Datenmarktplätze nun auch vermehrt innerhalb von Organisationen umgesetzt. Zur Erleichterung des Datenaustausches bieten Datenmarktplätze verschiedene Funktionalitäten. Dazu gehört eine Infrastruktur, um Datenprodukte oder Datendienste sicher zu teilen und zu nutzen. Datenmarktplätze schaffen einen Rahmen für Teilnahmeregeln, Effizienzsteigerung und Transaktionsabwicklung durch standardisierte Schnittstellen und Dienste (Jahnke et al., 2024). Zur Population eines Datenmarktplatzes registrieren Datenbereitstellende im ersten Schritt die von ihnen zur Verfügung gestellten Datenprodukte. Datenmarktplätze werden üblicherweise von einem Betreiber koordiniert. Dieser gilt als Vertrauensanker, indem er beispielsweise die Qualität der angebotenen Daten sicherstellt.

Andererseits können Marktplatzbetreibende auch zusätzliche Services wie Datenintegrationsdienste anbieten. Datenmarktplätze können entweder die Daten direkt replizieren (ähnlich eines Daten-Hub) oder lediglich den Zugang zu diesen in Drittsystemen managen. Vorteile eines Datenmarktplatzes liegen in der Schaffung einer Übersicht über die verfügbaren Daten bzw. Datenprodukte und in der Automatisierung von Transaktionsprozessen, beispielsweise hinsichtlich des Datenabrufs. Somit können datenbasierte Innovationen stimuliert werden. Die erweiterte Verfügbarkeit von Daten ermöglicht ihrerseits die Verbesserung von Produkten, Dienstleistungen oder Prozessen (Eichler et al., 2023). Im Gegensatz zu Daten-Hubs muss dabei nicht zwangsläufig eine physische Integration der Daten erfolgen. Abschnitt 2.6 beschreibt die Nutzung eines Datenmarktplatzes als Architekturmuster im Kontext von Data Mesh.

Tabelle 2: Vergleich der Konzepte zur Verfügbarmachung von Daten

Charakteristik	Daten-Hub	Daten Marktplatz	Data Mesh
Vorteile	Zentrale Datenintegration, Echtzeit-Datenzugriff, Flexibilität und Kollaboration	Erleichterter Datenaustausch, Kollaboration, Skalierbarkeit	Skalierbarkeit, Verantwortung & Fachwissen, Agilität
Anwendungsbereich	Daten-Hub wird zur Verfügbarmachung qualitativ hochwertiger Daten zu Analyse-zwecken verwendet. Daten-Hubs eignen sich besonders für die Zusammenführung von Daten gleicher fachlicher Bereiche.	Orchestrierung der Mehrfachnutzung von Daten und Vereinfachung von Datentransaktionen ⁵ . Daten müssen dazu als Produkte (inklusive umfassender Metadaten) bereitgestellt werden.	Große und komplexe Organisationen mit Domänen, die unabhängig arbeiten und dennoch Bedarfe nach der gemeinsamen Bewirtschaftung analytischer Daten besitzen.
Architektur	Zentrale Schnittstellen für den Datenabruf, basierend auf physischer Datenintegration.	Zentrale oder dezentrale Architektur möglich. Metadaten werden üblicherweise an zentraler Stelle gehalten.	Dezentralisierter Ansatz mit Fokus auf Autonomie einzelner Domänen und Förderung durch gemeinsame Gouvernanz und Dateninfrastrukturkomponenten.
Gouvernanz	Stark kontrollierte Datenformate und Datenqualität.	Betreiber eines Datenmarktplatzes als Koordinator und Orchestrator des Datenaustausches.	Föderierte Gouvernanz, in der geteilte Datenprodukte einen hohen Grad der Kontrolle durch die Datenbereitstellenden aufweisen.

2.6. Data Mesh-Topologien

Bei der Umsetzung von Data Mesh ergeben sich verschiedene Interpretationen durch Organisationen. Stets basierend auf den vier Kernprinzipien von Data Mesh werden konkrete Architekturen anhand der individuellen Anforderungen und der Datenstrategie einer Organisation entwickelt. Zu den Einflussfaktoren zählen dabei unter anderem die Größe und Komplexität der Organisationen, existierende Systemlandschaften für Stamm- und Transaktionsdaten, die

⁵ Eine Datentransaktion umfasst die gesamte Interaktion zwischen Datenbereitstellenden und Datennutzenden zur Ermöglichung der Mehrfachnutzung von Daten. Bestandteile einer Datentransaktion sind unter anderem die Identifizierung nutzbarer Daten, die Datenanfrage, die Prüfung der Anfrage sowie die Freigabe und der eigentliche Transfer der Daten von Datenbereitstellenden zu Datennutzenden.

Kompetenzen der Mitarbeitenden oder die Bedürfnisse nach zentralisierter Übersicht und Kontrolle. Eine Einordnung individueller Data Mesh-Architekturen kann anhand der Zuordnung zu Data Mesh-Topologien⁶ (Strengtholt, 2023) erfolgen. Diese können, wie in Abbildung 6 gezeigt, entlang des Spektrums von dezentral zu zentral klassifiziert werden.

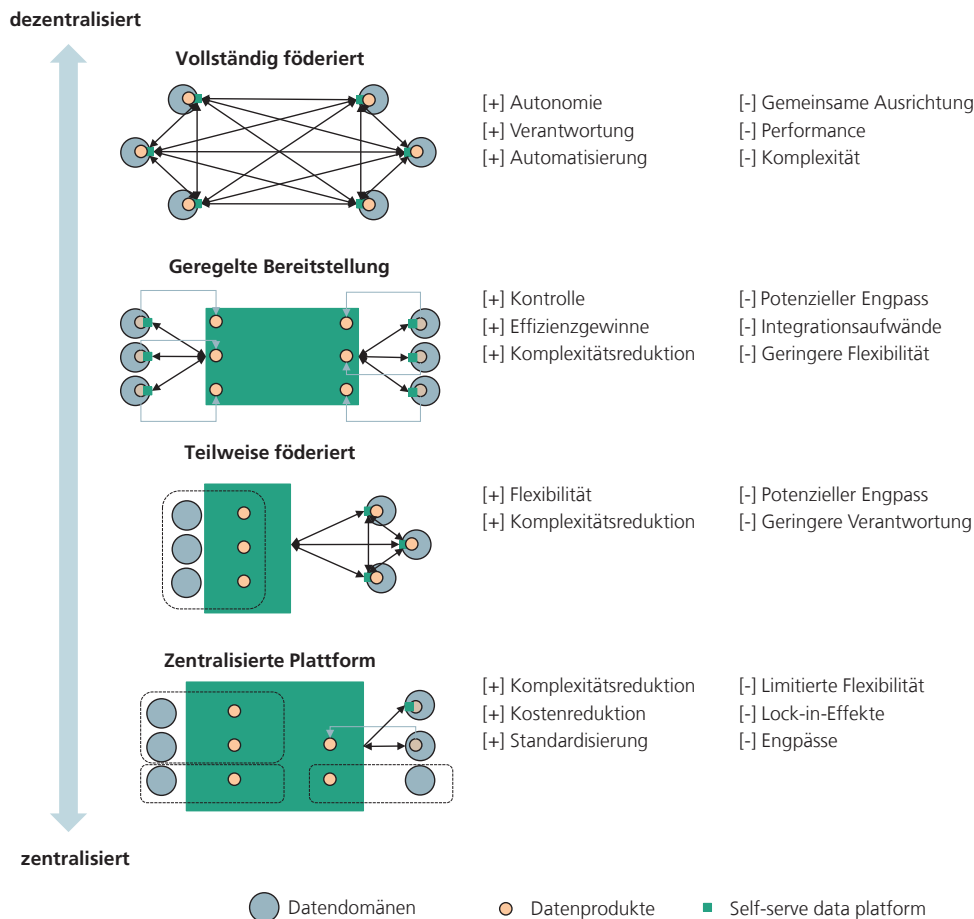


Abbildung 6: Data Mesh-Topologien

In einem *vollständig föderierten* Data Mesh erhalten einzelne Domänen vollständige Autonomie über ihre Organisation und die Art und Weise, wie Datenprodukte bereitgestellt werden, solange diese gemeinsamen Standards und Richtlinien entsprechen. Dabei existiert kein zentraler technischer oder organisatorischer Anker. Daten werden direkt zwischen den einzelnen Domänen ausgetauscht. Die Nutzung gemeinsamer Dateninfrastrukturdienste fokussiert hier auf die Sicherstellung von Datenschutz und Datensicherheit. Diese Topologie setzt die individuelle Verantwortung einzelner Domänen für Daten voraus und fördert die automatisierte Bereitstellung von Daten. Andererseits ergeben sich Herausforderungen bei der Abstimmung eines gemeinsamen Zielbildes zwischen den Datendomänen. Die notwendige Dezentralisierung und Automatisierung führen zwangsläufig zu einer erhöhten Komplexität. Diese spiegelt sich einerseits in höheren Kosten – jede Domäne nutzt ihre favorisierten Dienste – und andererseits in hohen Anforderungen hinsichtlich der innerhalb der Domäne benötigten Datenmanagement-Kompetenzen wider. Gleichermäßen kann es aufgrund der isolierten Datenbehandlung zu Mehraufwänden bei Datenabruf oder Datenintegration kommen.

⁶ Der Begriff »Topologie« entstammt dem Feld der Netzwerktechnologien. Eine Topologie beschreibt die Objekte eines Netzes und deren Beziehungen untereinander.

Durch die Topologie einer *geregelten Bereitstellung* können letztgenannte Herausforderungen angegangen werden. Dabei werden dezentral entwickelte Datenprodukte über eine zentrale Datenaustauschebene zur Verfügung gestellt. Diese kann beispielsweise, in Anlehnung an einen Daten-Hub, die Auffindbarkeit von Daten vereinfachen und gewisse Datenqualitätsstandards oder Nutzungsrichtlinien durchsetzen. Weitere Zusatzservices können auf dieser zentralisierten Zugriffsebene aufsetzen. Entsprechend ergeben sich größere Möglichkeiten zur Einflussnahme auf die bereitgestellten Datenprodukte. Zusätzliche Dienste können durch eine gemeinsame Nutzung für Effizienzgewinne sorgen. Weiterhin ergibt sich insbesondere für Datennutzende eine Reduktion der Komplexität durch einen zentralen Kontaktpunkt. Andererseits wird die Autonomie und Flexibilität der individuellen Datendomänen reduziert. Für Datenbereitstellende entstehen zusätzliche Aufwände zur Integration ihrer Daten. Gleichermaßen kann die zentrale Schicht einen Bottleneck bei der Bereitstellung von Zusatzdiensten oder für Integrationsservices darstellen.

Im *teilweise föderierten Ansatz* verwenden einzelne Domänen – üblicherweise Datenbereitstellende – von einem geteilten Datenteam unterstützt, welches eine für diese Domänen zentralisierte Infrastruktur nutzt. Dies geschieht üblicherweise in Organisationen mit einer großen Anzahl von Legacy-Anwendungen oder geringeren Datenkompetenzen. Datendomänen mit größeren Datenmanagement-Kompetenzen – üblicherweise Datennutzende – können hierbei weiterhin im Sinne von Data Mesh als eigenständige Domäne agieren und ihre Autonomie und Flexibilität wahren. Dieser Ansatz reduziert die Komplexität für Domänen mit geringem Datenmanagement-Reifegrad. Allerdings ergeben sich durch die teilweise eingeführte Zentralisierung des Datenmanagements mögliche Bottlenecks bei der Datenbereitstellung und -nutzung. Zudem wird die Verantwortung für die Bereitstellung von Datenprodukten hin zu zentralen Einheiten verschoben.

Schlussendlich ist es auch möglich, Data Mesh durch Nutzung einer *gemeinsamen zentralisierten Plattform* zu realisieren. Diese wird üblicherweise von Cloud-Services-Providern oder dedizierten Anbietenden von Datenmanagement-Lösungen wie Snowflake, Databricks oder anderen bereitgestellt. Auf der Plattform werden Aufgaben wie die Datenintegration, die Datentransformation oder die Datenbereitstellung ausgeführt. Organisationen versuchen, diese Aufgaben zu automatisieren und Self-Service-Ansätze zu verfolgen, um die zentralisierten Datenplattform-Teams zu entlasten. Einzelne Domänen werden durch die Nutzung verschiedene Konzepte wie Cloud User, Container oder Projekte voneinander abgegrenzt. Zudem tragen einzelne Domänen üblicherweise die für sie individuellen Integrationsaufwände. Die Aufgabe der Domänen in diesem Ansatz fokussiert sich auf die Übernahme von Verantwortung für die bereitgestellten Daten und das Einbringen in die Festlegung von Gouvernanz-Richtlinien. Dieser Ansatz ermöglicht das Verfolgen von Data Mesh-Ansätzen auch für weniger reife Datenmanagement-Organisationen durch starke Komplexitätsreduktion für einzelne Domänen. Zudem können Kosten durch die Nutzung einer gemeinsamen Plattform vereinfacht überwacht und optimiert werden. Durch die Nutzung gemeinsamer Services und Datenmodelle erfolgt eine Standardisierung der Datenmanagementfunktionen, die zu einer verbesserten Interoperabilität von Daten führt. Auf der anderen Seite stehen Datendomänen unabhängig von ihrer Maturität nur eine festgeschriebene und begrenzte Auswahl von Diensten zur Datenbereitstellung zur Verfügung. Weiterhin können sich durch die starke Fokussierung auf einen Plattformanbieter Lock-in-Effekte ergeben. Wie bei anderen zentralisierten Ansätzen existiert insbesondere bei der zentralisierten Plattform das Risiko von Engpässen auf Seiten des Plattformteams.

2.7. Data Mesh-Architekturmuster

Bei der konkreten Realisierung von Data Mesh in Organisationen lassen sich einige technische und organisatorische Architekturmuster identifizieren. Diese Architekturmuster werden im Folgenden beschrieben.

Anwendung von »Bounded Contexts« zur Entwicklung von Datendomänen

Die Zuordnung der dezentralen Verantwortlichkeiten zu Daten stellt eine der fundamentalen Herausforderungen der Datengouvernanz im Data Mesh dar. Im Data Mesh sind Domänenteams bestehend aus technischen und fachlichen Rollen für das Management von Daten zuständig. Bei der Definition der Datendomänen und der Zuordnung von Datenobjekten zu diesen wird auf das Prinzip der »Bounded Contexts« aus der Softwarearchitektur zurückgegriffen. Bounded Contexts stellen abgrenzbare Bereiche einer Organisationsstruktur dar, in welchen ein einheitliches Verständnis über die verwendeten Konzepte, Regeln und Geschäftslogik herrschen. Diese Bounded Contexts entsprechen dabei nicht zwangsweise den Fachbereichen einer Organisation, sondern können auch Untergruppen oder Zusammenschlüsse aus diesen darstellen.

Abbildung 7 zeigt ein Beispiel solcher Bounded Contexts anhand eines imaginären Fernbusunternehmens Dreamline Buses. Zentrales Datenobjekt von Dreamline Buses stellt die Tabelle »Fahrten« dar, welche unter anderem Informationen zu Routen, Zeitplänen, Reisenden und Fahrzeugen enthält. Das Datenobjekt besitzt dabei für verschiedene Kontexte (bspw. den Kundendienst, das Controlling oder die Wartung) eine Relevanz. Beispielsweise wird das Datenobjekt im Bereich des Kundenmanagements zur Durchführung von Buchungen benötigt, während es im Bereich Wartung für die Planung von Wartungsintervallen notwendig ist. Diese Domänen haben dabei ein eigenes Verständnis über gewisse Konzepte innerhalb des Datenobjekts. Beispielsweise könnte ein Kunde für das Kundenmanagement eine natürliche Person sein, die eine Fahrt gebucht hat, während die Wartungsabteilung hingegen ihre Leistungen an interne Kunden vermittelt. Weiterhin existieren unterschiedliche Bedürfnisse an den Detailgrad der Informationen. Während für einen Kunden oder eine Kundin üblicherweise die Dauer einer Fahrt als Information ausreicht, sind für das Wartungsteam konkrete Informationen zu Geschwindigkeiten oder Streckenbeschaffenheit notwendig. Entsprechend werden diese vormals gemeinsam gepflegten Informationen nun zwischen den Bereichen aufgeteilt, sodass diese unabhängig voneinander und anhand der individuellen Anforderungen verwaltet werden können. Wichtig ist dabei, die Verantwortung für Daten angemessen aufzuteilen und Dateninteroperabilität zwischen den Bereichen zu schaffen, sodass die Daten bei Bedarf weiterhin miteinander geteilt werden können. Im Fall von Dreamline Buses sind beispielsweise Daten aus dem Bereich der Wartung für die Routenoptimierung nützlich. Im Rahmen einer Data Mesh-Einführung müssen die Datendomänen entsprechend der Kontexte definiert und »geschnitten« werden.

Dreamline Buses



Abbildung 7: Das Konzept »Bounded Context« anhand des Beispiels Dreamline Buses

Föderierte Gouvernanz zur Festlegung von Regeln und Richtlinien

Zur Festlegung einer gemeinsamen Interpretation von Kernkonzepten sowie von globalen Regeln und Richtlinien für den Umgang mit Daten wird im Data Mesh der Ansatz einer föderierten Gouvernanz verfolgt. Dazu kommen Verantwortliche aus den einzelnen Datendomänen in Form eines Komitees zusammen, um Entscheidungen zu treffen, die die Gesamtfunktionalität des Data Mesh betreffen. Der föderierte Ansatz ermöglicht eine gleichberechtigte Entscheidungsfindung. Gleichzeitig erhalten einzelne Domänen Autonomie bei Themen, die nicht auf globaler Ebene relevant sind. Beispielsweise können die Domänenteams weiterhin selbst entscheiden, unter welchen Bedingungen Daten abgerufen werden können. Auf globaler Ebene werden hierzu lediglich die Standards zur Beschreibung dieser Bedingungen festgelegt. Die föderierte Gouvernanz lässt sich mit dem föderalistischen System von Ländern wie der Schweiz vergleichen: Auf Bundesebene werden hier ausschließlich Themen behandelt, welche die Kantone nicht eigenständig lösen können. Abbildung 8 zeigt eine schematische Darstellung des Konzepts einer föderierten Gouvernanz.

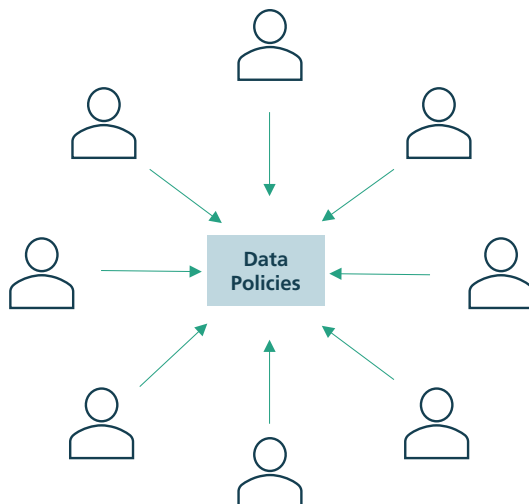


Abbildung 8: Schematische Darstellung der föderierten Gouvernanz

Event-Streaming-Plattform zur Bereitstellung von Daten in Echtzeit

Die Event-Streaming-Plattform stellt eines von zwei in diesem Dokument beschriebenen Mustern zur Bereitstellung von Daten im Data Mesh dar. Mit Hilfe von Event-Streaming-Plattformen werden die innerhalb von Datenprodukten generierten Ereignisse (z.B.: Creates, Updates, etc.) in Echtzeit den relevanten Datennutzenden zur Verfügung gestellt. Datenprodukte senden ihre Events über sogenannte Topics, die von den Datennutzenden abonniert werden. Ein Topic kann dabei ein oder mehrere Datenprodukte umfassen. Die über ein Topic erhaltenen Daten können von den Datennutzenden abgerufen und in weiteren Systemen langfristig gespeichert werden. Durch Daten-Streams entstehen ebenso Streams an Metadaten. Auch diese können über die Event-Streaming-Plattform weitergegeben werden. Der Einsatz von Event-Streaming-Plattformen eignet sich besonders für sich schnell ändernde Daten und vermeidet ständig wiederkehrende Datenabfragen.

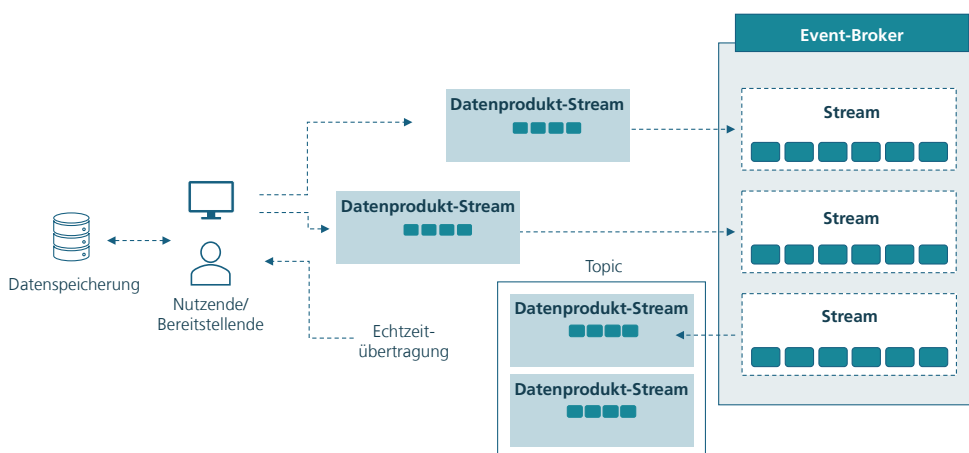


Abbildung 9: Schema einer Event-Streaming-Plattform

API-Gateway zur Bereitstellung von Daten in Batches

Ein weiteres Pattern zur Bereitstellung von Daten ist das API-Gateway. API-Gateways eignen sich insbesondere für Data Mesh-Umgebungen, in denen Daten ausschließlich zu bestimmten Zeitpunkten (Batches) aktualisiert werden und bei denen Echtzeitanalysen nicht benötigt werden. API-Gateways ermöglichen als zentraler Kontaktpunkt den Zugriff auf verschiedene Datenprodukte. Datenprodukte müssen sich dazu im ersten Schritt im API-Gateway registrieren. Das API-Gateway implementiert die Zugriffsverwaltung und steuert, welche Datennutzenden auf die APIs der Datenprodukte zugreifen können. Es stellt eine Abstraktionsschicht dar, indem es die Zugriffsanfragen an die entsprechenden Datenprodukte weiterleitet und die Daten bei Bedarf aggregiert. Ebenso ist es in der Lage, die Leistung der APIs zu überwachen und bei Bedarf Fehlermeldungen auszugeben.

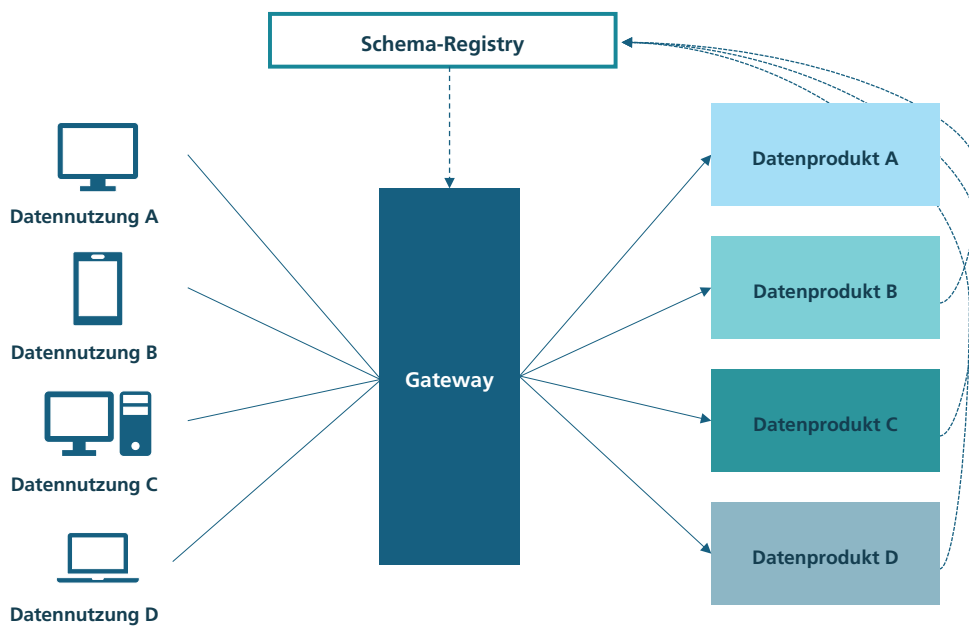


Abbildung 10: Schema eines API-Gateways

Datenmarktplätze und Datenkataloge zur Entdeckung von Datenprodukten

Eine wichtige Komponente von Data Mesh-Implementierungen stellen Datenmarktplätze beziehungsweise zugrundeliegende Datenkataloge dar. Datenkataloge speichern Informationen über Datenprodukte und machen diese im Data Mesh auffindbar. Dazu gehören technische Metadaten, Informationen über die Besitzer der Datenprodukte, Datenherkunftsinformationen und Informationen über die in den Daten beschriebenen Geschäftsobjekte. Datenprodukte lassen sich durch definierte Schnittstellen registrieren. Sind Datenmarktplätze, Datenprodukte und Identitätsmanagementsystem interoperabel, so kann der Datenmarktplatz die Rolle eines Zugriffssorchestrators einnehmen. Als üblicherweise zentralisierte Komponente eines Data Mesh folgen Datenkataloge und Datenmarktplätze dem Hub-and-Spoke-Muster, indem sie relevante Metadaten über Datenprodukte aggregieren und an potenzielle Datennutzende weitergeben.

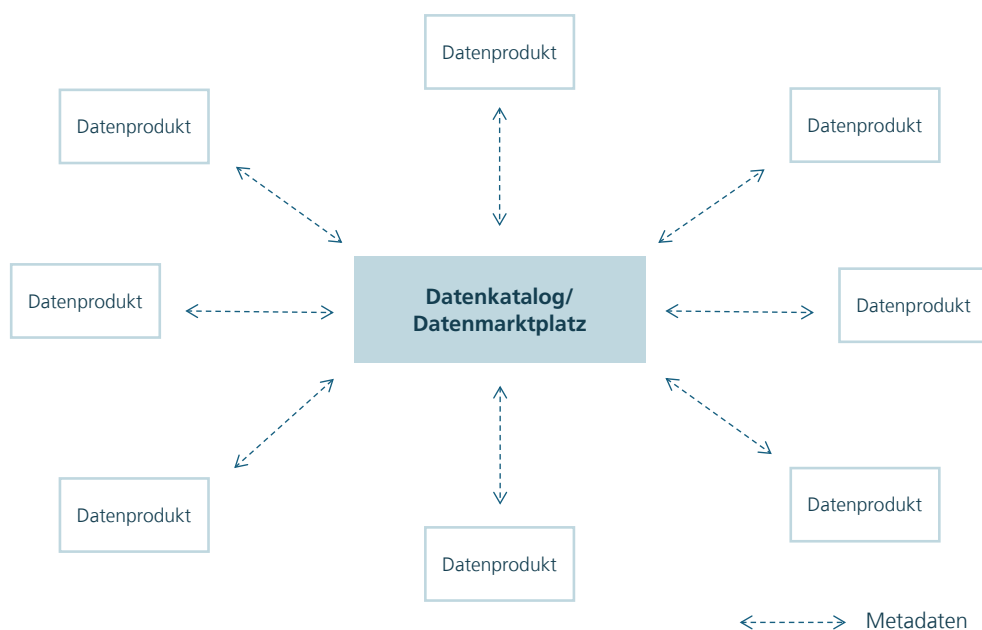


Abbildung 11: Schema eines Datenkatalogs und eines Datenmarktplatzes

3. Datenräume in der Schweiz – Befähiger zur Mehrfachnutzung von Daten

3.1. Das Datenökosystem Schweiz – Herausforderungen und Ziele

Die Mehrfachnutzung von Daten ermöglicht die Förderung umfassender Potenziale in Wirtschaft, Politik, Verwaltung, Wissenschaft und durch Individuen mit wirtschaftlichem und gesellschaftlichem Mehrwert. Beispielsweise können Hotelbuchungen, Wegkonditionen, Auslastungs- und Mobilitätsdaten zusammengeführt und zur Steuerung touristischer Ströme zum Schutz von Natur und Umwelt sowie zur Verbesserung der Touristenzufriedenheit eingesetzt werden. Dabei macht man sich zu Nutze, dass einmal erhobene Daten potenziell nahezu kostenfrei vervielfältigt werden können.

In der Schweiz werden die Potenziale geteilter Daten – wie auch in anderen europäischen Ländern – derzeit nur in wenigen Fällen gehoben. Dies ist auf verschiedenste rechtliche, organisatorische, semantische und technische Hindernisse zurückzuführen. Fehlendes Vertrauen und mangelnde Selbstbestimmung führen aufgrund unzureichender Transparenz und Kontrolle zu einer kritischen Einstellung gegenüber der Freigabe von Daten. Weiterhin ist die Qualität vorliegender Daten oft unzureichend, da der Reifegrad der beteiligten Organisationen zum Teilen von Daten zu gering ist. Entsprechend sind größere Aufwände bei der Vorbereitung der Daten für die Mehrfachnutzung notwendig. Es fehlt zudem an einer organisationsübergreifenden Semantik, was zu Verständnisproblemen bei der Datennutzung führt. Aus technischer Perspektive mangelt es an geeigneten, wiederverwendbaren Komponenten und an einer Einigung über gemeinsame Standards zur Mehrfachnutzung.

Diese Herausforderungen sollen mit der Entstehung (oftmals sektorspezifischer) Datenräume angegangen werden. Datenräume definieren Regeln und Standards für den Datenaustausch, und setzen diese in technischen und organisatorischen Strukturen um. Das Datenökosystem Schweiz zielt auf die Formung eines vertrauenswürdigen und interoperablen Rahmens für Datenräume ab, sodass Daten verschiedener Akteure für gesellschaftlichen und sozialen Wohlstand und zum wissenschaftlichen Fortschritt genutzt werden können. Dazu muss eine Reihe von Voraussetzungen geschaffen werden. Zu diesen zählen:

- Vertrauen, Transparenz und Selbstbestimmung einzelner Akteure,
- Interoperabilität unter anderem hinsichtlich rechtlicher, organisatorischer und technischer Kernaspekte – national und international anschlussfähig,
- ausreichende Datenmanagement-Kompetenzen der beteiligten Organisationen.

Das Datenökosystem Schweiz unterstützt diese Aspekte, indem es als vertrauenswürdige staatliche Instanz Transparenz gewährleistet, Fairness in der Datennutzung und Interoperabilität

fördert sowie eine Machtkonzentration bei einzelnen Akteuren verhindert. Darüber hinaus erleichtert es ein gemeinsames Verständnis von Daten und verringert den Aufwand für die Entwicklung und den Ausbau von Datenräumen. Zudem wird die Wiederverwendbarkeit von Komponenten aus Datenräumen gefördert, was die Effizienz und Zusammenarbeit weiter verbessert.

3.2. Datenräume als Befähiger der Datenmehrfachnutzung

Während Datenräume von Seiten der Wirtschaft, Politik und Forschung als wichtiger Lösungsansatz für die Realisierung der Mehrfachnutzung von Daten angesehen werden, ergibt sich hinsichtlich des konkreten Umfangs und deren Eigenschaften noch kein einheitliches Bild. Der Begriff »Datenraum« wurde vor etwa 20 Jahren als organisationsinternes Datenintegrationskonzept eingeführt (Franklin et al., 2005). Anders als herkömmliche Ansätze, die eine physische Zusammenführung von Daten in einer gemeinsamen Datenbank erfordern, erfolgt in Datenräumen die Datenintegration primär auf semantischer Ebene. Daten verbleiben dadurch bis zu ihrem Abruf an ihrem Ursprung und werden ausschließlich bei Bedarf integriert, was Flexibilität und die Vermeidung von Datenredundanzen fördert. Datenräume können zudem verschachtelt oder überlappend sein. Daten können somit in verschiedenen Datenräumen verfügbar sein und zwischen ihnen geteilt werden. Zur Applikation des Konzepts auf den interorganisationalen Kontext wurden im Rahmen der Arbeiten der International Data Spaces Association⁷ zudem die Elemente Datensouveränität und Vertrauen als weitere zentrale Anforderungen definiert. Im Kern ähnelte dabei die Idee eines Datenraums bereits der Vision eines Data Mesh, wenngleich zunächst vor allem technische Aspekte fokussiert wurden.

In der Geschäftswelt wird der Begriff Datenraum oft im Kontext der organisationsübergreifenden Zusammenarbeit mit Daten verstanden. Organisationen aus verschiedenen Branchen sehen Datenräume als ein Format zur Erreichung gemeinsamer Ziele durch die kooperative Nutzung von Daten.

Weiterhin rücken Datenräume und ihre zugrundeliegenden Software-Infrastrukturen vermehrt in den Fokus regulatorischer Anforderungen. Datenräume müssen Aspekte wie Vertrauen, Interoperabilität und Datenübertragbarkeit unterstützen, Datensouveränität gewährleisten und diskriminierungsfrei sein. In diesem Kontext fungieren Datenräume als Datenintermediäre, die beispielsweise unter die Maßgaben des EU Data Governance Act fallen (Otto, 2022). Zudem existieren vielfältige politische Initiativen zur Förderung von Datenräumen.

Das Data Spaces Support Centre (DSSC) als EU-geförderte Institution zur Konsolidierung und Förderung von Datenraum-Aktivitäten übernimmt dabei die Definition von CEN/CENELEC:

*[EN] Data space
interoperable framework, based on common governance principles, standards, practices and enabling services, that enables trusted data transactions between participants.*

*[DE] Datenraum
Interoperabler Rahmen, der auf Basis gemeinsamer Gouvernanz-Prinzipien, Standards, Praktiken und Diensten vertrauenswürdige Datentransaktionen zwischen Teilnehmenden ermöglicht. (CEN/CENELEC, S. 6)*

⁷ <https://internationaldataspaces.org/>

Kernelement eines Datenraums ist die Herstellung von Interoperabilität auf den Ebenen Technologie, Semantik, Organisation und Recht. Dazu legt ein Datenraum Gouvernanz-Prinzipien, Standards, Prozessmodelle und Datenraum-Dienste fest. Weiterhin wird Vertrauen als wichtiges Element in den Vordergrund gerückt. Vertrauen in die Datenraum-Dienste, die Datenraum-Teilnehmenden sowie schlussendlich die bereitgestellten Daten sind unerlässliche Voraussetzungen für den Datenaustausch.

In der Schweiz hat die Bundesverwaltung eine eigene Definition für den Begriff »Datenraum« formuliert, um einen einheitlichen Rahmen für regulatorische Aktivitäten zu setzen:

Ein Datenraum ist eine technische und organisatorische Struktur, welche Bereitstellung, Austausch und Bezug von Daten aus verschiedenen Quellen und von verschiedenen Akteuren ermöglicht und regelt. Oftmals sektorenspezifisch organisiert und durch Zweck, klare Regeln und Standards definiert. Wird von der Datenraumbetreiberin [Datenraumträgerschaft] verwaltet. (UVEK & EDA, 2022, S. 46)

Damit erhält die Datenraumträgerschaft eine zentrale Rolle. Sie soll als Hauptakteur die Organisation eines Datenraums übernehmen. Zudem wird eine sektorspezifische Gliederung von Datenräumen vorgesehen, die unter anderem spezifische technische Anforderungen und semantische Standards der jeweiligen Branchen berücksichtigt. Entgegen der ursprünglichen technischen Definition von Datenräumen wird eine dezentrale Architektur und Datenhaltung bei beiden Definitionen nicht explizit erwähnt. Entsprechend ergeben sich für die Stakeholder des Datenökosystems Schweiz und die entstehenden Datenräume größere Freiheiten hinsichtlich der Gestaltung ihrer Architekturen.

Bei der Umsetzung von Datenökosystemen ist zwischen zwei Arten von Akteuren zu unterscheiden: Dies sind die Akteure eines Datenraums und die Akteure eines Datenökosystems. Während Datenraum-Akteure technisch integriert sind, müssen Akteure des Datenökosystems nicht zwangsweise technisch integriert sein. Ein Beispiel hierfür sind natürliche Personen, die Mobilitätsdienste nutzen. Durch ihre Interaktion mit Mobilitätsanbietern, beispielsweise einem Verkehrsunternehmen, produzieren sie Daten und profitieren von datenbasierten Diensten. Die Personen stellen ihre Nutzungsdaten allerdings nicht direkt über die Infrastruktur eines Datenraums bereit, sondern geben diese ausschließlich an ihren Mobilitätsanbieter weiter, der diese wiederum in einen Mobilitätsdatenraum integriert. Für den Mobilitätsanbieter ergibt sich andersherum durch die Teilnahme an Datenräumen eine neuartige Grundlage zur Schaffung digitaler Serviceangebote in Datenökosystemen (Möller et al., 2024).

3.2.1. Bestandteile von Datenräumen

Konkreten Aufschluss über die Bestandteile von Datenräumen liefert das DSSC im Rahmen des Data Spaces Blueprint (DSSC, 2024)⁸. Der Data Spaces Blueprint liefert eine Übersicht zu Datenraumbausteinen. Ein Datenraumbaustein ist eine grundlegende und in sich geschlossene Komponente, die innerhalb eines Datenraums umgesetzt und mit anderen Komponenten kombiniert wird, um die Funktionalität eines Datenraums zu gewährleisten. Die konkrete Umsetzung dieser Bausteine hängt von den individuellen Zielen und Umständen des umzusetzenden Datenraums ab. Eine Übersicht der Bausteine ist in Abbildung 12 gezeigt. Eine detaillierte Übersicht über die Datenraumbausteine im Kontext des Datenökosystems Schweiz liefert eine Studie der Bundeskanzlei – Digitale Transformation und IKT Lenkung (2025).

⁸ Grundlage dieser Publikation ist die Version 1.5 des DSSC Blueprints. Künftige Versionen des Blueprints können von diesem Stand abweichen.



Abbildung 12: Datenraumbausteine gemäß Bundeskanzlei – Digitale Transformation und IKT Lenkung (2025)

Das DSSC unterscheidet zwischen den geschäftlichen und organisatorischen Bausteinen sowie den technischen Bausteinen. Die geschäftlichen und organisatorischen Bausteine beziehen sich auf die Werterzeugung innerhalb von Datenräumen, deren Organisationsstruktur und rechtlich konforme Gestaltung. Die technischen Bausteine beziehen sich auf die technischen Fähigkeiten, die zur Umsetzung von Datenräumen benötigt werden. Die geschäftlichen und organisatorischen Bausteine unterteilen sich in Bausteine zur Charakterisierung des Wertversprechens eines Datenraums (Geschäftsbetrieb), zur Definition von Struktur, Entscheidungsprozessen und Beteiligungsformen (Gouvernanz) und zur Etablierung von rechtlicher und vertraglicher Compliance (Rechtsrahmen). Die technischen Bausteine gliedern sich nach den übergeordneten Datenraumfähigkeiten, die zur Mehrfachnutzung von Daten erforderlich sind. Dazu gehören die Herstellung von Dateninteroperabilität, die Erzeugung von Vertrauen und die Gewährleistung digitaler Selbstbestimmung sowie die Ermöglichung von Datenwertschöpfung. Diese Fähigkeiten werden mithilfe technischer Lösungen und Standards realisiert.

4. Von Data Mesh lernen, um Datenräume zu fördern

Auf Basis der vorangegangenen Beschreibungen der Konzepte von Data Mesh und Datenräumen werden verschiedene Gemeinsamkeiten der beiden Konzepte sichtbar, die an dieser Stelle noch einmal hervorgehoben werden: Beide Konzepte fokussieren auf das Teilen von Daten in komplexen Umgebungen, bei denen den einzelnen Teilnehmenden oder Domänen weitreichende Autonomie hinsichtlich ihrer internen Organisation und Infrastruktur gegeben wird. Beide Ansätze verfolgen dabei einen ganzheitlichen soziotechnischen Systemansatz. In diesem werden sowohl die bereitgestellte technische Struktur, die Organisation und Gouvernanz sowie kulturelle Aspekte als Grundpfeiler zur Ermöglichung der Mehrfachnutzung entwickelt. Obwohl es einige Unterschiede zwischen den Anforderungen und Grundprinzipien der beiden Konzepte gibt – Datenräume müssen aufgrund ihres Fokus auf das interorganisationale Teilen von Daten insbesondere stärker Datensouveränitätsaspekte und rechtliche Vorgaben hinsichtlich der Mehrfachnutzung von Daten fokussieren – lassen sich Ansätze aus dem Bereich Data Mesh umfassend und gewinnbringend auf den Bereich Datenräume übertragen.

Dazu gehören insbesondere technische und organisatorische Mechanismen zur Realisierung der Prinzipien »Data as a Product«, »Federated computational Governance« und »Self-serve Infrastructure«. Das Prinzip »Domain-driven Design« kann nur in geringen Umfängen auf die Datenraumentwicklung appliziert werden, da die Teilnehmenden in Datenräumen üblicherweise von sich aus eigenständige Organisationen sind, auf deren interne Struktur die Gestaltung des Datenraums keinen direkten Einfluss nehmen kann. Die Implikationen einer Anwendung der zuvor genannten drei Prinzipien auf Datenräume wird im Folgenden konkretisiert.

4.1. Behandlung von Daten »as a Product« zur Ermöglichung von Generativität und Mehrfachnutzung

Im Kontext des modernen Datenmanagements gewinnen die Begriffe »Data Product« und »Data as a Product« an Bedeutung. In Datenräumen liegt der Fokus stark auf der Bereitstellung von Datenprodukten als spezifisches in sich abgeschlossenes Resultat eines Datenverarbeitungsprozesses, das auf bestimmte Anwendungsfälle zugeschnitten ist. Data Mesh-Konzepte gehen darüber hinaus und fördern durch die Behandlung von Daten als Produkt die umfassende Verwaltung von Daten entlang ihres Lebenszyklus unter besonderer Berücksichtigung der Anforderungen von potenziellen Datennutzenden.

Ein Datenprodukt ist ein verwaltendes Artefakt, das wiederkehrende Informationsbedürfnisse befriedigt und durch die Umwandlung und Aufbereitung relevanter Datenelemente in eine konsumierbare Form einen Mehrwert schafft (Hasan & Legner, 2023). Datenprodukte beziehen sich auf auffindbare, verständliche, qualitativ hochwertige, gebrauchsfertige und wiederverwendbare Datenbestände, die Menschen auf verschiedene geschäftliche Herausforderungen anwenden können (Bode et al., 2023; Hackl & Winter, 2023). Datenprodukte fokussieren sich dabei insbesondere auf die technische Bereitstellung von Funktionen, wie beispielsweise deren Entdeckbarkeit und Analysierbarkeit durch Datennutzende mittels Metadaten sowie deren einfache Abrufbarkeit. Für die Datenbereitstellenden ergeben sich Vorteile durch die Analysierbarkeit der tatsächlichen Nutzung ihrer Daten und der daraus folgenden Möglichkeit, die Wertbemessung der Daten zu fördern. Ein wichtiges Konzept zur Realisierung von Datenprodukten stellen sogenannte Datenverträge dar. Dabei handelt es sich nicht um Verträge im rechtlichen Sinne. Vielmehr beschreiben Datenverträge die Erwartungen an ein Datenprodukt, beispielsweise hinsichtlich der Datenqualität, des Personenbezugs, der Aktualität oder der Lizenzen, die die Datenbereitstellenden gedenken zu erfüllen. Mit einem Datenvertrag verspricht der Anbietende dem potenziellen Nutzenden ein bestimmtes Leistungsniveau. Datenprodukte werden üblicherweise zu deren Auffindbarkeit und Vergleichbarkeit in Datenkatalogen registriert.

In Datenräumen werden Datenprodukte üblicherweise auf Basis zuvor definierter Anwendungsfälle und deren Datenanforderungen erzeugt (Top-Down-Prinzip). Die Generativität der Daten zur Nutzung in weiteren Anwendungen und für bislang nicht vorgesehene Anwendungsfälle wird dabei eingeschränkt. Insbesondere bei Daten öffentlicher Einrichtungen, beispielsweise im Bereich Smart Cities, wird dieser Ansatz der Erfüllung vorgegebener Checklisten nicht mehr ausreichen, um Datenwertschöpfung zu fördern und eine echte Mehrwertnutzung von Daten zu gewährleisten (Krystek, 2023). Daten müssen so bereitgestellt werden, dass diese eine möglichst hohe Generativität erzielen und in vielfältigen Anwendungen verwendet werden können. In dieser Hinsicht bieten die Verfahrensweisen des Managements von Daten als Produkt aus dem Data Mesh wertvolle Best Practices.

Daten als Produkt ist ein Ansatz in der Datenverwaltung und -analyse, bei dem Datensätze als eigenständige Produkte behandelt werden, die mit Blick auf die Endnutzenden entworfen, erstellt und gepflegt werden (Mucci & Stryker, 2024). Entlang des Datenlebenszyklus werden dabei Aspekte des Produktmanagements wie etwa die Entwicklung von Produktstrategien, die Zusammenarbeit mit verschiedenen Teams oder die Kommunikation mit Stakeholdern zur Erfassung von Anforderungen und Feedback auf die Entwicklung von Datenprodukten angewendet. Somit werden potenzielle Nutzungsmöglichkeiten der Daten antizipiert und eine einfache Verwendbarkeit durch potenzielle Datennutzende gewährleistet. Insbesondere ermöglicht das Management von Daten als Produkt auch eine Fokussierung der Aufwände in der Datenverwaltung auf besonders nachgefragte Datenprodukte. Die Behandlung von Daten als Produkt geht somit über die einfache technische Bereitstellung von Datenprodukten basierend auf vorgegebenen Kriterien hinaus.

Ein grundlegendes Prinzip des datenproduktbasierten Ansatzes ist ein klar definierter Produktionsprozess, der alle relevanten Beteiligten (insbesondere Datensubjekte und Datennutzende) mit einbezieht (Hasan & Legner, 2024). Die Datenbereitstellenden müssen die wesentlichen geschäftlichen Anforderungen identifizieren und analysieren, welche Daten in welcher Qualität und zu welchen Zeitpunkten bereitzustellen sind. Darauf basierend muss er die relevanten internen Daten identifizieren und das passende Datenprodukt bestehend aus Daten und Metadaten entwickeln und testen. Funktionen der Datenproduktentwicklung sind dabei möglichst zu automatisieren. Erwartungen an das Datenprodukt wie beispielsweise die Semantik, Qualität und Nutzungsbedingungen werden in maschinenlesbaren Datenverträgen dokumentiert, die von den genutzten technischen Komponenten automatisch analysiert und durchgesetzt werden können. Datennutzende erhalten somit die Möglichkeit, das Datenprodukt zu analysieren, dieses

abzurufen und im Rahmen der festgelegten Nutzungsbedingungen zu verwenden. Die Datenbereitstellenden erhalten Transparenz in die Datenverwendung und sind in der Lage, durch das Feedback der Datennutzenden das Datenprodukt kontinuierlich zu verbessern.

Eine wichtige Rolle kommt dabei den Datenproduktverantwortlichen zu. Sie fungieren als maßgebliche Stimme der Organisation und wägen die Bedürfnisse, Anliegen und Rückmeldungen der Datennutzenden gegenüber denen der Datenbereitstellenden ab. Sie sorgen zudem für den Austausch wichtiger Kenntnisse über Daten zwischen den beteiligten Parteien. Dieser Informationsfluss ist entscheidend, um Anforderungen richtig zu verstehen, Ressourcen sinnvoll zuzuweisen und Datenprodukte im Einklang mit den Zielen der Datennutzenden und den Anforderungen an Datensouveränität der Datenbereitstellenden zu gestalten. Dabei sollten die Datenproduktverantwortlichen ebenso entscheiden, ob ein Datenprodukt frei oder gegen monetäre Gegenleistung verfügbar gemacht werden soll (Krystek, 2023).

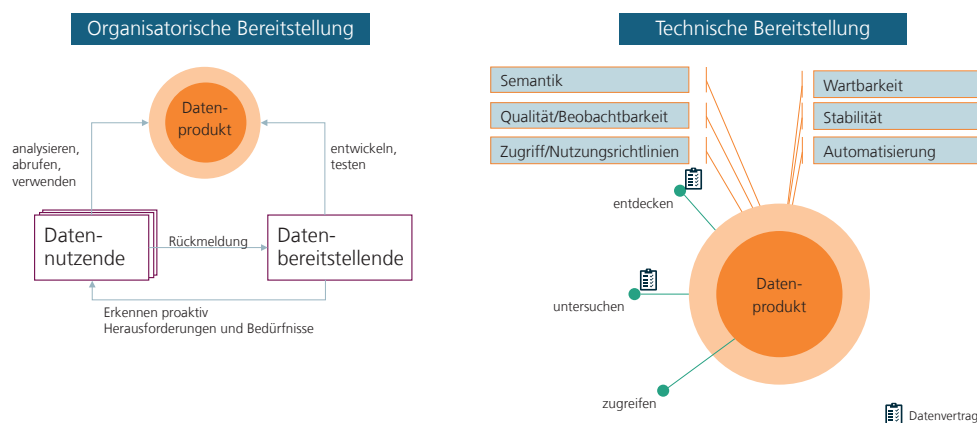


Abbildung 13: Management von Daten als Produkt aus Sicht der Datenbereitstellenden in Datenräumen

Wenngleich die Behandlung von Daten als Produkt an erster Stelle durch die Datenbereitstellenden selbst, beispielsweise durch die Ausrichtung interner Strukturen, Prozesse und Verhaltensweisen, umgesetzt werden muss, können gewisse Gestaltungsaspekte von Datenräumen ebenso die Behandlung von Daten als Produkt incentivieren. Dazu gehören beispielsweise:

- die gemeinsame Festlegung verpflichtender Beschreibungsattribute von Datenprodukten sowie die Definition gemeinsamer KPIs,
- die Bereitstellung von Mehrwertdiensten zur einfachen Beschreibung von Datenqualität und Ermöglichung von Datenbeobachtbarkeit durch die Teilnehmenden,
- die Schaffung von Services, die Transparenz der Datennutzung ermöglichen,
- die Schaffung einer offenen Kultur des Informationsaustausches zwischen Beteiligten und eines gemeinsamen Verständnisses über die grundlegenden Herausforderungen der Datenraumteilnehmenden,
- die Schaffung einer Plattform für Austausch und Kommunikation zwischen den Stakeholdergruppen,
- die Kommunikation erfolgreich umgesetzter Anwendungen als gemeinsame Leistung der beitragenden Parteien.

Der Übergang zu einem Ansatz, bei dem Daten mit der gleichen Sorgfalt und dem gleichen strategischen Denken behandelt werden wie traditionelle Produkte, eröffnet die Möglichkeit, das Innovationspotenzial von Datenräumen zu maximieren und gleichzeitig die Zusammenarbeit zwischen den Teilnehmenden zu fördern. Diese Transformation erfordert ein Umdenken hinsichtlich des Teilens von Daten – weg von isolierten Datensilos hin zu einer Kultur, die den gemeinsamen Nutzen und die Wertschöpfung aus Daten in den Mittelpunkt stellt.

4.2. Institutionalisierte Bereitstellung der dezentralisierten technischen Dateninfrastruktur

Die technische Dateninfrastruktur stellt die fundamentale Grundlage zur Ermöglichung von Interoperabilität und Datenaustausch und damit für die Mehrfachnutzung von Daten dar (Estermann et al., 2018). Sie ist zudem ein wichtiger Anreizmechanismus für den Austausch von Daten und die Teilnahme an Datenräumen, da sie unter anderem die Datensicherheit, die Datenkontrolle und das Vertrauen zwischen den beteiligten Akteuren beeinflusst (Gelhaar et al., 2021). Bestandteile der technischen Dateninfrastruktur sind beispielsweise das Identitätsmanagement-System, das Teilnehmendenverzeichnis, der Datenkatalog sowie weitere Mehrwertdienste. Die Dateninfrastruktur eines Datenraums abstrahiert, analog zur Self-Service Dateninfrastruktur im Data Mesh, die individuellen Dateninfrastrukturen der Unternehmen und erleichtert die Erstellung von Datenprodukten. Die Anforderungen an die gemeinsam genutzte Dateninfrastruktur sind vielfältig. Sie umfassen unter anderem Vertrauenswürdigkeit, Nutzungsfreundlichkeit, Kosteneffizienz und Skalierbarkeit sowie die Wahrung der digitalen Selbstbestimmung der Teilnehmenden. Daraus entstehen üblicherweise vielfältige Zielkonflikte, die es im Rahmen der Entwurfs- und Implementierungsphasen eines Datenraums zu lösen gilt. Im Gestaltungsprozess sind insbesondere die Voraussetzungen der potenziellen Datenraumteilnehmenden (kleine und mittelständische Unternehmen, Großunternehmen, öffentliche Institutionen) zu berücksichtigen, um faire Kollaborations- und Wettbewerbsbedingungen im Umgang mit Daten zu schaffen.

Während in Datenräumen generell verschiedene Optionen zum Betrieb der Dateninfrastruktur möglich sind, werden diese im Data Mesh von einem dedizierten Plattform-Team zur Verfügung gestellt. Die bereitgestellten Dienste lassen sich durch die Domänenteams einfach im Rahmen eines Self-Service nutzen. Für die Domänen ergibt sich dadurch eine erhebliche Komplexitätsreduktion. Diese Form der Bereitstellung wird ebenso in verschiedenen Initiativen des organisationsübergreifenden Datenaustausches erprobt (Falconi & Plebani, 2023; Krystek, 2023).

In Anbetracht dieser Erkenntnis sollten Datenräume ein oder mehrere dedizierte Unternehmen mit der Entwicklung und Bereitstellung der Infrastruktur für die Teilnehmenden betrauen. Die Infrastrukturbetreibenden können dabei unabhängig von der Datenraumträgerschaft sein und entweder durch diese oder durch die einzelnen Datenraumteilnehmenden beauftragt werden. Betrieb und Architektur der Dienste sind hierbei voneinander unabhängig. Eine dezentrale Architektur kann nach wie vor von einem oder mehreren dedizierten Betreibern unterstützt werden. Die Beauftragung von dedizierten Organisationen hat dabei unter anderem folgende Vorteile:

- Datenbereitstellende und Datennutzende können sich auf ihre Kernaufgaben konzentrieren und benötigen keine umfassenden Kenntnisse im Management und im Betrieb der Dateninfrastrukturdienste.

- Effizienzen bei der Bereitstellung von Dateninfrastrukturen werden gehoben, indem die Leistung gleichartiger Tätigkeiten durch die Teilnehmenden reduziert wird.
- Gemeinsam genutzte Infrastrukturdienste erhöhen die Vereinheitlichung der Datenprodukte und verringern mögliche Interoperabilitätsprobleme.
- Eine Weiterentwicklung der Dateninfrastruktur erfolgt auf Basis gemeinsam definierter Anforderungen und nicht auf Basis individueller Interessen.

Bei der konkreten Umsetzung der Self-Service Dateninfrastruktur sollten einige Berücksichtigungen getroffen werden. Die Betreiberorganisation muss eine neutrale und vertrauenswürdige Partei sein, die keine Eigeninteressen bei der Verwertung von Daten verfolgt. Die Betreibenden sollten an einer gemeinsamen Weiterentwicklung der Dienste interessiert sein und nicht nur bestehende Services verkaufen wollen, um eine kontinuierliche Weiterentwicklung und Expansion des Datenraums zu ermöglichen. Die Weiterentwicklung der Dateninfrastruktur sollte dabei möglichst unabhängig von einzelnen Anwendungsfällen erfolgen und Generativität ermöglichen. Die bereitgestellte Dateninfrastruktur sollte die zugrundeliegende Komplexität verbergen und deren Services sollten möglichst einfach über wohl definierte Schnittstellen nutzbar sein. Wiederkehrende Aufgaben sollten zudem automatisiert werden. Weiterhin ist die Autonomie einzelner Teilnehmenden durch die Verringerung zentraler Instanzen zu wahren, um mögliche Bottlenecks zu reduzieren, und den Teilnehmenden den notwendigen Entscheidungsspielraum zu gewähren.

4.3. Föderierte Gouvernanz in Datenräumen

Zur Etablierung von Datenräumen sind vielfältige Gouvernanzentscheidungen zu treffen. Dazu gehören insbesondere die Festlegung der Organisationsstruktur sowie von Regeln, Richtlinien und Standards zum Umgang mit Daten entlang ihres Lebenszyklus. Im interorganisationalen Bereich sind insbesondere Entscheidungen hinsichtlich der in Tabelle 3 beschriebenen Aspekte notwendig.

Tabelle 3: Organisationsübergreifende Aspekte der Datengouvernanz nach Ribeiro et al. (2024)

Aspekt der Datengouvernanz	Beschreibung
Organisationsübergreifende Datensicherheit	Definition sicherer Mechanismen für den Zugriff auf Daten und die Gewährleistung von Transparenz bei datenbezogenen Vorgängen.
Organisationsübergreifende Datenarchitektur	Definition von Mechanismen zur Förderung der Datenintegration von mehreren Partnern im Rahmen eines Architekturmodells.
Organisationsübergreifende Metadaten	Definition organisationsübergreifender Metadatenstandards und Entwicklung von Verfahren für die dezentrale Metadatenkuration.

Aspekt der Datengouvernanz	Beschreibung
Organisationsübergreifendes Dateneigentum und Vertrauen	Modelle zur Definition von Dateneigentum und zur Gewährleistung von Vertrauen in Daten, Dienste und Teilnehmende.
Organisationsübergreifender Datenlebenszyklus	Entwicklung von gemeinsamen Datenaustauschstandards.
Organisationsübergreifender Datenschutz	Prüfung gemeinsam genutzter Datenbestände auf personenbezogene Informationen und Gewährleistung der Einhaltung der für die verschiedenen Partner geltenden Regeln und Richtlinien.

Um die beschriebenen Entscheidungen zu treffen, bedarf es der Festlegung einer effizienten Entscheidungsstruktur. Zur Berücksichtigung der Anforderungen und Fähigkeiten einzelner Domänen wird in Data Mesh ein föderiertes Gouvernanz-Modell genutzt. Im Rahmen der föderierten Gouvernanz kommen Vertretende der einzelnen Datendomänen zusammen, um grundlegende Entscheidungen über die mindestens notwendigen Mechanismen zur Interoperabilität der einzelnen Domänen und zur Hebung gemeinsamer Effizienzen zu treffen, die von der zugrundeliegenden Dateninfrastruktur technisch umgesetzt werden müssen. Gleichzeitig wird den einzelnen Domänenteams ein ausreichender Freiraum gelassen, um eine schnelle Datenproduktentwicklung zu gewährleisten (Dolhopolov et al., 2024). Einzelne Domänen können dabei nur Teil des Data Mesh werden, sofern sie die definierten Regeln und Richtlinien erfüllen.

Ein föderiertes Modell der Datengouvernanz ist zur Gewährleistung der Autonomie einzelner Organisationen auch im Bereich von Datenräumen sinnvoll. Abbildung 14 visualisiert die Ebenen föderierter Gouvernanz im Kontext des Datenökosystem Schweiz. Im Datenökosystem Schweiz ergeben sich drei Ebenen föderierter Gouvernanz. Auf datenraumübergreifender Ebene, hier Data Spaces Mesh genannt, ist ein minimaler Satz gemeinsamer Regeln, Richtlinien und Technologien festzulegen, um die Interoperabilität zwischen einzelnen Datenräumen grundlegend zu gewährleisten. Diese Grundsatzentscheidungen sollten gemeinsam durch die Datenraumträgerschaften getroffen und weiterentwickelt werden. Die Anlaufstelle Datenökosystem Schweiz kann hier als Orchestrator dienen. Auf Ebene eines individuellen Datenraums sollten darauf aufbauende (domänenspezifische) Regeln, Richtlinien, Standards und Technologien festgelegt werden, um eine Zusammenarbeit zwischen den einzelnen Teilnehmenden im Kontext der bereitgestellten Daten zu ermöglichen. Diese Aspekte sollten gemeinsam von den teilnehmenden Organisationen definiert werden, um den Bedürfnissen der Nutzenden gerecht zu werden. Diese sind zudem teilweise selbst von den Datenraumteilnehmenden umzusetzen, um eine Teilnahme zu berechtigen. Einzelne Akteure sind weiterhin in der Lage zu bestimmen, wie sie ihre interne Infrastruktur gestalten, ob sie an einem Datenraum teilnehmen, welche Datenprodukte sie zur Verfügung stellen und wer unter welchen Umständen auf die Daten zugreifen kann.

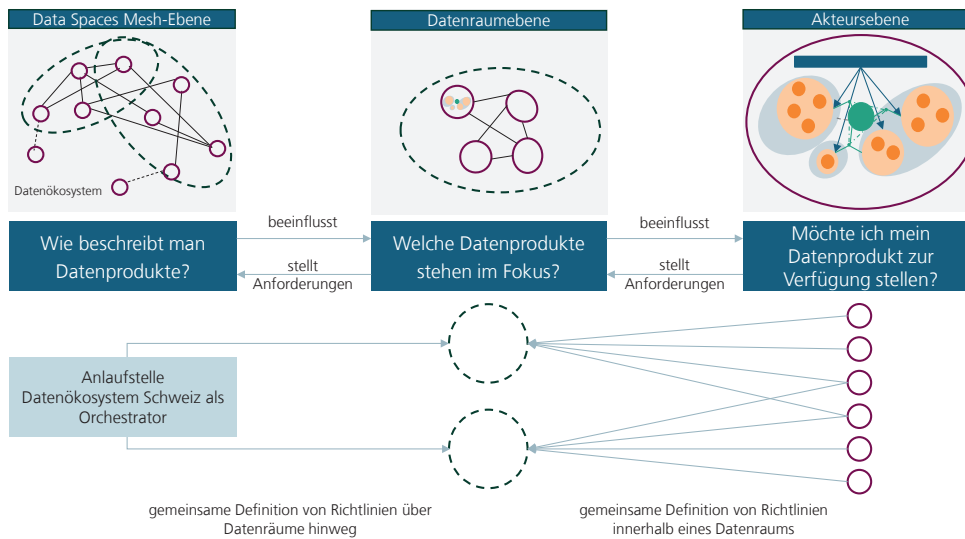


Abbildung 14: Föderierte Datengouvernanz in Datenräumen

4.4. Herausforderung: Erfüllung der Anforderungen unterschiedlicher Anspruchsgruppen in der Umsetzung

Analog zur Umsetzung von Data Mesh existieren auch zur Umsetzung von Datenräumen verschiedene Topologien aus dem Spektrum von zentralisiert zu dezentralisiert. Je nach Topologie ergeben sich verschiedene Erfüllungsgrade, die einen Einfluss auf die für die Teilnehmenden eines Datenraums relevanten Entscheidungsfaktoren besitzen⁹. Zu den Erwartungen der Teilnehmenden an einen Datenraum gehören:

- **Souveränität:** Möglichkeit der Datenbereitstellenden, selbstbestimmt über Datenspeicherung und Datennutzung zu entscheiden.
- **Resilienz:** Fähigkeit des Datenraums, trotz unerwarteter Probleme funktionsfähig zu bleiben.
- **Skalierbarkeit:** Fähigkeit, die Anzahl der Teilnehmenden und Datenprodukte eines Datenraums zu erhöhen, ohne an Funktionalität einzubüßen.
- **Kontrolle:** Möglichkeit der Datenraumträgerschaft, die Teilnehmenden, angebotenen Datenprodukte und deren Nutzung zu kontrollieren.
- **Einfachheit:** Einfache Nutzung des Datenraums für Teilnehmende.
- **Auffindbarkeit:** Effizienz der Schritte zur Identifizierung von im Datenraum angebotenen Daten.

⁹ Weitere Informationen liefert das IDSA Rulebook: https://github.com/International-Data-Spaces-Association/IDSA-Rulebook/blob/main/documentation/3_Functional_Requirements.md

Abbildung 15 stellt den Erfüllungsgrad der genannten Erwartungen durch verschiedene Datenraumtopologien in schematischer Form dar. Auch bei der Umsetzung von Datenräumen muss eine Abwägung zwischen verschiedenen Entscheidungsfaktoren getroffen werden, um eine informierte Entscheidung hinsichtlich der genutzten allgemeinen Topologie beziehungsweise der Auslegung einzelner Komponenten zu treffen. Dabei sind insbesondere die unterschiedlichen Bedürfnisse der Teilnehmenden zu berücksichtigen. Zentraler Betrachtungsgegenstand sollte dabei die Dichotomie zwischen Datenbereitstellenden und Datennutzenden sein, da beide Gruppen unterschiedliche Erwartungen und Anforderungen an einen Datenraum stellen.

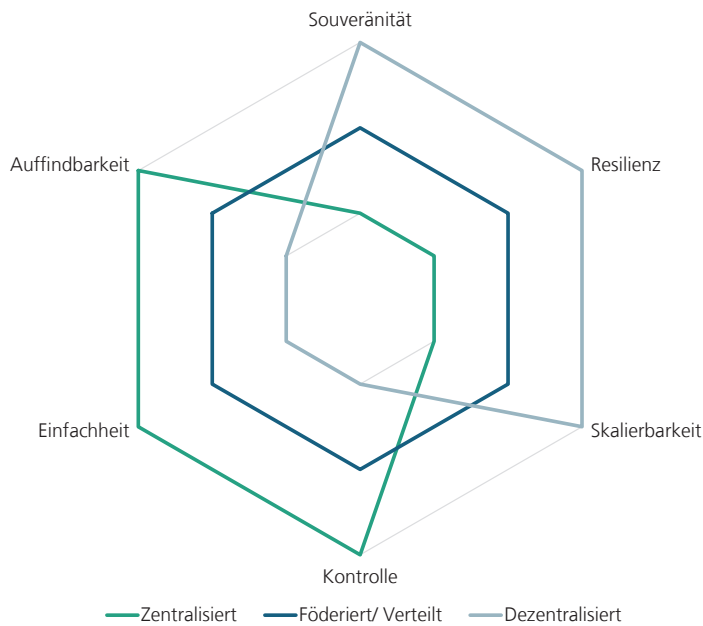


Abbildung 15: Topologien und deren Erfüllung von Kernerwartungen in Datenräumen gemäß IDSA

Datenbereitstellende legen Wert auf Aspekte wie Souveränität, Kontrolle über die Daten und Datenauffindbarkeit. Sie benötigen Mechanismen, die sicherstellen, dass ihre Daten geschützt sind und nur von autorisierten Akteuren genutzt werden können. Zudem ist die Transparenz über die Nutzung der bereitgestellten Daten für Bereitstellende von großer Bedeutung, um Vertrauen in den Datenaustausch zu erzeugen. Auf der anderen Seite stehen die Datennutzenden, die in der Regel an einem einfachen und schnellen Zugang zu qualitativ hochwertigen Daten interessiert sind. Sie erwarten, dass die Daten einfach zu finden, zu integrieren und zu verwenden sind. Werden die geteilten Daten zu einem zentralen Bestandteil des Wertversprechens der Datennutzenden, ergeben sich auch erhöhte Anforderungen an Resilienz und Skalierbarkeit der Datenraumdienste.

Die Einbindung beider Gruppen ist zentral, um eine nachhaltige Etablierung des Datenraums beziehungsweise des darauf basierenden Datenökosystems zu gewährleisten. Der Wert eines Datenökosystems steigt mit der Anzahl seiner Teilnehmenden (Netzwerkeffekte). Für Datenbereitstellende bedeutet ein größeres Netzwerk eine potenziell größere Menge von Datennutzenden. Für die Datennutzenden ergibt sich ein Zugang zu einer größeren Vielfalt an Datenquellen. Für die Trägerschaft eines Datenraums ergeben sich Skaleneffekte bei der Leistungserbringung. Steigt die Anzahl der Datentransaktionen in einem Datenraum, so können die Kosten pro Transaktion sinken, was sowohl den Datenbereitstellenden als auch den Datennutzenden durch niedrigere Transaktionsgebühren oder verbesserte Services zugutekommen kann.

4.5. Praxisbeispiel: Data Mesh-Ansätze im Schweizer Datenraum Gesundheit

Eine Veranschaulichung und Evaluation der zuvor erläuterten Data Mesh-Ansätze für Datenräume wird folgend anhand eines konzeptuellen Anwendungsbeispiels des Schweizer Datenraums Gesundheit vorgenommen. Mit dem Datenraum Gesundheit sollen allen Patientinnen, Patienten und Gesundheitsfachpersonen jederzeit verlässliche, detaillierte und vernetzte Gesundheitsdaten zur Verfügung stehen. Dabei stehen die Kerninteressen der Stakeholder wie Datenschutz, Datensicherheit und Interoperabilität im Vordergrund. Der Schweizer Gesundheitsdatenraum¹⁰ ist wie folgt definiert:

*Der **Datenraum Gesundheit** umfasst die Daten des Sektors Gesundheit (Behandlung, Abrechnung, Behörden, Forschung) und den rechtlichen, organisatorischen und technischen Rahmen. Der Rahmen regelt Aufbau, Betrieb und Nutzung des Datenraumes sowie die Gewährleistung von Sicherheit und Vertrauenswürdigkeit. Der Datenraum Gesundheit ist Bestandteil des Datenökosystems der Schweiz, national und international interoperabel.*

Eine potenzielle Anwendung des Datenraums Gesundheit stellt die Bereitstellung von Daten für die Krebsregister dar. Die Krebsregister in der Schweiz erfassen alle Fälle von Krebs und seinen Vorstufen sowie von bestimmten gutartigen Tumoren bei Personen mit Wohnsitz in der Schweiz, um Wissen über Tumorerkrankungen zu generieren. Die Daten werden dazu beispielsweise von den Spitälern an die kantonalen Krebsregister übersendet. Anschließend laufen die Daten bei der nationalen Krebsregistrierungsstelle zusammen, welche die Analyse der Daten vornimmt.

Im Rahmen der aktuellen Auslegung des Krebsregisters ergeben sich einige Herausforderungen. So sind Daten zum Austausch über alle beteiligten Akteure des Gesundheitswesens zu standardisieren. Zudem muss ein gemeinsamer Ende-zu-Ende Prozess über alle Akteure festgelegt werden, sodass ein einheitlicher Gesamtprozess entsteht. Weiterhin ergeben sich kulturelle Barrieren bei der Erhebung und Weitergabe der Daten. Beispielsweise erfolgt die Erfassung von Daten als manueller Prozess basierend auf (digitalen) Formularen und nicht im Hinblick auf die tatsächlich für die Analyse benötigten Daten. Sind zusätzlich relevante und über den Mindeststandard hinausgehende Daten vorhanden, werden diese oftmals nicht zur Mehrfachnutzung freigegeben, sondern verbleiben beim jeweiligen Akteur.

Für die Umsetzung des Anwendungsfalls Krebsregister im Schweizer Gesundheitsdatenraum ergeben sich darauf basierend folgende Zielstellungen: Die Erfassung und Bereitstellung der notwendigen Daten sollte auf Akteurebene als vollautomatisierter Prozess umgesetzt werden, der keine manuellen Eingriffe mehr erfordert. Vom Gesundheitsdatenraum wird erwartet, dass er die Transformation von einem fallbasierten hinzu einem patientenorientierten Denken durch die Verfügbarmachung und Zusammenführung von Daten unterstützt. Der Abruf von Gesundheitsdaten soll den Akteuren im Sinne eines Self-Service ermöglicht werden und somit auf vordefinierten Richtlinien und Akteurscharakteristika basieren anstatt auf langwierigen manuellen Verhandlungen.

¹⁰ Zum Zeitpunkt der Veröffentlichung dieses Dokuments befindet sich der Datenraum Gesundheit in der Aufbauphase. Die in dieser Diskussion herausgestellten Inhalte bewegen sich daher auf konzeptueller Basis und spiegeln nicht zwangsläufig die tatsächliche Implementierung des Datenraums Gesundheit wider.

Durch die Zusammenarbeit von Datenraumbetreibenden und Datengebenden sind dazu das Dateneigentum klar festzulegen und Datenprodukte auf technischer und organisatorischer Ebene zu definieren.

Behandlung von Daten »as a Product«

Die Anwendung eines Datenprodukt Denkens im Anwendungsfall Krebsregister impliziert die Entwicklung mehrerer Typen von Datenprodukten. Um den Anwendungsfall zu erfüllen, wird ein Eintrag in das Krebsregister in Form einer klinischen Krebsdokumentation einer Patientin oder eines Patienten als Datenprodukt benötigt. Dieses Datenprodukt orientiert sich an den Bedarfen der Datennutzenden wie beispielsweise den Kantonen und wird als nutzungsorientiertes Datenprodukt realisiert. Zu dessen Bereitstellung sind mehrere quellenorientierte Datenprodukte notwendig. Hierzu zählen beispielsweise Diagnosen und Medikamentengaben. Diese sind von den Spitälern bereitzustellen. Um die Stabilität, Qualität und ordnungsgemäße Nutzung der Datenprodukte sicherzustellen, werden (maschinenlesbare) Datenverträge zwischen Datenbereitstellenden und Datennutzenden realisiert. Für die datennutzenden Kantone ergibt sich die Notwendigkeit, die Datenverträge mit den Datenbereitstellenden so zu gestalten, dass möglichst geringe Aufwände zur Datenintegration von Nöten sind. Datenbereitstellende möchten üblicherweise ihrerseits geringe Aufwände zur Datentransformation aufbringen und ihre Datenangebote flexibel gestalten und erweitern. Durch die Verwendung von Produktmanagementprinzipien können die quellenorientierten Datenprodukte kontinuierlich weiterentwickelt werden – auch um andere Anwendungsfälle zu stützen und somit die Mehrfachnutzung von Daten weiter zu intensivieren. Im Falle des Krebsregisters muss jedoch angemerkt werden, dass die Flexibilität aufgrund der regulatorischen Pflicht zur Datenbereitstellung in einem gewissen Maße eingeschränkt wird. Im Sinne der Datenproduktentwicklung sollte ebenso eine kontinuierliche Einschätzung des Datenprodukterfolgs und die Möglichkeit zur Lieferung von Feedback durch die Datenproduktbereitstellenden eingerichtet werden. Beispielsweise könnten zusätzliche Akteure, wie die Forschung, weitere Bedarfe an die Daten anmelden.

Bereitstellung der Dateninfrastruktur durch Datenraumträgerschaft

Um die Bereitstellung von Datenprodukten, die Mehrfachnutzung von Daten und letztendlich die Realisierung von gewinnbringenden Anwendungen im Gesundheitsdatenraum zu ermöglichen, sind verschiedene Dateninfrastrukturdienste erforderlich. Dazu gehören Basisdienste wie ein Identitätsmanagement oder ein Datenproduktkatalog. Um das Datenproduktmanagement zu ermöglichen, benötigt es zudem dedizierte Unterstützungsdienste. Hierzu gehören beispielsweise eine Feedback-Plattform zur Weiterentwicklung der Datenprodukte sowie Dienste zur Überwachung des Datenprodukterfolgs. Neben technischen Diensten bedarf es zudem minimalen Standards zur Erreichung von Interoperabilität. Diese beinhalten sowohl technische Standards (beispielsweise Datenprodukt-Templates) als auch inhaltliche Standards (beispielsweise semantische Modelle). Diese Elemente sollten von der Trägerschaft des Gesundheitsdatenraums verfügbar gemacht werden, um die Komplexität der Datenbereitstellung zu reduzieren, insbesondere vor dem Hintergrund des unterschiedlichen Digitalisierungsgrades der beteiligten Akteure der Gesundheitswirtschaft.

Umsetzung von föderierter Gouvernanz

Wird eine föderierte Form der Gouvernanz im Datenraum Gesundheit eingeführt, so entstehen drei Gouvernanzebenen. Individuelle Akteure erhalten Freiheiten zu entscheiden, welche Gesundheitsdaten von ihnen zusätzlich zu den rechtlich verpflichtenden Daten bereitgestellt werden. Zudem sind zusätzliche Freiräume bei der Gestaltung der Datenprodukte hinsichtlich deren Beschreibung oder Kommunikation sowie der Verwendung nicht-verpflichtender Standards denkbar. Auf Datenraumebene kommen die Akteure, moderiert durch die Datenraumträgerschaft, zusammen, um Elemente wie die gesundheitsspezifische Semantik, die speziell benötigten Datenraumdienste und die zu verwendenden Austauschformate und Prozesse zu spezifizieren. Auf einer übergreifenden Ebene über verschiedene Datenräume hinweg (Data

Spaces Mesh) sollten gemeinsam minimale Standards, Dienste und Prozesse spezifiziert werden, um eine datenraumübergreifende Interoperabilität zu schaffen. Zu diesen Elementen gehören beispielsweise die Festlegung einer minimalen Datenproduktbeschreibung, eines gemeinsamen Datenlebenszyklus oder von gemeinsamen Identitäts- und Vertrauensdiensten. Hierzu könnten die relevanten Datenraumträgerschaften unter dem Schirm der Anlaufstelle Datenökosystem Schweiz zusammenkommen. Interoperabilität zwischen Datenräumen ermöglicht die einfache Umsetzung bereichsübergreifender Anwendungen. Beispielsweise könnten Daten des Krebsregisters mit Wetter- und Klimadaten wie UV-Strahlung, Temperaturen oder Feinstaubkonzentrationen zusammengeführt werden. Daraus ergeben sich erweiterte Potenziale sowohl für die Erforschung dieser Effekte als auch für eine daraus folgende datenbasierte Steuerung von politischen Maßnahmen.

5. Diskussion und Fazit

Eine Anwendung von Data Mesh-Ansätzen im Kontext des Gesundheitsdatenraums Schweiz offenbart verschiedene Stärken, Schwächen, Chancen und Risiken der Nutzung von Data Mesh-Praktiken in Datenräumen. Eine SWOT-Analyse fasst diese Implikationen zusammen. Abbildung 16 stellt die Ergebnisse der SWOT-Analyse dar.

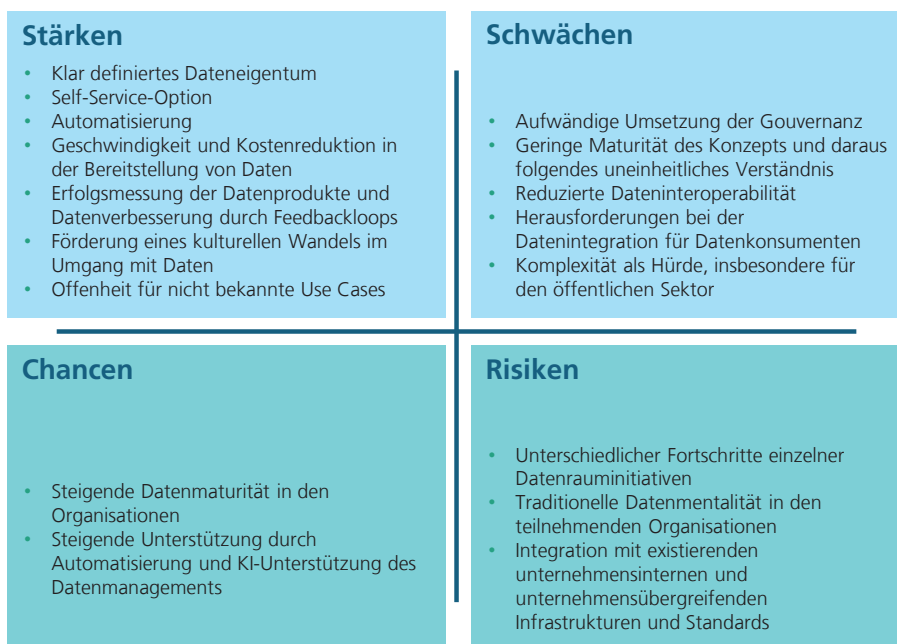


Abbildung 16: SWOT-Analyse von Data Mesh-Ansätzen in Datenräumen

Zu den eindeutigen Stärken des Ansatzes zählt die verpflichtende Übernahme von Dateneigentum und Verantwortung durch die Datenbereitstellenden. Daten müssen gemäß der in den Datenverträgen beschriebenen Eigenschaften verfügbar gemacht werden. Ebenso übernehmen Datennutzende die Verantwortung für die Erstellung nutzungsorientierter Datenprodukte und die für den jeweiligen Anwendungsfall notwendige Datentransformationen. Daten sowie die benötigten Infrastrukturkomponenten können im Rahmen eines Self-Service bezogen werden. Die Grundlage dafür ist die Automatisierung von Datenbereitstellungs- und Datenfreigabeprozessen, bei denen Akteure regelbasierten Zugriff auf Daten erhalten, sofern sie die Datenzugriffsbedingungen einhalten. Entsprechend erhöhen sich auch Datensicherheit und Datenschutz. Datenbereitstellende sind durch Data Mesh in der Lage, Daten einfacher und mit höherer Geschwindigkeit bereitzustellen, da die Aufwände für Datentransformationen und Datenbereinigungen reduziert werden. Gleichermassen kann die Qualität der bereitgestellten Datenprodukte durch die Ermöglichung von Feedbackloops zielgerichteter verbessert werden. Zudem wird eine Erfolgsmessung der Datenprodukte möglich. Die angebotenen Datenprodukte erlauben zudem eine einfache Datenmehrfachnutzung, da sie einfach auffindbar, analysierbar und nutzbar sind. Nicht zuletzt fördert die Umsetzung eines Datenraums in Form eines Data Mesh den kulturellen Wandel der einzelnen Akteure. Anstelle von Silodenken werden Offenheit, Selbstsouveränität und Vertrauen in der Datenbereitstellung und Datennutzung von den einzelnen Akteuren verlangt.

Allerdings ergeben sich auch Schwächen einer Nutzung Data Mesh-Konzepts im Rahmen von Datenräumen. Dazu gehört eine zeitintensive Umsetzung der föderierten Gouvernanz aufgrund des hohen Grads an Abstimmungsbedarfen und der Notwendigkeit zur Konsensfindung. Weiterhin ergeben sich aus der Neuartigkeit des Konzepts und fehlenden Best-Practices insbesondere im organisationsübergreifenden Rahmen derzeit noch Verständnisprobleme zwischen den verschiedenen Akteuren eines Datenraums hinsichtlich der Kernkonzepte von Data Mesh und deren Umsetzung. Durch die geringere Standardisierung der Daten und höhere Freiheitsgrade in Bezug auf die genutzten Technologien bei den Akteuren ist von einer reduzierten Dateninteroperabilität auszugehen. Entsprechend ergeben sich insbesondere für Datennutzende höhere Aufwände zur Integration von Daten. Durch die hohe Komplexität des Data Mesh-Ansatzes ergeben sich neue Aufgaben und Prozesse für alle Teilnehmenden eines Datenraums. Für Akteure mit einer geringeren Datenmaturität, wie beispielsweise im Bereich des öffentlichen Sektors, kann die Erfüllung der im Rahmen eines Data Mesh gestellten Anforderungen insgesamt zu erheblichen Änderungsbedarfen führen und eine Teilnahmehürde darstellen.

Chancen für die verstärkte Einbindung von Data Mesh-Ansätzen in Datenräumen liefert einerseits der kontinuierlich steigende Datenreifegrad in den Organisationen. Immer mehr Organisationen setzen eigene Datenstrategien und Datenmanagement-Konzepte um. Entsprechend steigen das Verständnis und die Bereitschaft, Daten im Sinne einer Data Mesh-Organisation auch mit anderen Akteuren in Datenräumen zu teilen. Weiterhin trägt die (KI-gestützte) Automatisierung wiederkehrender Datenmanagement-Tätigkeiten, beispielsweise der Beschreibung von Datenprodukten oder der Definition von Datennutzungsbedingungen, zu einer einfacheren und schnelleren Umsetzung von Data Mesh-Aspekten bei. Im Sinne der Weiterentwicklung des unternehmensinternen Datenmanagements sind hier auch für den organisationsübergreifenden Datenaustausch relevante Weiterentwicklungen zu erwarten.

Demgegenüber stehen allerdings auch Risiken, die es bei der Umsetzung von Datenräumen in Form eines Data Mesh zu beachten gilt. So ergeben sich insbesondere hinsichtlich der datenraumübergreifenden Definition von Diensten und Richtlinien Schwierigkeiten aufgrund des unterschiedlichen Fortschrittsgrades der Datenrauminitiativen im Datenökosystem Schweiz. Gegebenenfalls ist es daher nicht immer möglich, eine datenraumübergreifende Interoperabilität zu ermöglichen. Zumindest innerhalb einzelner Datenräume sollte entsprechend berücksichtigt werden, dass bei der Umsetzung der föderierten Gouvernanz eine ausgewogene Berücksichtigung der unterschiedlichen Interessen erfolgt. Weiterhin stehen traditionelle Mentalitäten im Umgang mit Daten der Einführung von Data Mesh-Ansätzen in Datenräumen entgegen. Beispielsweise ist im Gegensatz zu herkömmlichen Datenaustauschprozessen eine aktivere Kollaboration der Datennutzenden gefordert, um Datenprodukte gemeinsam umzusetzen und weiterzuentwickeln. Zudem könnten sich, wie im Praxisbeispiel des Krebsregisters gezeigt, zusätzliche Aufwände zur Datenintegration auf Seiten der Datennutzenden ergeben.

Hier sind entsprechend Widerstände zu erwarten, die auf kultureller Ebene angegangen werden müssen. Zudem stellt die Integration bestehender Datenaustauschsysteme mit oftmals zentralisierter Architektur und vorhandener Standards in das Data Mesh-Konzept eine Herausforderung dar.

Abschließend ergeben sich aus der gemeinsamen Betrachtung von Data Mesh und Datenräumen zudem geteilte Fragestellungen, die es im Rahmen der zukünftigen Entwicklungen zu beantworten gilt. Hierzu zählt der Aufbau einer Kultur, die das Teilen von Daten fördert und auf Offenheit, Vertrauen und Zusammenarbeit basiert. Es gilt, geeignete Change-Management-Strategien zu entwickeln, um den kulturellen Wandel zu unterstützen und potenzielle Widerstände zwischen den Datenaustauschpartnern zu überwinden. Weiterhin ist eine Abwägung zwischen den Vorteilen zentralisierter und dezentralisierter Architekturen zu treffen, um einen Ausgleich zwischen den Ansprüchen der diversen involvierten Stakeholder zu schaffen.

Letztendlich steht und fällt zudem der Erfolg von Data Mesh- und Datenraumimplementierungen mit der erfolgreichen Umsetzung wertschaffender Dienste und Produkte. Erfolgreiche Anwendungsfälle incentivieren Datenbereitstellende und Datennutzende zur Teilnahme am Data Sharing und fördern die Akzeptanz der gewählten Lösungen. Darüber hinaus sind im Bereich der Datenräume zusätzliche Mechanismen zur Vertrauensbildung zwischen potenziell unbekannten Parteien und zur Sicherung der Datensouveränität der Datenbereitstellenden bei Datenweitergabe zu berücksichtigen, die über die im Rahmen von Data Mesh angestellten Überlegungen hinausgehen.

Zusammenfassend lässt sich festhalten, dass der Einsatz von Data Mesh-Konzepten im Sinne von Datenräumen einen vielversprechenden Ansatz darstellt, um die Autonomie der Akteure zu fördern, die Geschwindigkeit der Datenbereitstellung zu erhöhen und ein nutzenorientiertes Datenmanagement zu ermöglichen. Die konkrete Umsetzung, insbesondere im Hinblick auf den Grad der Dezentralisierung und Autonomie, sollte jedoch in den einzelnen Datenräumen vor dem Hintergrund der anvisierten Akteurszielgruppe, der rechtlichen und gesellschaftlichen Rahmenbedingungen und der zu realisierenden Anwendungsfälle genau geprüft werden.

Datenraumträgerschaften kommt dabei die Rolle zu, einen Ausgleich zwischen den Interessen digitaler und weniger digitaler Akteure zu schaffen und beide Gruppen gleichermaßen zur Teilnahme an einem Datenraum zu befähigen. Die Akteure selbst sollten zur Umsetzung von Data Mesh ihre eigene Datenmanagement-Maturität erhöhen und somit die Grundlage für ein Teilen von Daten in Datenräumen nach Data Mesh-Prinzipien schaffen. Zudem sollten diese kontinuierlich prüfen, welche Mehrwertanwendungen auf geteilten Daten erzeugt werden können und die Datenbedarfe an die relevanten Akteure in den Datenräumen weitergeben. Schlussendlich sollte die Anlaufstelle Datenökosystem Schweiz intensiv an der Interoperabilität zwischen Datenräumen auch im Hinblick auf die genutzten Dienste hinwirken und den Austausch zwischen einzelnen Datenräumen fördern, um die Anschlussfähigkeit der Daten zu gewährleisten.

6. Anhang

6.1. Glossar

Das Glossar bietet einen Überblick über relevante Begriffe im Bereich Datenräume und Data Mesh in deutscher und englischer Sprache, um die Orientierung in dieser Studie und der weiterführenden Literatur zu erleichtern.

Englisch	Deutsch (CH)
Data Advocacy	Datenanwaltschaft
Data Analysts	Datenanalyst/-innen
Data Discovery	Datenentdeckung
Data Domain	Datendomäne
Source-aligned Domain	Quellenorientierte Domäne
Aggregate Domain	Aggregierende Domäne
Consumer-aligned Domain	Nutzungsorientierte Domäne
Data Ecosystem	Datenökosystem
Data Fabric	Data Fabric
Data Hub	Daten-Hub
Data Governance	Datengouvernanz
Data Lake	Data Lake
Data Lakehouse	Data Lakehouse
Data Leadership	Datenführung
Data Lineage	Datenherkunft
Data Literacy	Datenkompetenz
Data Management	Datenmanagement
Data Marketplace	Datenmarktplatz
Data Mesh	Data Mesh
Data Observability	Datenbeobachtbarkeit
Data Owner	Verantwortliche für die Datenbearbeitung
Data Ownership	Dateneigentum
Data Product	Datenprodukt
Data Provider	Datenbereitstellende
Data Scientists	Datenwissenschaftler/-innen
Data Space	Datenraum
Data Spaces Building Blocks	Datenraumbausteine
Data Spaces Participant	Datenraumteilnehmende
Data User	Datennutzende
Data Warehouses	Data Warehouses
Governance	Gouvernanz
Knowledge Graphs	Wissensgraphen
Metadata	Metadaten
Participants	Teilnehmende
Policies	Richtlinien
Sovereignty	Souveränität

7. Literaturverzeichnis

Bean, R. (2021). *Why Is It So Hard to Become a Data-Driven Company?* <https://hbr.org/2021/02/why-is-it-so-hard-to-become-a-data-driven-company>

Bitkom e. V. (Hrsg.). (2022). *Data Mesh – Datenpotenziale finden und nutzen: Ein Praxisleitfaden des Digitalverbands Bitkom*. https://www.bitkom.org/sites/main/files/2022-06/220531_LF_Data_Mesh.pdf

Blohm, I., Wortmann, F., Legner, C. & Köbler, F. (2024). Data products, data mesh, and data fabric. *Business & Information Systems Engineering*, 66(5), 643–652. <https://doi.org/10.1007/s12599-024-00876-5>

Bode, J., Kühl, N., Kreuzberger, D., Hirschl, S. & Holtmann, C. (2023). *Towards Avoiding the Data Mess: Industry Insights from Data Mesh Implementations*. <https://doi.org/10.48550/arXiv.2302.01713>

Bühlmann, S. (2024). *Data Hub: Definition und Bedeutung*. <https://dataminds.ch/glossar/data-hub-definition-und-bedeutung/>

Bundeskanzlei – Digitale Transformation und IKT Lenkung. (2025). *Bausteine von Datenräumen: Datenökosystem Schweiz*. https://www.bk.admin.ch/bk/de/home/digitale-transformation-ikt-lenkung/datenoekosystem_schweiz/grundlagen-fuer-datenraeume.html

CEN/CENELEC. *CWA 18125 - Trusted Data Transaction*. https://www.cencenelec.eu/media/CEN-CENELEC/CWAs/RI/2024/cwa18125_2024.pdf

Dehghani, Z. (2022). *Data mesh: Delivering data-driven value at scale*. O'Reilly.

Dolhopolov, A., Castelltort, A. & Laurent, A. (2024). Implementing Federated Governance in Data Mesh Architecture. *Future Internet*, 16(4), 115. <https://doi.org/10.3390/fi16040115>

DSSC. (2024). *Data Spaces Blueprint v1.5*. Data Spaces Support Centre. <https://dssc.eu/space/bv15e/766064046/Building+Block+Overview>

Eichler, R., Gröger, C., Hoos, E., Stach, C., Schwarz, H. & Mitschang, B. (2023). Introducing the enterprise data marketplace: a platform for democratizing company data. *Journal of Big Data*, 10(1). <https://doi.org/10.1186/s40537-023-00843-z>

Estermann, B., Fraefel, M., Neuron, A. C. & Vogel, J. (2018). Conceptualizing a national data infrastructure for Switzerland. *Information Polity*, 23(1), 43–65. <https://doi.org/10.3233/IP-170033>

Falconi, M. & Plebani, P. (2023). Adopting Data Mesh principles to Boost Data Sharing for Clinical Trials. In C. K. Chang (Hrsg.), *2023 IEEE International Conference on Digital Health: IEEE ICDH 2023 : hybrid conference, Chicago, Illinois, 2-8 July 2023 : proceedings* (S. 298–306). IEEE. <https://doi.org/10.1109/ICDH60066.2023.00051>

- Franklin, M., Halevy, A. & Maier, D. (2005). From databases to dataspace: a new abstraction for information management. *ACM Sigmod Record*, 34(4), 27–33. <https://doi.org/10.1145/1107499.1107502>
- Gelhaar, J., Gürpınar, T., Henke, M. & Otto, B. (2021). Towards a Taxonomy of Incentive Mechanisms for Data Sharing in Data Ecosystems. *Twenty-fifth Pacific Asia Conference on Information Systems*. <https://aisel.aisnet.org/pacis2021/121/>
- Hackl, T. & Winter, R. (2023). Data Mesh at Scale: Exploration of current practices in large organizations. *University of St.Gallen (HSG)*. https://www.researchgate.net/publication/379513337_Data_Mesh_at_Scale_-_Exploration_of_current_practices_in_large_organizations
- Hasan, M. R. & Legner, C. (2023). Understanding Data Products: Motivations, Definitions, and Categories. *ECIS(229)*. https://aisel.aisnet.org/ecis2023_rp/229
- Hasan, M. R. & Legner, C. (2024). Improving Consumer-Provider interaction with Data Products: Insights from traditional industries. *ECIS*, 2024. https://aisel.aisnet.org/ecis2024/track07_busanalytics/track07_busanalytics/17/
- Jahnke, N., Schoormann, T., Jussen-Lengersdorf, I. & Möller, F. (2024). Designing Federated Data Marketplaces in Industrial Production: *Findings from a Prototypical Implementation*. <https://aisel.aisnet.org/wi2024/61/>
- Krystek, M. (2023). Introducing Data Mesh paradigm for Smart City platforms design. *Proceedings of the 56th Hawaii International Conference on System Sciences 2023*. https://aisel.aisnet.org/hicss-56/st/smart_app_development/2/
- Lefebvre, H. & Legner, C. (2022). HOW COMMUNITIES OF PRACTICE ENABLE DATA DEMOCRATIZATION INSIDE THE ENTERPRISE. *Thirtieth European Conference on Information Systems (ECIS 2022)*. https://aisel.aisnet.org/ecis2022_rp/141/
- Machado, I. A., Costa, C. & Santos, M. Y. (2022). Data Mesh: Concepts and Principles of a Paradigm Shift in Data Architectures. *Procedia Computer Science*, 196, 263–271. <https://doi.org/10.1016/j.procs.2021.12.013>
- Möller, F., Jussen, I., Springer, V., Gieß, A., Schweihoff, J. C., Gelhaar, J., Guggenberger, T. & Otto, B. (2024). Industrial data ecosystems and data spaces. *Electronic Markets*, 34(1). <https://doi.org/10.1007/s12525-024-00724-0>
- Mucci, T. & Stryker, C. (2024). *What is data as a product (DaaP)?* <https://www.ibm.com/topics/data-as-a-product>
- Otto, B. (2022). The Evolution of Data Spaces. In B. Otto, M. ten Hompel & S. Wrobel (Hrsg.), *Designing Data Spaces* (S. 3–15). Springer International Publishing. https://doi.org/10.1007/978-3-030-93975-5_1
- Priebe, T., Neumaier, S. & Markus, S. (2021). Finding Your Way Through the Jungle of Big Data Architectures. In Y. Chen (Hrsg.), *2021 IEEE International Conference on Big Data: Dec 15-Dec 18, 2021, virtual event : proceedings* (S. 5994–5996). IEEE. <https://doi.org/10.1109/BigData52589.2021.9671862>

Ribeiro, V., Barata, J. & Da Rupino Cunha, P. (2024). A Maturity Model for Data Governance in Decentralized Business Operations: Architecture and Assessment Archetypes. In *International Conference on Information Systems Development, Proceedings of the 32nd International Conference on Information Systems Development*. University of GdańskGdańsk, Poland. <https://doi.org/10.62036/ISD.2024.5>

Serra, J. (2024). *Deciphering data architectures: Choosing between a modern data warehouse, data fabric, data, data lakehouse, and data mesh* (First edition). O'Reilly Media, Inc.

Strengtholt, P. (2023). *Data Management at Scale: Modern data architecture with data mesh and data fabric* (Second edition). O'Reilly.

Sutherland, M. K. & Cook, M. E. (2017). Data-Driven Smart Cities. In C. C. Hinnant (Hrsg.), *ACM Digital Library, Proceedings of the 18th Annual International Conference on Digital Government Research* (S. 471–476). ACM. <https://doi.org/10.1145/3085228.3085239>

Ulich, E. (2015). Arbeitssysteme als Soziotechnische Systeme – eine Erinnerung. *Journal Psychologie des Alltagshandelns*, 6(1). http://www.allgemeine-psychologie.info/cms/images/stories/allgpsy_journal/Vol%206%20No%201/Arbeitssystem_Ulich.pdf

UVEK & EDA. (2022). *Schaffung von vertrauenswürdigen Datenräumen basierend auf der digitalen Selbstbestimmung*. <https://www.news.admin.ch/news/message/attachments/70835.pdf>



Kontakt

Nils Jahnke
Mobility & Smart Cities
Tel. +49 231 97677-467
nils.jahnke@isst.fraunhofer.de

Fraunhofer-Institut für Software- und
Systemtechnik ISST
Speicherstraße 6
44147 Dortmund
www.fraunhofer.de