

Technologieprogramm Edge Datenwirtschaft
des Bundesministeriums für Wirtschaft und Klimaschutz



DATENINFRASTRUKTUREN FÜR DATENÖKOSYSTEME

Eine Entscheidungshilfe für die Gestaltung der Systemarchitektur von Dateninfrastrukturen für das Data Sharing

Impressum

Diese Entscheidungshilfe wurde im Auftrag des Bundesministeriums für Wirtschaft und Klimaschutz im Rahmen der Begleitforschung zum Technologieprogramm „Edge Datenwirtschaft“ erstellt.

Autor

Nils Jahnke

Fraunhofer-Institut für Software- und Systemtechnik ISST
Dortmund

Herausgeber

Peter Gabriel

Dr. Nicole Wittenbrink

Begleitforschung Edge Datenwirtschaft
Institut für Innovation und Technik (iit)
in der VDI / VDE Innovation + Technik GmbH
Berlin

Datum

März 2025

Layout

PRpetuum GmbH

Executive Summary

Mithilfe von neuartigen, datengetriebenen Anwendungen generieren Unternehmen umfassende wirtschaftliche Mehrwerte und adressieren gleichzeitig einige der großen Herausforderungen der heutigen Zeit wie den Fachkräftemangel oder die Einsparung von natürlichen Ressourcen. Um die Potenziale datengetriebener Anwendungen zu heben, sind Unternehmen häufig auf die Bereitstellung von Daten durch externe Akteure – das so genannte Data Sharing – angewiesen. Datenbereitstellende erhalten im Gegenzug für die Weitergabe ihrer Daten Kompensationen. Eine Kompensation kann die Nutzung datengetriebener Anwendungen oder die vereinfachte Erfüllung regulatorischer Vorgaben umfassen. Besitzen die Datenaustauschbeziehungen zwischen den Akteuren eine netzwerkartige Konstellation, entstehen Datenökosysteme, in denen Unternehmen rund um Daten als zentrale Ressource miteinander kollaborieren und konkurrieren.

Zur Umsetzung datengetriebener Anwendungen müssen die Daten zur richtigen Zeit und in der benötigten Qualität verfügbar sein. Dies ist die Aufgabe einer Dateninfrastruktur. Eine anforderungsorientierte Dateninfrastruktur schafft die Grundlage für Effizienz und Leistungsfähigkeit der auf ihr basierenden Dienste und fördert die Skalierbarkeit der umzusetzenden Geschäftsmodelle.

Die vorliegende Entscheidungshilfe unterstützt Unternehmen und Industriekonsortien dabei, eine auf Basis der organisatorischen Rahmenbedingungen und der geplanten Anwendungsfälle passende Systemarchitektur der Dateninfrastruktur für das organisationsübergreifende Data Sharing auszuwählen. Insbesondere adressiert die Entscheidungshilfe Architekturteams und Entscheidungsbefugte in den jeweiligen Initiativen, die sich mit dem Neuaufbau oder der Weiterentwicklung der Dateninfrastruktur befassen. Mittels der Entscheidungshilfe werden zentrale, föderierte, dezentrale und verteilte Architekturen gegenübergestellt. Anhand von individuell abgeleiteten Entscheidungskriterien wird als Resultat die passendste Gestaltungsoption der Dateninfrastruktur identifiziert.

Basierend auf etablierten Methoden aus dem Bereich der Unternehmensarchitektur erläutert die Entscheidungshilfe vier Phasen des Entscheidungsprozesses:

- 1) Umweltanalyse: Identifizierung, Beschreibung und Evaluation von Einflüssen und Interessen zur Gestaltung der Dateninfrastruktur aus der Umwelt
- 2) Geschäftsmodellanalyse: Ableitung und Charakterisierung relevanter Daten, Datenflüsse und Systeme aus dem Geschäftsmodell
- 3) Anforderungsdefinition: Identifikation und Priorisierung relevanter Anforderungen an die Dateninfrastruktur als Entscheidungsbasis
- 4) Grobdesign: Erarbeitung und Evaluation möglicher Designoptionen und Herausstellung der bestmöglichen Alternative



Als zentrales Hilfsmittel bietet diese Entscheidungshilfe einen Kriterienkatalog zur Evaluation der Designoptionen von Dateninfrastrukturen. Dieser Kriterienkatalog erläutert die Entscheidungsfaktoren für Dateninfrastrukturen, ihre Relevanz für spezielle Data Sharing-Konstellationen sowie mögliche Interdependenzen zwischen den Faktoren und liefert schließlich eine Einschätzung, inwieweit die Entscheidungsfaktoren durch die ausgewählten Gestaltungsoptionen erfüllt werden.



Inhalt

1	Einleitung	4
2	Grundlegende Dienste und Gestaltungsoptionen von Dateninfrastrukturen.....	6
2.1	Bestandteile von Dateninfrastrukturen.....	6
2.2	Gestaltungsoptionen von Dateninfrastrukturen.....	8
3	Vorgehensmodell zur Auslegung von Dateninfrastrukturdiensten.....	15
3.1	Phase 1 – Umweltanalyse	16
3.2	Phase 2 – Geschäftsmodellanalyse	18
3.3	Phase 3 – Anforderungsdefinition.....	21
3.4	Phase 4 – Grobdesign.....	24
4	Kriterienkatalog für Dateninfrastrukturdienste.....	25
5	Ausblick: die nächsten Schritte	35
	Referenzen.....	36

1 Einleitung

Neuartige, datenbasierte Anwendungen ermöglichen umfassende Verbesserungen von Prozessen, Produkten und Serviceleistungen mit ökonomischem und gesellschaftlichem Mehrwert. Eine Studie des MIT zeigt, dass Unternehmen, in denen Abläufe datenbasiert, automatisiert und in Echtzeit stattfinden, ein um mehr als 50 Prozent höheres Wachstum erreichen als deren weniger digitalisierte Konkurrenz (Weill et al., 2024). Beispiele für datenbasierte Anwendungen sind die Messung und Optimierung der Energieverbräuche von Produktionsprozessen oder die intelligente Einbindung dezentraler Kleinanlagen zur Stabilisierung des Stromnetzes. Aufgrund der hohen Komplexität erfolgt die Wertschöpfung aus Daten in diesen Anwendungen zu meist nicht alleine durch einen Akteur. Stattdessen schließen sich mehrere Akteure zur kollaborativen Leistungserbringung zusammen.

Zentrale Komponente dieser Kollaboration ist das Teilen von Daten (Data Sharing). Durch die Datenaustauschbeziehungen der Akteure entstehen sogenannte Datenökosysteme. Datenökosysteme sind eine neuartige, komplexe Organisationsform, in denen autonome Akteure mit gemeinschaftlichen und gegensätzlichen Interessen miteinander kollaborieren. Daten müssen zur Umsetzung der wertschöpfenden Anwendungen über die Grenzen einzelner Standorte und über die Einflussbereiche einzelner Akteure hinweg nutzbar gemacht werden.

Je nach Anwendung existieren unterschiedliche Anforderungen an das Data Sharing. So müssen für manche Anwendungen Daten mit besonders geringen Latenzzeiten zur Verfügung gestellt werden, wohingegen andere Anwendungen besonderen Wert auf den Schutz von personenbezogenen Daten bei der Datenbereitstellung legen.

Grundlage des Data Sharings ist eine für die Austauschpartner langfristig zur Verfügung stehende, zuverlässige und vertrauenswürdige Dateninfrastruktur, welche die Entwicklung einer Vielzahl neuartiger Anwendungen befähigt. Dateninfrastrukturen stellen technische Dienste, gemeinsame Standards, Regeln und Richtlinien sowie Vertrauensmechanismen bereit (Dodds & Wells, 2019). Konkrete Dienste von Dateninfrastrukturen sind beispielsweise ein Teilnehmendenverzeichnis, ein Identitätsmanagement oder Kataloge zur Beschreibung und Entdeckung von verfügbaren Daten und Services.

Entwurf und Umsetzung der Dateninfrastruktur können entweder durch ein einzelnes Unternehmen oder in einem kollaborativen Prozess mehrerer beteiligter Organisationen erfolgen. Um eine für die Anwendungszwecke passende Lösung zu entwickeln, müssen eine Reihe von Fragen aus den Bereichen der Planung, der Gestaltung, der Implementierung und der Wartung von Dateninfrastrukturen beantwortet werden (Otto et al., 2022). Das Design einer Dateninfrastruktur muss zwischen den Anforderungen der Datenbereitstellenden, Datengebenden und der Allgemeinheit abwägen, um möglichst viele Akteure von einem Teilen ihrer Daten zu überzeugen. Eine wichtige Fragestellung ist daher, ob Dateninfrastrukturdienste in zentralisierter oder dezentralisierter Form gestaltet werden sollen. Angesichts der Neuartigkeit des Themenkomplexes fällt es Verantwortlichen jedoch häufig schwer, eine informierte Entscheidung hinsichtlich der Auslegung der Dateninfrastrukturdienste zu treffen. Insbesondere vor dem Hintergrund des organisationsübergreifenden Teilens von Daten ergeben sich komplexe Anforderungen und Rahmenbedingungen, die von herkömmlichen unternehmensinternen Entwurfs- und Entwicklungsprozessen abweichen. Dazu zählt beispielsweise die Notwendigkeit, eine Balance zwischen individuellen und gemeinschaftlichen Interessen der Akteure zu finden.

Die vorliegende Dokument geht diese Herausforderung an. Unternehmen, die gemeinsam mit anderen Akteuren eine auf gemeinsam genutzten Daten basierende Anwendung technisch umsetzen wollen, erhalten eine Entscheidungshilfe für die Gestaltung von Dateninfrastrukturdiensten. Ziel der Entscheidungshilfe ist es, die bestmögliche Gestaltungsoption für die Dateninfrastruktur auf Basis der Rahmenbedingungen und konkret umzusetzenden Anwendungen zu identifizieren. Insbesondere werden in dieser Entscheidungshilfe die Anliegen kleiner und mittlerer Unternehmen und kleinerer Dateninfrastrukturinitiativen berücksichtigt. Konkret richtet sich die Entscheidungshilfe an Systemarchitektinnen und -architekten sowie an Entscheidungsbeauftragte, die sich in ihren Unternehmen mit der Umsetzung von kollaborativen datenbasierten Anwendungen befassen.

Die Entscheidungshilfe führt zunächst durch eine Beschreibung von relevanten Dateninfrastrukturdiensten und deren Gestaltungsoptionen in die Thematik von Dateninfrastrukturen ein. Anschließend präsentiert sie ein Vorgehensmodell zur Auswahl einer passenden Designoption von Dateninfrastrukturdiensten für kollaborative Anwendungsvorhaben in der Datenwirtschaft. Das Vorgehensmodell inkludiert die vier Phasen „Umweltanalyse“, „Geschäftsmodellanalyse“, „Anforderungsdefinition“ und „Grobdesign“ und orientiert sich an etablierten Konzepten aus dem Bereich der Unternehmensarchitektur. Zentrales Element dieser Entscheidungshilfe ist ein Kriterienkatalog zur Evaluation der Gestaltungsoptionen von Dateninfrastrukturdiensten. Der Kriterienkatalog beschreibt die wesentlichen Entscheidungsfaktoren für Dateninfrastrukturen, ihre Bedeutung für spezifische Data Sharing-Szenarien und legt dar, inwiefern einzelne Gestaltungsoptionen die Entscheidungsfaktoren erfüllen können.

Diese Entscheidungshilfe ist wie folgt strukturiert: Abschnitt 2 erläutert die grundlegenden Dienste von Dateninfrastrukturen und gibt eine Übersicht zu möglichen Gestaltungsoptionen. Abschnitt 3 präsentiert ein Vorgehensmodell zur Evaluation der Gestaltungsoptionen von Dateninfrastrukturdiensten. Der zugehörige Kriterienkatalog für Dateninfrastrukturdienste ist in Abschnitt 4 dargestellt. Der abschließende Ausblick empfiehlt weiterführende Literatur zu einzelnen Referenzarchitekturen als Ausgangspunkt für die technische Realisierung (Abschnitt 5).

Die Zusammenarbeit mit den Experten aus den Early-Adopter-Projekten des Technologieprogramms Edge Datenwirtschaft hat maßgeblich dazu beigetragen, praxisnahe Erkenntnisse bei der Entwicklung von Dateninfrastrukturen in diese Entscheidungshilfe zu integrieren und die Anforderungen unterschiedlicher Geschäftsbereiche und Unternehmensgrößen umfassend zu adressieren. Das Autorenteam möchte sich daher an dieser Stelle noch einmal herzlich und ausdrücklich bei allen Interviewpartnern für ihre Unterstützung bei der Erstellung dieser Entscheidungshilfe bedanken:

- Lars Nolting, Tennet TSO GmbH, Projekt DEER
- Florian Remark, Strategion GmbH, Projekt SECAI
- Bruno Ristok, C&S Computer und Software GmbH, Projekt Careful-Edge-X
- Heinrich Traphöner, Empolis GmbH, Projekt EASY

Die Verantwortung für den Inhalt diese Entscheidungshilfe liegt ausschließlich bei den Autoren.

Die Entscheidungshilfe wurde im Rahmen der Begleitforschung zum Technologieprogramm Edge Datenwirtschaft im Auftrag des Bundesministeriums für Wirtschaft und Klimaschutz erstellt. Das Programm umfasst zehn Projekte, die neue Formen von Edge-Cloud-Systemen für wichtige Sektoren der deutschen Wirtschaft entwickeln.

2 Grundlegende Dienste und Gestaltungsoptionen von Dateninfrastrukturen

2.1 Bestandteile von Dateninfrastrukturen

Dateninfrastrukturen umfassen üblicherweise drei Arten von Bausteinen zur Ermöglichung von Data Sharing-Anwendungen: Technische Dienste befähigen Datenaustausch und Datenintegration zwischen physisch separierten Systemen und sorgen für die Sicherheit vor unbefugtem Zugriff oder Datenverlust. Standards sorgen für die Vereinheitlichung von Daten und Prozessen und damit für organisatorische, semantische und technische Interoperabilität. Regeln und Richtlinien definieren ein gemeinsames Rahmenwerk für die Nutzung der Dateninfrastruktur und die Gestaltung des Data Sharings.

Im Fokus aktueller Entwicklungen im Bereich von Dateninfrastrukturen steht insbesondere die Entwicklung von Dateninfrastrukturdiensten (Demchenko et al., 2014; Otto & Burmann, 2021). Sowohl die Europäische Kommission als auch die Bundesregierung haben aufgrund des hohen Potenzials von Dateninfrastrukturen für die Erzeugung datengetriebener Wertschöpfung in den letzten Jahren öffentliche Fördermittel bereitgestellt, um die Konzeptualisierung und Entwicklung von Dateninfrastrukturdiensten in unterschiedlichen Initiativen voranzutreiben. Daraus entstanden sind Vereine wie die International Data Spaces Association (IDSA)¹ oder Gaia-X², öffentliche Initiativen wie SIMPL³ oder 8ra⁴, open-source Entwicklungsarbeiten unter dem Schirm der Eclipse Dataspaces Working Group⁵ oder CKAN⁶ sowie weitere private Dateninfrastrukturinitiativen. Die Inhalte der Initiativen unterscheiden sich im dabei Hinblick auf ihre Fokusbereiche und die erbrachte Leistung (beispielsweise Konzeptualisierung gegenüber Implementierung). Allerdings lassen sich über die Initiativen hinweg gemeinsame Dateninfrastrukturdienste identifizieren, die sich zwischen den Initiativen durch ihre jeweiligen Architekturansätze und Technologiestacks unterscheiden. Zu diesen Diensten gehören ein Teilnehmendenverzeichnis, Komponenten für das Identitäts- und Zugriffsmanagement sowie ein Daten- und Dienstekatalog. Zusätzlich sind je Initiative weitere individuelle Dienste, beispielsweise zur Erzeugung von Datensouveränität oder Datenintegrität, denkbar.

Teilnehmendenverzeichnis

Ein Teilnehmendenverzeichnis (engl. Registry) operationalisiert die Verwaltung der am Data Sharing beteiligten Akteure sowie der erreichbaren Endpunkte, beispielsweise von intelligenten Diensten oder intelligenten Geräten. Es listet grundlegende Informationen über die Teilnehmenden auf und macht diese Informationen auf globaler Ebene auffindbar. Eine Eintragung in das Teilnehmendenverzeichnis erfolgt üblicherweise, nachdem ein automatisierter oder manueller

¹ <https://internationaldataspaces.org/>

² <https://gaia-x.eu/>

³ <https://digital-strategy.ec.europa.eu/en/policies/simpl>

⁴ <https://www.8ra.com/>

⁵ <https://dataspace.eclipse.org/projects/>

⁶ <https://ckan.org/>

Onboarding-Prozess durch den Akteur durchlaufen wurde. Eine Listung im Teilnehmendenverzeichnis erzeugt somit eine grundlegende Vertrauensbasis zwischen den Datenaustauschpartnern und ist eine notwendige Bedingung für die Nutzung weiterer Dateninfrastrukturdienste und für weitere Interaktionen zwischen den Teilnehmenden. Übliche Anforderungen an ein Teilnehmendenverzeichnis sind die Integrität, Verfügbarkeit und Unverfälschbarkeit der gespeicherten Daten.

Exkurs: Akteure und Teilnehmende im Kontext von Dateninfrastrukturen

Als Akteur wird in diesem Dokument jegliche Organisation oder natürliche Person mit Bezug zur Realisierung von Mehrwerten auf Basis geteilter Daten verstanden. Im Bereich des Data Sharing sind Akteure in der Regel Organisationen. Wenn eine Organisation die notwendigen Onboarding-Prozesse erfolgreich durchlaufen hat, wird sie zu einem Teilnehmenden einer Dateninfrastruktur. Die Teilnehmenden einer Dateninfrastruktur können je nach Anwendungsfall eine oder mehrere Rollen einnehmen. Als Datenbereitstellende stellen sie verfügbare Daten über die Dateninfrastruktur zur Verfügung und erhalten dazu monetäre oder nicht-monetäre Kompensationen. Als Datennutzende konsumieren sie die Daten der Datengebenden über die Dateninfrastruktur, um Mehrwerte zu realisieren und stellen Kompensationen für andere Teilnehmende zur Verfügung. Darüber hinaus können Teilnehmende die Rolle von Betreibenden der Dateninfrastrukturdienste einnehmen und damit die Funktionalität der Dateninfrastruktur sicherstellen. Für den Betrieb von Dateninfrastrukturdiensten werden üblicherweise monetäre Anreize geschaffen.

Identitäts- und Zugriffsmanagement

Eine Identität wird als die Summe von Eigenschaften eines Akteurs verstanden, um diesen in einem spezifischen Kontext identifizieren zu können. Zu einer Identität gehören ein eindeutiger Bezeichner und zusätzliche Attribute. Je nach Akteurstyp können diese zusätzlichen Attribute unterschiedlicher Art sein. Für natürliche Personen sind z.B. Name, Adresse, Alter oder Organisationszugehörigkeit typische Attribute. Für Organisationen sind z.B. Organisationsname, Rechtsform, juristische Adresse und erhaltene Zertifizierungen denkbar. Die Identität eines Akteurs wird benötigt, um festzustellen, welche Anwendungen und Dienste genutzt werden dürfen und welche Privilegien der Akteur während der Nutzung hat. Mögliche Attribute zur eindeutigen Identifizierung eines Akteurs müssen daher für den jeweiligen Kontext festgelegt werden.

Wichtige Funktionen eines Identitäts- und Zugriffsmanagements (Identity and Access Management) sind die Authentifizierung und die Autorisierung von Akteuren. Im Rahmen der Authentifizierung wird sichergestellt, dass es sich bei der Identität wirklich um den Akteur handelt, den sie vorgibt zu sein. Die Autorisierung umfasst die Festlegung, welche Berechtigungen ein Akteur auf Basis seiner Attribute erhält. Daher ist ein gewisses Vertrauen in die Richtigkeit der Attribute notwendig. Teilweise wird diese Vertrauenskomponente des Identitätsmanagements auch explizit als „Identity & Trust“ bezeichnet, um die besondere Rolle für die Herstellung von Vertrauen zwischen den Akteuren hervorzuheben. Das Identitäts- und Zugriffsmanagement ist essenziell, um den Zugriff auf andere Dienste der gleichen Dateninfrastruktur, den Abruf und die Bereitstellung von Daten sowie die Nutzung von intelligenten Diensten durch einzelne Akteure zu verwalten.

Daten- und Dienstekatalog

Der Katalogdienst (Catalog) listet Informationen über Daten und Serviceangebote von Teilnehmenden einer Dateninfrastruktur mithilfe von Metadaten. Metadaten werden oft als „Daten über Daten“ beschrieben und umfassen strukturierte Informationen über technische Eigenschaften, organisatorische Hintergründe und Konzepte sowie die Herkunft und Verwendung der Daten.

Metadaten sind damit üblicherweise weniger sensitiv als die zur Wertschöpfung verwendeten eigentlichen Daten.

Die das Daten- und Serviceangebot beschreibenden Metadaten werden üblicherweise von den Anbietenden im Katalog gelistet. Teilnehmende können die Metadaten innerhalb eines Katalogs einsehen, wenn sie die vom Datenanbietenden definierten und vom Betreibenden des Katalogs umgesetzten Zugriffsrichtlinien erfüllen. Kataloge erfüllen damit wichtige Funktionen entlang der Datenwertschöpfungskette. Durch die Möglichkeiten zur Inventarisierung und Entdeckung der Angebote führen Datenkataloge Angebot und Nachfrage im Bereich des Data Sharing zusammen. Zudem werden verfügbare Daten und Dienste bereits vor ihrer Nutzung analysier- und bewertbar. Auf Basis der im Katalog kuratierten Informationen kann zudem eine Nutzungsanfrage oder ein Abruf der angebotenen Daten oder Dienste stattfinden.

Zusatzdienste

Je nach Kontext und Einsatzbereich der Dateninfrastruktur sind weitere zusätzliche Dateninfrastrukturdienste denkbar. Hierzu zählen Verzeichnisdienste für standardisierte semantische Datenmodelle. Semantische Datenmodelle legen fest, welche Konzepte durch Daten beschrieben werden, wie diese Konzepte gruppiert und klassifiziert werden und welche Beziehung sie zueinander haben. Semantische Datenmodelle tragen somit zu einem besseren Verständnis der Daten bei.

Als Standardisierte Endpunkte zu den Systemen der Datenanbietenden und Datennutzenden können Datenkonnektoren als weitere Zusatzdienste eingesetzt werden. Datenkonnektoren sind in der Lage, Daten und Metadaten zu empfangen und bereitzustellen, Nutzungsbedingungen während des Datenaustauschs zu prüfen und teilweise auch technisch umzusetzen. Datenkonnektoren abstrahieren somit die Komplexität des technischen Datenaustausches zwischen den Teilnehmenden einer Dateninfrastruktur.

Eine weitere denkbare Kategorie von Zusatzdiensten stellen Orchestratorservices dar. Diese dienen als Koordinatoren, um Aufgaben über verschiedene Dateninfrastrukturdienste zu verteilen und zu aggregieren. Orchestratorservices tragen somit dazu bei, Workflows über die verschiedenen Dienste einer Dateninfrastruktur nahtlos und mit möglichst geringer manueller Interaktion auszuführen.

2.2 Gestaltungsoptionen von Dateninfrastrukturen

Eine grundlegende Entscheidung in der Gestaltung eines Systems ist die Festlegung der Systemarchitektur. Die Systemarchitektur beschreibt die Anordnung funktionaler Elemente, die Realisierung der funktionalen Elemente durch physische Komponenten und die Schnittstellen zwischen den Komponenten. Für Dateninfrastrukturdienste impliziert dies insbesondere die Beantwortung der Fragen „Wie häufig benötige ich einen Dienst?“, „Wer stellt diesen Dienst bereit?“ und „Wie ist die Beziehung zwischen diesen Diensten?“. Die Beantwortung dieser Fragestellungen muss in Abhängigkeit von den Zielstellungen der Dateninfrastruktur und von den vorherrschenden Rahmenbedingungen erfolgen.

Die zuvor beschriebenen Dateninfrastrukturdienste können durch verschiedene grundlegende Gestaltungsoptionen realisiert werden (Schleimer et al., 2023). Als Gestaltungsoption wird eine

spezifische Möglichkeit zur Erreichung der gewünschten Funktionalität verstanden. Der folgende Absatz charakterisiert vier grundlegende Gestaltungsoptionen von Dateninfrastrukturdiensten zur Erreichung der gewünschten Funktionalität.

Vollständig zentralisierte Gestaltung

Bei der zentralisierten Gestaltung eines Dateninfrastrukturdienstes wird dieser genau einmal für die gesamte Dateninfrastruktur bereitgestellt (s. Abbildung 1). Die Einmaligkeit des Dienstes bezieht sich auf die Sicht der Dateninfrastrukturnutzenden. Innerhalb eines Dienstes kann eine für die externen Nutzenden unsichtbare Replikation von Funktionen wie beispielsweise der Datenspeicherung stattfinden, um die Ausfallsicherheit oder die Wiederherstellungsfähigkeit zu erhöhen. Die Bereitstellung des Dienstes kann z.B. durch eine dedizierte neutrale Partei oder, in seltenen Fällen, durch das Unternehmen mit dem größten Interesse an einem Funktionieren der Dateninfrastruktur geleistet werden. Zur Population des Dateninfrastrukturdienstes werden Daten bzw. Metadaten aus den Einflussbereichen der beteiligten Akteure an den Dateninfrastrukturdienst übertragen. Dem Betreibenden des Dateninfrastrukturdienstes wird vertraut, die Daten im Sinne der datenbereitstellenden Teilnehmenden zu verwenden. Dies ist insbesondere für Dienste relevant, die Daten zwischen den Teilnehmenden einer Dateninfrastruktur vermitteln. Beispielsweise sollte ein Katalogdienst im Sinne der Datengebenden neutral agieren und den Nutzenden keine Datenangebote bevorzugt präsentieren.

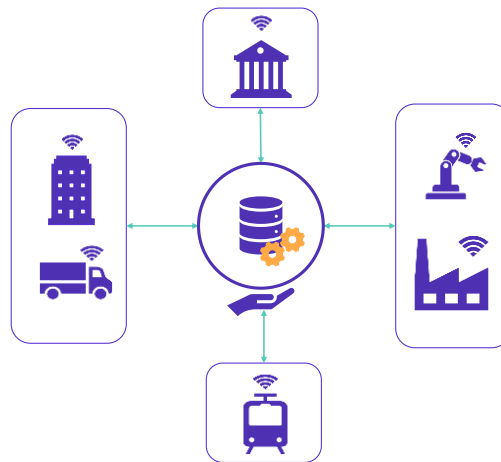


Abbildung 1: Schematische Darstellung zentralisierter Dateninfrastrukturdienste

Beispiel Daten- und Dienstekatalog:

Ein einziger Betreibender stellt den Daten- und Dienstekatalog zur Verfügung. Die Akteure übertragen ihre Metadaten an den zentralen Katalogdienst. Dieser macht die Daten- und Dienstangebote der Akteure transparent und für andere Akteure durchsuchbar. Bei Wünschen nach Aktualisierungen oder Löschungen von Daten- oder Serviceangeboten muss der zentrale Dienst durch die Akteure kontaktiert werden. Ein Beispiel für einen zentralisierten Datenkatalog, der einmalig für eine Dateninfrastruktur implementiert wird, stellt der „Metadata Broker“ im IDSA-Referenzarchitekturmodell für Datenräume⁷ dar. Auch SIMPL sieht einen zentralisierten Katalog vor⁸.

⁷ <https://github.com/International-Data-Spaces-Association/metadata-broker-open-core>

⁸ <https://simpl-programme.ec.europa.eu/book-page/simpl-open-architecture>

Vollständig dezentralisierte Gestaltung

Das vollständig dezentralisierte Design von Dateninfrastrukturdiensten (s. Abbildung 2) impliziert, dass ein Dienst mehrfach für dieselbe Dateninfrastruktur implementiert wird, ohne dass Dienste desselben Typs miteinander kommunizieren können. Dies kann unter anderem der Fall sein, wenn unterschiedliche Anforderungen und Erwartungen hinsichtlich eines Dienstes existieren oder Akteure davon absehen möchten, Daten an einen zentralisierten Dienst zu übertragen. Durch die mehrfache Ausführung des Dienstes sind ebenso verschiedene Bereitstellungsoptionen innerhalb einer Dateninfrastruktur denkbar. Beispielsweise können Betreibende der Dateninfrastruktur den Dateninfrastrukturdienst dediziert als Serviceleistung für einen Teilnehmenden oder für mehrere Teilnehmende als Shared Service bereitstellen. Zusätzlich können auch einzelne Teilnehmende den Dienst selbst für ihre intelligenten Geräte oder Dienste bereitstellen, um sich von Abhängigkeiten zu entkoppeln (s. mittlerer Teil der Abbildung 2). Eine vollständig dezentralisierte Gestaltung bedingt, dass potenziell alle Dateninfrastrukturdienste eines gleichen Typs angefragt werden müssen, um ein vollständiges Bild der Dateninfrastruktur zu erhalten. Daher ist zu erwarten, dass bei dieser Konstellation eine Konkurrenzsituation zwischen Diensten gleichen Typs entstehen wird, bei der es final üblicherweise zu einer Konsolidierung hinzu einigen wenigen Diensten des gleichen Typs kommt (Bader et al., 2023).

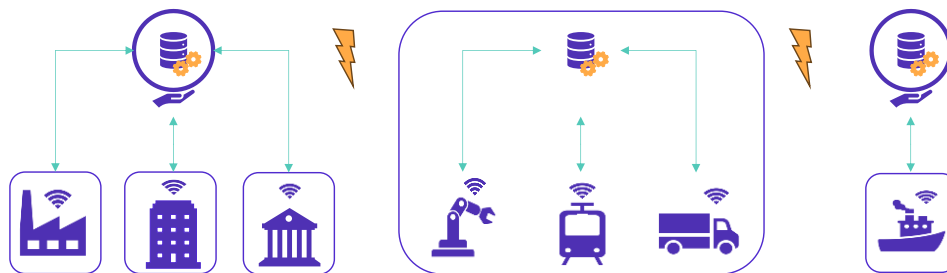


Abbildung 2: Schematische Darstellung dezentralisierter Dateninfrastrukturdienste

Beispiel Daten- und Dienstekatalog:

Mehrere nicht miteinander verbundene Katalogdienste stehen zur Verfügung. Ein Akteur muss sich daher vor der Registrierung seiner Daten- und Dienstangebote entscheiden, ob er sich einem bestehenden Katalogdienst anschließen möchte oder seinen eigenen Katalog (eventuell unter Beauftragung eines Dienstleisters) bereitstellen möchte. Der Akteur überträgt seine Metadaten an den entsprechenden Katalog oder nutzt seinen eigenen Katalog. Potenziell Datennutzende müssen entweder wissen, wo potenziell relevante Datenangebote gefunden werden können, oder eine Vielzahl der verfügbaren Kataloge durchsuchen. Auch im Falle einer dezentralisierten Gestaltung des Katalogdienstes können der IDSA Metadata Broker und der SIMPL-Katalog als Beispiele dienen. Beide Initiativen sehen potenziell die mehrfache, parallele Nutzung des Katalogdienstes in ihren Architekturdokumenten vor.

Föderierte Gestaltung

Die föderierte Umsetzung eines Dateninfrastrukturdienstes (s. Abbildung 3) besteht grundsätzlich aus einer Vielzahl von unabhängigen Dateninfrastrukturdiensten gleichen Typs, die untereinander interoperabel sind und durch den Austausch von (Meta-)Daten miteinander kooperieren. Die Interoperabilität wird durch gemeinsame Standards und Protokolle sichergestellt, die von den einzelnen Diensten unterstützt werden. Jeder Dienst kann autonom entscheiden, welche Metadaten er mit welchem anderen Dienst teilt und selbst festlegen, wie die erhaltenen Metada-

ten dargestellt werden sollen. Entsprechend erhält jeder Dienst eine individuelle Sicht der Dateninfrastruktur. Zusätzlich ist es für einzelne Dienste möglich, flexibel dem System der föderierten Dienste beizutreten oder dieses bei Bedarf wieder zu verlassen. Wie auch in der vollständig dezentralisierten Gestaltung, können die einzelnen Dateninfrastrukturdienste des gleichen Typs von den Akteuren selbst oder von einem durch die Akteure beauftragten Dienstleister betrieben werden. Die umfassendste Selbstbestimmung erhalten die Akteure im Falle eines Eigenbetriebs des Dienstes.

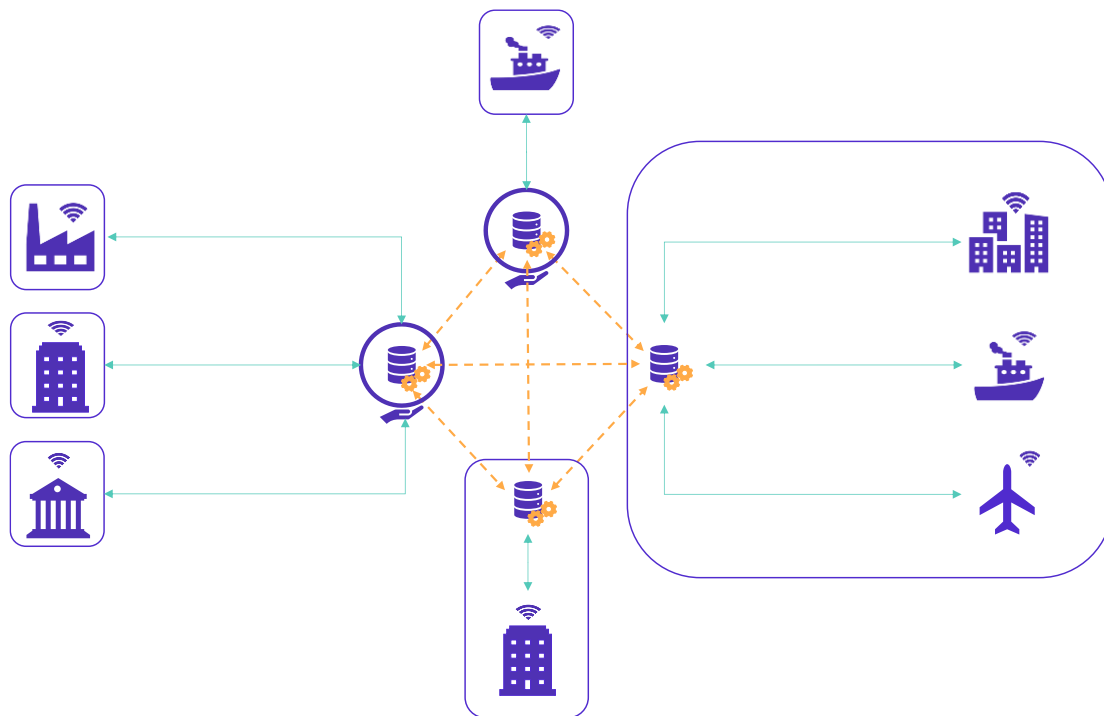


Abbildung 3: Schematische Darstellung föderierter Dateninfrastrukturdienste

Beispiel Daten- und Dienstekatalog:

Ein Akteur setzt einen eigenen Katalogdienst auf, um am System föderierter Kataloge teilzunehmen. Der Katalog muss die festgelegten Standards und Protokolle befolgen. Im Idealfall steht zur einfachen Umsetzung eine offene Referenzimplementierung des Katalogs zur Verfügung. Im ersten Schritt befüllt der Akteur seinen eigenen Katalog mit seinen Daten- und Dienstangeboten. Im nächsten Schritt werden, je nach Synchronisierungsmethode, die Angebote weiterer Kataloge angefragt. Geben diese Kataloge ihr Einverständnis, so können ihre Angebote abgerufen und im eigenen Katalog gezeigt werden. Andersherum können auch die weiteren Kataloge die Angebote des Teilnehmenden abrufen, sofern sie die Zugriffsbedingungen erfüllen. Jeder der föderierten Kataloge erhält durch die Datenreplikation ein individuelles Bild über die verfügbaren Daten- und Serviceangebote. Um ein Entfernen oder eine Aktualisierung der Angebote zu veranlassen, müssen die externen Kataloge benachrichtigt werden und die Anfrage umsetzen. Eine föderierte Umsetzung eines Datenkatalogs ist beispielsweise Teil der Eclipse Dataspace Components (EDC)⁹.

Verteilte Gestaltung

Zur Realisierung einer verteilten Gestaltung von Dateninfrastrukturen existieren verschiedene Konzepte und Subtypen, die sich anhand ihrer Topologie unterscheiden. Die gewählte Topologie beeinflusst die Eigenschaften und das Verhalten der Dateninfrastruktur wie beispielsweise

⁹ <https://eclipse-edc.github.io/>

das Synchronisierungsverhalten der einzelnen Dateninfrastrukturdienste gleichen Typs. Die folgende Tabelle fasst eine Auswahl möglicher Topologien zur Gestaltung verteilter Systeme zusammen.


Topologie	Beschreibung	Vorteile	Nachteile
 Ring	Durch die Verbindung eines Knotens mit zwei anderen Knoten ergibt sich ein geschlossener ringförmiger Datenübertragungspfad. Daten wandern entlang des Pfades, bis sie ihr Ziel erreichen.	<ul style="list-style-type: none"> - Gleichberechtigung der Knoten - Einfache Fehlererkennung 	<ul style="list-style-type: none"> - Ausfallanfälligkeit - Begrenzte Skalierbarkeit
 Linie	Jeder Knoten ist nacheinander mit einem anderen Knoten verkettet, ein Ringschluss wird jedoch nicht vorgenommen, was in offenen Enden resultiert.	<ul style="list-style-type: none"> - Einfache Implementierung - Leichte Erweiterung 	<ul style="list-style-type: none"> - Ausfallanfälligkeit - Datenkollisionen
 Baum	Hierarchische Anordnung einzelner Knoten in Form der Äste eines Baumes mit aufsteigendem Zentralisierungsgrad je Hierarchieebene.	<ul style="list-style-type: none"> - Einfache Verwaltung - Gute Skalierbarkeit 	<ul style="list-style-type: none"> - Hohe Verbindungskosten - Ausfall des zentralen Knotens beeinträchtigt das gesamte Netzwerk
 Stern	Es existiert ein zentraler Knoten der als Orchestrator des verteilten Systems agiert. Alle Knoten sind mit diesem Orchestrator verbunden und tauschen Daten mit ihm aus.	<ul style="list-style-type: none"> - Einfache Fehlersuche - Hohe Ausfallsicherheit - Leichte Erweiterung 	<ul style="list-style-type: none"> - Hohe Verbindungskosten - Ausfall des zentralen Knotens beeinträchtigt das Netzwerk
 Vollständig verbunden	Jeder Knoten ist mit jedem anderen Knoten verbunden. Die Knoten besitzen üblicherweise eine gleichberechtigte Funktion.	<ul style="list-style-type: none"> - Hohe Ausfallsicherheit - Direkte Verbindungen - Hohe Redundanz 	<ul style="list-style-type: none"> - Sehr hohe Verbindungskosten - Komplexe Implementierung und Verwaltung

Tabelle 1: Topologien verteilter Architekturen im Vergleich

Im Rahmen dieser Entscheidungshilfe wird die Ausprägung eines verteilten Systems auf Basis von Distributed Ledger-Technologie (DLT) betrachtet (s. Abbildung 4). Ein Distributed Ledger ist essenziell ein Informationsspeicher, der über die am System beteiligten Knoten geteilt wird. Über einen Konsensmechanismus wird sichergestellt, dass die Informationen über die Knoten hinweg synchron gehalten werden können. Jeder Knoten eines DLT-Systems hält entsprechend zu jedem Zeitpunkt den gleichen Satz an Informationen. In Anlehnung an ein Grundbuch werden Informationen in DLT-Systemen üblicherweise ausschließlich hinzugefügt. Ein DLT-System kann zudem Geschäftslogiken basierend auf sogenannten „Smart Contracts“ umsetzen. DLT besitzen aufgrund ihrer Datenreplikation und Nutzung von Konsensmechanismen besondere Relevanz bei der Herstellung von Transparenz und Vertrauen in Dateninfrastrukturen. Sie werden derzeit insbesondere in den Bereichen der Finanzwirtschaft oder des Lieferkettenmanagements erprobt und eingesetzt.

Zur Umsetzung eines Dateninfrastrukturdienstes basierend auf DLT finden sich Teilnehmende der Dateninfrastruktur, beauftragte Dienstleistende und sonstige Interessensgruppen (z.B. öf-

fentliche Organisationen) zusammen, die jeweils einen Knoten des verteilten Systems betreiben. Der Datenbestand und die Geschäftslogik der Smart Contracts werden durch die beteiligten Knoten repliziert. Möchte zu einem späteren Zeitpunkt ein weiterer Akteur einen Knoten betreiben und sich am System beteiligen, so muss dazu ein Konsens unter den existierenden Knoten gefunden werden. Zur Interaktion mit der Dateninfrastruktur können Akteure jedoch auch den Einstieg über einen bereits vorhandenen Knoten wählen, sofern es von diesem zugelassen wird. Entschließen sich einzelne Beteiligte des DTL-Systems zu einer Abschaltung ihrer Knoten, wird der Dienst von den anderen Knoten aufrechterhalten und die vorhandenen Daten bleiben verfügbar.

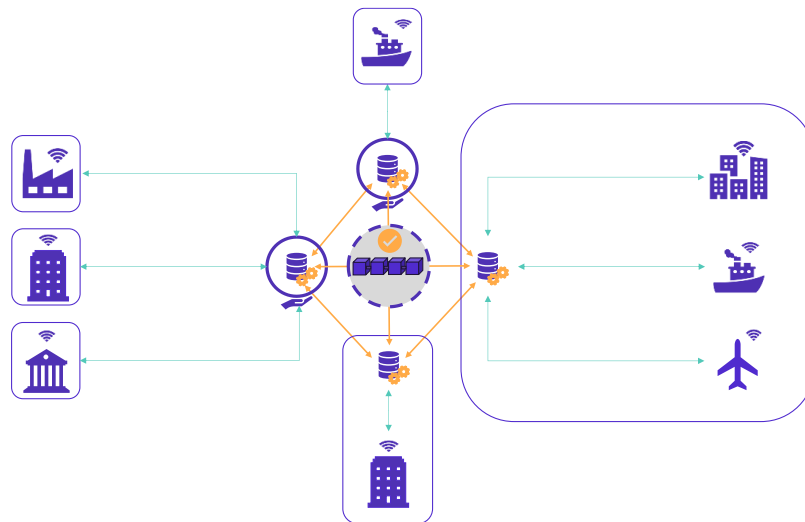


Abbildung 4: Schematische Darstellung verteilter Dateninfrastrukturdienste auf Basis von Distributed Ledger-Technologie

Beispiel Daten- und Dienstekatalog:

Zur Umsetzung des Katalogdienstes schließen sich mehrere Akteure und Dienstleister zusammen, die jeweils einen Knoten des verteilten Systems betreiben. Jeder Knoten ist dabei gleich und verhält sich nach den gemeinsam definierten Regeln. Zur Listung seiner Angebote stellt ein Akteur die Metadaten seiner Daten und Dienste über einen eigenen Knoten oder einen Knoten der anderen Netzwerkteilnehmenden dem verteilten Dateninfrastrukturdienst zur Verfügung. Wird zwischen den Knoten ein Konsens zur Aufnahme erreicht, werden die Angebote durch den verteilten Katalogdienst übernommen. Jeder der Knoten erhält ein synchrones Abbild der gelisteten Angebote. Einmal in den verteilten Ledger geschrieben, können die Angebote nicht wieder entfernt, sondern ausschließlich als inaktiv oder abgelaufen gekennzeichnet werden. Fällt einer der Knoten aus, ist nach wie vor ein aktuelles Abbild in allen anderen Knoten verfügbar. Ein Beispiel für die Umsetzung eines Daten- und Dienstekatalogs auf Basis von DLT ist im Rahmen von Ocean Protocol¹⁰ umgesetzt. Datenangebote werden hier als Non-Fungible Token repräsentiert.

Charakteristika der Gestaltungsoptionen im Vergleich

Eine zusammenfassende Gegenüberstellung der vorgestellten Gestaltungsoptionen wird in der folgenden Tabelle gezeigt. Dienste gleichen Typs können entweder einmalig oder mehrmalig innerhalb einer Dateninfrastruktur vorhanden sein. Die Bereitstellung der Dienste kann entweder durch dedizierte Dienstleistende ohne Interesse an einer Beteiligung am Data Sharing oder die Datengebenden und Datennutzenden selbst erfolgen. Bei der mehrfachen Bereitstellung

¹⁰ <https://docs.oceanprotocol.com/>

von Diensten desselben Typs können die Dienste vollständig voneinander isoliert sein, die Daten vollständig untereinander replizieren oder autonom entscheiden, welche Daten welchem Dienst zur Verfügung gestellt werden.

Gestaltungsoption	Anzahl der Dienste gleichen Typs	Bereitstellung der Dienste	Beziehung zwischen den Diensten gleichen Typs
Vollständig zentralisiert	1	Dienstleistende	-
Vollständig dezentralisiert	>1	Dienstleistende und/oder Datengebende/Datennutzende	Vollständig isoliert
Föderiert	>1	Dienstleistende und/oder Datengebende/Datennutzende	Selbstbestimmtes Data Sharing
Verteilt	>1 (virtuell 1)	Dienstleistende und/oder Datengebende/Datennutzende	Vollständig repliziert

Tabelle 2: Gegenüberstellung der Gestaltungsoptionen

Zusätzlich zu den beschriebenen Architekturoptionen existieren noch weitere Mischformen oder spezifische Merkmalsvariationen, die im Rahmen dieser Entscheidungshilfe aus Komplexitätsgründen nicht erläutert werden. Verschiedene Eigenschaften und Rahmenbedingungen können für oder gegen die Nutzung einer spezifischen Gestaltungsoption sprechen. In jedem Fall sollte für den spezifischen Anwendungsbereich die passende Architekturoption ausgewählt werden, um die geplanten Anwendungen bestmöglich zu unterstützen und möglichst viele Akteure von einer Beteiligung an den geplanten Anwendungen zu überzeugen.

3 Vorgehensmodell zur Auslegung von Dateninfrastrukturdiensten

Das im folgenden vorgestellte Vorgehensmodell unterstützt Unternehmen bei der Auslegung von Dateninfrastrukturdiensten, indem es in aufeinander aufbauenden Phasen die Aktivitäten und Ergebnisse beschreibt, die benötigt werden, um zu einer informierten Entscheidung hinsichtlich des Designs der Dateninfrastruktur zu gelangen. Das Vorgehen geht von der neuen Errichtung einer Dateninfrastruktur zwischen den beteiligten Akteuren des Data Sharing aus (Greenfield-Ansatz). Das finale Ergebnis stellt die Auswahl einer der vorgestellten Designoption zur Realisierung der Dateninfrastrukturdienste dar. Das Vorgehensmodell orientiert sich dazu an etablierten Methoden im Bereich der Unternehmensarchitektur, d.h. der modellhaften Betrachtung eines Unternehmens aus der Sicht der Informationstechnik. Dafür werden Konzepte der Generalized Enterprise Reference Architecture (ISO, 2019) und des Business Motivation Model (Object Management Group, 2015) an die Auslegung von Dateninfrastrukturdiensten adaptiert.

Ausgangspunkt für die Anwendung des Vorgehensmodells ist die vorhandene Spezifikation eines oder mehrerer Geschäftsmodelle, die auf Data Sharing-Mechanismen basieren. So sollten Elemente wie das Wertangebot, die Zielgruppe, die Datenaustauschpartner, die Schlüsselressourcen und mögliche Ertragsmechanismen vor dem Dateninfrastrukturentwurf bekannt sein. Diese Aspekte werden im Folgenden genutzt, um konkrete Rahmenbedingungen und Anforderungen abzuleiten, die ihrerseits die Grundlage für eine reflektierte Auslegungsentscheidungen darstellen.

Abbildung 5 zeigt einen Überblick über den Lebenszyklus von Dateninfrastrukturdiensten. Jeder der Blöcke symbolisiert eine Lebenszyklusphase von der Umweltanalyse bis hin zur Dekommissionierung der Dateninfrastruktur. Die jeweiligen Phasen zeichnen sich durch eine logische Abhängigkeit aus. Vorausgehende Phasen geben einen Auftrag im Sinne von Standards, Anforderungen oder Rahmenbedingungen an die folgenden Phasen. Andersherum erhalten die vorausgehenden Phasen Feedback hinsichtlich der tatsächlichen Umsetzbarkeit, Unterstützbarkeit oder Nützlichkeit. Im Rahmen eines kontinuierlichen Verbesserungsprozess ermöglichen die Feedbackschleifen eine fortlaufende Anpassung der Architektur auf Basis relevanter interner oder externer Veränderungen.

Für die Entscheidungsfindung zur Auslegung der Dateninfrastruktur und ihrer Dienste sind die ersten vier Lebenszyklusphasen relevant:

- Umweltanalyse: In dieser initialen Phase werden die relevanten Stakeholder bzw. Einflussnehmer und deren Erwartungen identifiziert und bewertet.
- Geschäftsmodellanalyse: Diese Phase identifiziert die relevanten Daten und Systeme und skizziert die zur Leistungserbringung notwendigen Datenflüsse basierend auf dem Geschäftsmodell.
- Anforderungsdefinition: Diese Phase leitet aus der Umweltanalyse und Geschäftsmodellanalyse die Anforderungen an die Dateninfrastruktur ab und priorisiert diese.

- Grobdesign: Diese Phase erarbeitet die möglichen Designalternativen zur Umsetzung der Dateninfrastruktur und identifiziert die bestmögliche Alternative basierend auf den zuvor definierten Anforderungen.

Diese vier Phasen werden im Folgenden in dieser Entscheidungshilfe beschrieben. Die an das Grobdesign anschließenden Phasen „Detaildesign“, „Umsetzung“, „Betrieb und Wartung“ sowie „Dekommissionierung“ sind aufgrund ihrer geringen Generalisierbarkeit nicht Teil dieser Entscheidungshilfe.

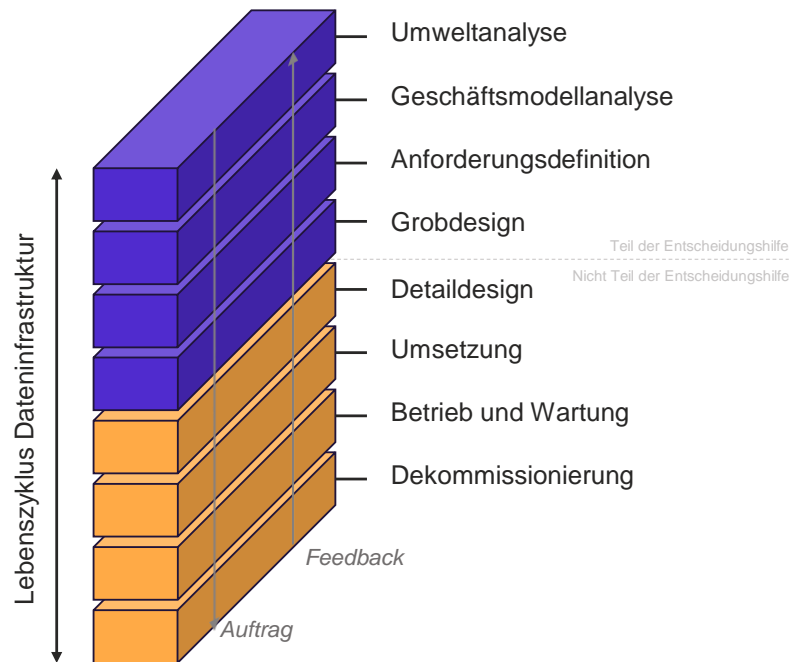


Abbildung 5: Vorgehensmodell zur Entscheidungsfindung für Dateninfrastrukturdienste

3.1 Phase 1 – Umweltanalyse



Jede durch die Dateninfrastruktur umgesetzte Anwendung und damit auch die Dateninfrastruktur selbst ist in einem Umfeld situiert, das bestimmte Erwartungen und Einflüsse an den Entwurf der Dateninfrastruktur richtet. Andersherum verursacht die Dateninfrastruktur selbst Umwelteffekte, die durch direkte Interaktion mit der Umwelt oder indirekt über den Lebenszyklus der Dateninfrastruktur entstehen (z.B. Energieverbräuche). Ziel der Umweltanalyse ist es, die in der Umwelt befindlichen Stakeholder zu identifizieren und Einflussfaktoren auf die Gestaltung der Dateninfrastruktur abzuleiten. Ein Stakeholder ist eine Person, Organisation oder Gruppe, die einen Anspruch oder ein Interesse an der entstehenden Dateninfrastruktur hat und von deren Aktivitäten oder Entscheidungen betroffen ist. Stakeholder versuchen, Einfluss auf die Gestaltung der Dateninfrastruktur zu nehmen. Einflussfaktoren umfassen grundsätzlich alle Aspekte, die ohne die unmittelbare Ausübung von Zwang oder Macht eine Einwirkung auf die Gestaltung der Dateninfrastruktur ausüben können (Object Management Group, 2015). Einflussfaktoren können beispielsweise technologische, geschäftliche, organisatorische, politische, regulatorische, ökologische oder soziale Hintergründe haben.

Eine Identifizierung und Skizzierung der Umwelt ist in vielerlei Hinsicht relevant. Durch die Umweltanalyse lassen sich übergeordnete gesellschaftliche Erwartungen an die Lösung ableiten, die ihrerseits die Grundlage für die die Einflussnahme verschiedener Stakeholder darstellen. Insbesondere bei der Entwicklung von Lösungen, die eine breite Tragweite besitzen und eine Vielzahl von Akteuren betreffen, treten Faktoren der langfristigen gesellschaftlichen und ökologischen Verantwortung stärker in den Vordergrund. Neben den gesamtgesellschaftlichen Erwartungen identifiziert die Umweltanalyse konkrete Stakeholdergruppen, welche die Entscheidungsfindung in eine bestimmte, von der Optimallösung abweichende Richtung lenken können. Zu diesen Gruppen gehören beispielsweise NGOs.

Im Rahmen der Umweltanalyse sind drei aufeinander aufbauende Themen zu bearbeiten. Startpunkt ist die Identifizierung relevanter Stakeholder. Anschließend werden deren Erwartungen und Einflüsse beschrieben. Abschließend findet eine Einschätzung der Auswirkungen dieser Erwartungen statt.

Zu den Stakeholdern einer Dateninfrastruktur gehören sowohl Akteure, die einen Einfluss auf die Dateninfrastruktur nehmen, als auch solche, die in ihren Handlungen durch die Gestaltung der Dateninfrastruktur beeinflusst werden. Im Rahmen der Umweltanalyse werden insbesondere externe Stakeholder fokussiert. Die externe Umwelt wird üblicherweise gemäß des englischen Akronyms PESTEL in die Bereiche Politik (Politics), Wirtschaft (Economy), Soziales (Social), Technologie (Technology), Ökologie (Environment) und Recht (Legal) unterteilt.

Die Interessen eines Stakeholders äußern sich in Form von Ansprüchen gegenüber der Entwicklung der Dateninfrastruktur. Diese Ansprüche entstammen aus der Perspektive des Stakeholders und fundieren auf seinem eigenen Wissen und seiner Erfahrung, seiner Verantwortung und seinem Auftrag. Die Ansprüche können allgemeiner Natur sein oder sich explizit auf Designaspekte der Dateninfrastruktur wie beispielsweise die genutzten Standards der Dienste beziehen. Ein Stakeholder kann einen oder mehrere Ansprüche besitzen. Ebenso kann derselbe Anspruch durch verschiedene Stakeholder geltend gemacht werden. Die identifizierten Stakeholder, deren Ansprüche und mögliche Einflüsse sind nicht nur zu benennen, sondern auch für weitere Beteiligte nachvollziehbar zu beschreiben. Bei der Beschreibung sollte ein möglichst neutraler Ton gewählt werden, der eine spätere Bewertung der Relevanz einzelner Stakeholder und ihrer Erwartungen zulässt. Die identifizierten Stakeholder, Erwartungen und Einflüsse können in Tabellenform zusammengefasst werden. Eine Übersicht über mögliche Stakeholder und deren Erwartungen und Einflüsse beim Aufbau von Dateninfrastrukturen ist in der folgenden Tabelle zusammengefasst.

Stakeholder	Beispielhafte Interessen, Ansprüche und Einflüsse
Regulatorik	Einhaltung regulatorischer Pflichten insbesondere im Bereich des Datenrechts, Kartellrechts und der Produkthaftung
Nutzende	Langfristige Verfügbarkeit der Lösung; Nutzung von etablierten Standards und Richtlinien; Erreichen relevanter Zertifizierungen
Wettbewerber	Benchmarking der entwickelten Lösung; Interoperabilität und Standardisierung; Coopetition
Technologiepartner	Nutzung auf dem Markt verfügbarer Ressourcen, Technologien und Serviceangebote; Berücksichtigung aktueller Trends und Ermöglichung zukünftiger Weiterentwicklungen
Gesellschafterinnen und Gesellschafter	Erhalt der Wettbewerbsfähigkeit; Wachstum des Unternehmens; Erreichen positiver sozialer und Nachhaltigkeitskennzahlen
Nichtregierungsorganisationen (NRO)	Reduktion von Umweltwirkungen digitaler Infrastrukturen; Gewährleistung von digitalen Rechten und Privatsphäre; Umsetzung fairer Geschäftspraktiken; Erfüllung zusätzlicher domänenspezifischer Belange (beispielsweise hinsichtlich Verbraucher- und Patientenschutz, Arbeitnehmerrechten oder Industrieinteressen)
Forschung und Entwicklung	Einblick in genutzte Lösungen für Forschende; Nutzung neuer Forschungserkenntnisse im Rahmen der Umsetzung der Dateninfrastruktur

Tabelle 3: Stakeholder und deren beispielhafte Erwartungen und Einflüsse im Kontext von Dateninfrastrukturen

Im Rahmen der abschließenden Beurteilung wird eingeschätzt, inwiefern der Stakeholder einen legitimen Einfluss auf die Gestaltung der Dateninfrastruktur nehmen kann. Daraus folgt essenziell, ob der durch den Stakeholder geltend gemachte Anspruch oder Einfluss bei der weiteren Gestaltung der Dateninfrastruktur berücksichtigt werden muss. Die Beurteilung ergibt sich dabei aus der Rolle eines Stakeholders sowie aus seinem finanziellen, sozialen oder rechtlichen Status. Eine Beurteilung kann beispielsweise im Sinne von Risiken oder Chancen erfolgen (Object Management Group, 2015). Nach dieser Vorgehensweise werden die finanziellen Chancen und Risiken bei einer Beachtung oder Missachtung der Interessen einzelner Stakeholder gegeneinander abgewogen und eine Priorisierung der Ansprüche aufgrund monetärer Kriterien erstellt.

3.2 Phase 2 – Geschäftsmodellanalyse



Der Aufbau einer Dateninfrastruktur erfolgt grundsätzlich zur Unterstützung einer oder mehrerer Anwendungen und den zugrundeliegenden Geschäftsmodellen. Entsprechend müssen die Anforderungen einer Dateninfrastruktur klar aus den Geschäftsmodellen hervorgehen. Dazu ist es notwendig, ein grundlegendes Geschäftsverständnis aufzubauen: Zuerst muss die Frage beantwortet werden, was genau getan werden muss, um daraus abzuleiten, wie etwas getan werden kann.

Die Identifizierung der Charakteristika der Geschäftsmodelle ist relevant, um die Anforderungen, die sich aus dem Geschäftsmodell ergeben, durch die Technologie abzubilden und zu vermeiden, dass eine Diskrepanz entsteht, die den Erfolg der geplanten Anwendungen behindert. Bei-

spielsweise kann das Geschäftsmodell einen Aufschluss über erwartete Gewinne und Einsparungen geben. Diese monetären Kennzahlen haben anschließend Einfluss darauf, ob die Kosten einer Dateninfrastruktur einen relevanten Einflussfaktor darstellen. Der Zusammenhang zwischen den Eigenschaften des Geschäftsmodells und den Anforderungen an die Dateninfrastruktur ist in Abbildung 6 visualisiert.

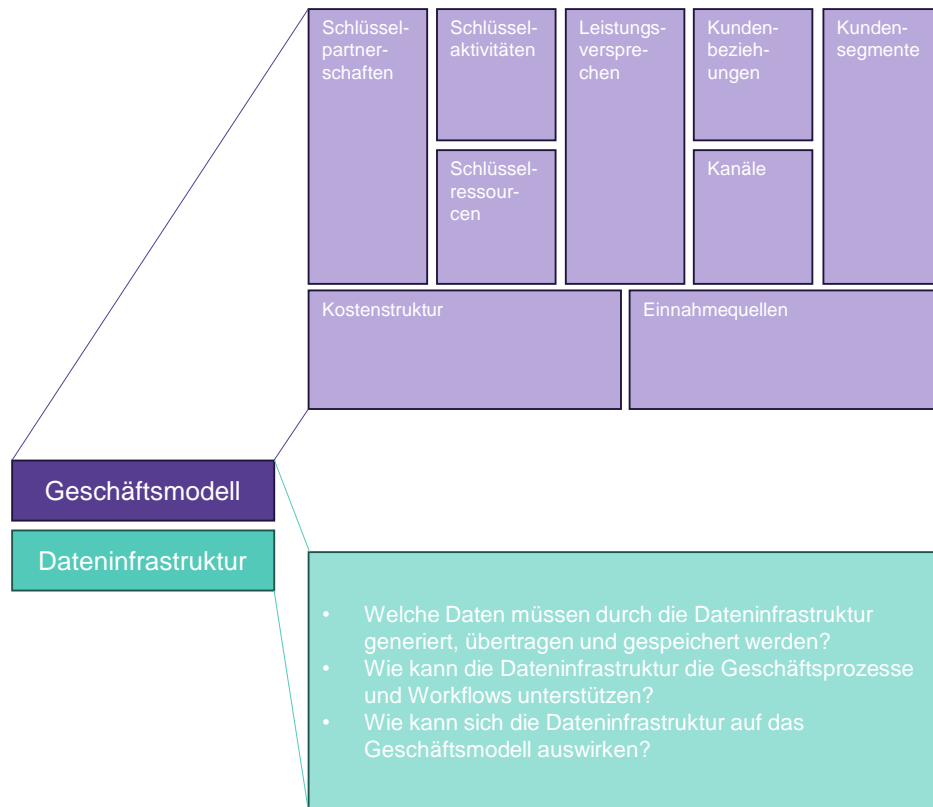


Abbildung 6: Zusammenhang zwischen Geschäftsmodell und Dateninfrastruktur in Anlehnung an Osterwalder et al. (2010)

Ziel der Phase „Geschäftsmodellanalyse“ ist daher die Ableitung relevanter Charakteristika der Geschäftsmodelle, die einen Einfluss auf die Gestaltung der Dateninfrastruktur besitzen. Dazu werden zunächst die Kernkomponenten des Geschäftsmodells identifiziert, anschließend die benötigten Daten als Wertschöpfungsressource tiefergehend charakterisiert und die Datenflüsse zwischen Quell- und Zielsystemen beschrieben.

Zur Beschreibung des Geschäftsmodells bieten sich etablierte Instrumente wie das „Business Model Canvas“ an (Osterwalder et al., 2010). Das Canvas hilft dabei, die wesentlichen Bausteine eines Geschäftsmodells zu identifizieren, zu strukturieren und ein gemeinsames Verständnis über die Bausteine zu generieren. Basierend auf dem Business Model Canvas sind in den letzten Jahren weitere ähnliche Hilfsmittel mit gleichem Konzept entstanden. Beispielhaft ist hier das „Data Cooperation Canvas“ (DS4SSCC, 2023) zu nennen, welches die generellen Konzepte von Geschäftsmodellen aufgreift und um spezifische Aspekte des Data Sharing wie die genutzten Daten und Interoperabilitätsstandards erweitert.

Unabhängig vom Einsatz des spezifischen Geschäftsmodellanalyse-Tools gilt es, Daten als zentraler strategischer Ressource in Data Sharing Geschäftsmodellen besondere Aufmerksamkeit zu schenken, um Einflüsse auf die Dateninfrastruktur ableiten zu können. Beispielsweise

geben Aktualisierungsintervalle der Daten Hinweise darauf, ob eine Echtzeitfähigkeit der Dateninfrastruktur benötigt wird. Folgende Charakteristika mit Bezug zu Daten sind daher genauer zu beschreiben:

- Metriken und Resultate, die durch Datenanalysen oder die Zusammenführung erzielt werden sollen
- Benötigte Daten und deren Datenhaltende (eigene Daten, Daten aus Geschäftsbeziehungen, offene Daten, gekaufte Daten)
- Relevante Datenquellen (datenhaltende Systeme) und Datensinken (beispielsweise die zu nutzenden Analysedienste)
- Sonstige Charakteristika der benötigten Daten (Frequenz, Qualität, Sensitivität, Formate, weitere Metadaten)

Abbildung 7 zeigt ein Template zur Beschreibung von Datenressourcen. Das Feld Datenliefernde beschreibt einzelne natürliche oder rechtliche Personen oder deren Gruppen, die Daten zur Realisierung von Mehrwertdiensten bereitstellen. Diese charakterisieren sich insbesondere durch ihre Verhandlungsstärke, die sich beispielsweise aus der Wichtigkeit ihrer Daten für den Anwendungsfall oder den rechtlichen Rahmenbedingungen, wie beispielsweise einer Pflicht zur Datenweitergabe, ergibt. Die Verhandlungsstärke der Datenliefernden stellt ein wesentliches Merkmal dar, um die Relevanz der Einflüsse bewerten zu können. Die Charakteristika der bereitgestellten Daten fassen Eigenschaften der Daten zusammen, aus denen sich Anforderungen an die Dateninfrastruktur ableiten lassen. Unter den Anreizen sind mögliche Entlohnungen für die Datenliefernden aufzuführen. Zudem wird eine Abschätzung vorgenommen, wie hoch der Wert des Anreizes durch die Datenliefernden eingeschätzt wird.

Datenliefernde							
Name	...						
Typ	Natürliche Person			Organisation			
Verhandlungsstärke	Stark		Mittel		Niedrig		
Charakteristika der bereitgestellten Daten							
Daten	...						
Datenquellen	...						
Personenbezug	Ja			Nein			
Datentyp	Strukturiert	Semi-strukturiert	Stream	Text	Audio	Bild	Video
Aktualisierung der Daten	Echtzeit		Intervall		Einmaliger Transfer		
Datenqualität	...						
Anreize für die Bereitstellung von Daten							
Anreiz	Geld	Virtuelle Werte	Daten	Dienstnutzung	Image	Rechtliche Compliance	Nichts
Anreizwert	Hoch		Mittel		Niedrig		

Abbildung 7: Template zur Charakterisierung relevanter Daten

Aus dem Geschäftsmodell können zudem weitere die Dateninfrastruktur betreffende Rahmenbedingungen abgeleitet werden. Dazu gehören beispielsweise die Anzahl und Charakteristika der zukünftigen Nutzenden, die bereits zuvor genannten Budgetbeschränkungen, die möglichen Technologiepartner und die Entlohnungsmechanismen für die Datenbereitstellung.

Sind die relevanten Daten, deren Quellen und die Datensenzen definiert, lassen sich diese im Rahmen von Datenflussdiagrammen visualisieren. Datenflussdiagramme unterstützen den Systementwurf, die Optimierung und die Dokumentation. Durch ihre visuelle Form unterstützen Datenflussdiagramme die Kommunikation zwischen einzelnen Beteiligten und machen die wichtigen Systemgrenzen sichtbar. Datenflussdiagramme zeigen zudem auf, in welchem Rahmen welche Unterstützungsleistungen durch die Dateninfrastruktur benötigt werden. Eine klassische Methodik zur Beschreibung von Datenflüssen stellt die Notation nach Gane & Sarson (1979) dar. Relevante Elemente dieser Notation für die Visualisierung von Datenflüssen in Data Sharing-Anwendungen sind in Abbildung 8 gezeigt.

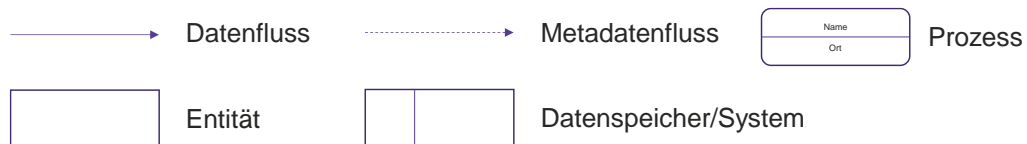


Abbildung 8: Auf Data Sharing-Anwendungen adaptierte Datenflussnotation nach Gane & Sarson (1979)

3.3 Phase 3 – Anforderungsdefinition



Basierend auf der vorausgegangenen Charakterisierung der Umwelt, des Geschäftsmodells und der Datenprozesse werden in dieser Phase die Anforderungen an die Dateninfrastruktur abgeleitet und bewertet. Dieses Vorgehen ermöglicht die Rückverfolgbarkeit der definierten Anforderungen auf bestimmte Rahmenbedingungen und liefert auch zu späteren Zeitpunkten Begründungen, warum eine spezifische Anforderung formuliert wurde. Darüber hinaus bietet dieses Vorgehen die Möglichkeit, bei sich ändernden Einflüssen, wie z.B. regulatorischen Rahmenbedingungen oder Anpassungen des Geschäftsmodells, die Anforderungsstrukturen kontrolliert anzupassen.

Das Ergebnis dieser Phase stellt ein Anforderungskatalog dar. Der Anforderungskatalog umfasst neben relevanten Entscheidungskriterien auch eine Priorisierung der Faktoren. Dazu werden im ersten Schritt die Anforderungen identifiziert. Anschließend wird die Bewertung der Faktoren vorgenommen.

Die Anforderungen an Dateninfrastrukturen lassen sich, wie allgemein üblich, in funktionale und nicht-funktionale Anforderungen unterteilen. Funktionale Anforderungen beschreiben das gewünschte Verhalten des Systems. Das Systemverhalten wird durch Dienste, Aufgaben oder Funktionen realisiert. Funktionale Anforderungen beschreiben das „Was“ der Dateninfrastruktur. Nicht-funktionale Anforderungen sind Kriterien, die die Qualität und die Eigenschaften eines Systems beschreiben, ohne sich auf spezifische Funktionen oder Verhaltensweisen zu beziehen. Sie beschreiben das „Wie“ der Dateninfrastruktur. Während sich die funktionalen Anforderungen somit auf die Funktionen der umzusetzenden Dateninfrastrukturdienste beziehen, entsprechen die nicht-funktionalen Anforderungen den für die Auswahl der Designoption relevanten Qualitäts- und Entscheidungskriterien. Entsprechend stehen die nicht-funktionalen Anforderungen im Fokus dieser Phase.

Zentrales Element dieser Entscheidungshilfe ist ein detaillierter Kriterienkatalog, der relevante Entscheidungskriterien erläutert und darstellt, unter welchen Umständen eine besondere Berücksichtigung eines Kriteriums ratsam ist und welche Zusammenhänge zwischen einzelnen

Kriterien bestehen. Der Kriterienkatalog findet sich in Abschnitt 4¹¹. Abbildung 9 zeigt eine zusammenfassende Übersicht relevanter Entscheidungskriterien für Dateninfrastrukturen. Die Entscheidungskriterien für die Umsetzung einer Dateninfrastruktur können wirtschaftlicher, organisatorischer oder technischer Natur sein. Wirtschaftlich motivierte Entscheidungskriterien ergeben sich aus den finanziellen Rahmenbedingungen der zu unterstützenden Geschäftsmodelle. Organisatorische Entscheidungskriterien beschreiben Entscheidungskriterien, die unter anderem durch organisatorische Maßnahmen gestützt werden. Die technischen Entscheidungskriterien ergeben sich aus der Art und Weise, wie eine Leistung technisch erbracht werden soll.

Organisatorisch			Technisch		
Vertrauenswürdigkeit	Datensouveränität	Transparenz	Portabilität	Generativität	Orchestrierung
Auditierbarkeit	Fairness	Vertraulichkeit	Skalierbarkeit	Datenintegrität und Datenverifizierbarkeit	Wiederherstellung
Autonomie	Rechtssicherheit	Interne Interoperabilität	Ressourcenintensität	Zuverlässigkeit	Informationssicherheit und Cybersicherheit
Wirtschaftlich			Wirtschaftlich		
Komplexität	Strategischer Fit	Externe Interoperabilität	Nachhaltigkeit	Operative Effizienz	Kostenamortisation

Abbildung 9: Entscheidungsfaktoren für Dateninfrastrukturen

Aus dieser Kriterienübersicht können auf Basis der zuvor erfolgten Umwelt- und Geschäftsmodellanalyse die relevanten Entscheidungskriterien identifiziert werden. Dazu ist im ersten Schritt zu prüfen, ob die Kriterien für den Erfolg der Dateninfrastrukturinitiative im Rückblick auf die geleisteten Vorarbeiten eine Relevanz besitzen. Hierzu bietet sich das Format sogenannter „Joint Requirements Planning Sessions“ an. Dabei handelt es sich um strukturierte Workshops, bei denen die relevanten Stakeholder zusammenkommen, um die Anforderungen für ein Projekt zu ermitteln und zu planen. Im gemeinsamen Format werden die in der Übersicht gelisteten Entscheidungsfaktoren konsolidiert und bei Bedarf ergänzt. Die Entscheidungsfaktoren sollten dabei auf konkrete Umweltbedingungen oder Erfordernisse der Geschäftsmodelle zurückgeführt und die Begründung dokumentiert werden, um langfristige Transparenz zu gewährleisten.

¹¹ Zudem ist eine editierbare Version des Kriterienkatalogs zur Durchführung einer Nutzwertanalyse für die Auswahl der Dateninfrastruktur-Designoption online unter <https://doi.org/10.5281/zenodo.15053526> verfügbar.

Durch die kooperative Einbeziehung der Stakeholder in die Identifikation der Entscheidungskriterien werden konfligierende oder redundante Anforderungen vermieden. Die folgende Tabelle zeigt einen individualisierten Kriterienkatalog als beispielhaftes Ergebnis dieses Schritts.

Kategorie	Entscheidungsfaktor	Beschreibung
Wirtschaftlich	Kostenamortisation	Die Kosten einer Dateninfrastruktur beziehen sich auf die gesamten finanziellen Aufwendungen, die für den Aufbau, die Einrichtung, den Betrieb und die Wartung der Dateninfrastruktur erforderlich sind. Im Rahmen der Amortisationsrechnung wird geprüft, ob und zu welchem Zeitpunkt die Kosten durch die Gewinne und Abschreibungen refinanziert werden.
Organisatorisch	Datensouveränität	Datensouveränität beschreibt die Fähigkeit eines Datenbereitstellenden, selbst über die Nutzung seiner Datenbestände (einschließlich der Metadaten (Bastiaansen et al. 2019)) während ihres gesamten Lebenszyklus und einschließlich der von Dritten durchgeführten Aktionen zu bestimmen (Scherenberg et al. 2024; Hellmeier und Scherenberg 2023). Anforderungen an die Datensouveränität werden üblicherweise in Verträgen oder Richtlinien formuliert und können derzeit nur zum Teil technisch durchgesetzt werden.
Organisatorisch	Vertrauenswürdigkeit	Vertrauen bezieht sich auf die Bereitschaft einer Partei, sich den Handlungen einer anderen Partei auszusetzen, unabhängig von der Möglichkeit, diese andere Partei zu überwachen oder zu kontrollieren. Grundlage ist dabei, dass die andere Partei eine für den Vertrauensgebenden wichtige Handlung ausführt (Mayer et al. 1995). Im Zusammenhang mit der gemeinsamen Nutzung von Daten betrifft das Vertrauen die Handlungen der Partner bei der gemeinsamen Nutzung von Daten, die Eigenschaften der Daten selbst und das Verhalten der Dateninfrastruktur.
Technisch	Ressourcenintensität	Die Ressourcenintensität einer Dateninfrastruktur umfasst beispielsweise die Nutzung von Speicher- und Rechenressourcen, die wiederum den Energieverbrauch der Lösung beeinflussen.
...

Tabelle 4: Beispielhafter Kriterienkatalog

Üblicherweise haben nicht alle der aufgenommenen Anforderungen die gleiche Relevanz für eine erfolgreiche Umsetzung der Dateninfrastruktur. Entsprechend sind die Anforderungskriterien zur Vorbereitung einer informierten Entscheidungsfindung zu priorisieren. Hierbei kann auf herkömmliche Priorisierungsmethoden aus dem Bereich der mehrkriteriellen Entscheidungsfindung zurückgegriffen werden. Häufig genutzte Priorisierungsmethoden sind beispielsweise der Analytical Hierarchy Process, der Hundert-Punkte-Test, die Nutzwertanalyse oder eine einfache Bewertung nach den Kriterien Hoch-Mittel-Niedrig. Während die drei erstgenannten Methoden auf Zahlenwerten basieren und eine Grundgesamtheit (100 Prozent der Anforderungspunkte) mathematisch auf die Entscheidungsfaktoren aufteilen, müssen bei einer Bewertung nach dem Schema Hoch-Mittel-Niedrig zunächst die zugehörigen Zahlenwerte für die einzelnen Faktoren definiert werden. Für jedes Entscheidungskriterium ist zudem festzulegen, ob sich dieses positiv oder negativ auf die Auswahlentscheidung auswirkt. Beispielsweise kann eine hohe Transparenz einer Dateninfrastruktur in einigen Fällen gewollt sein, in anderen jedoch nicht. Die folgende Tabelle visualisiert das beispielhafte Ergebnis einer Priorisierungsübung.

Kategorie	Entscheidungsfaktor	Beschreibung	Auslegung	Priorität
Wirtschaftlich	Kostenamortisation	...	Positiv	3
Organisatorisch	Datensouveränität	...	Positiv	2
Organisatorisch	Vertrauenswürdigkeit	...	Positiv	2
Technisch	Ressourcenintensität	...	Negativ	3
...

Tabelle 5: Beispielhafter gewichteter Kriterienkatalog

3.4 Phase 4 – Grobdesign



Auf Grundlage des individuellen Kriterienkatalogs wird in dieser Phase die passende Designoption der Dateninfrastruktur herausgestellt. Die Auswahl der Designoption stellt die wesentliche Ausgangslage für die anschließende Feinplanung und Umsetzung der Dateninfrastruktur dar. Zwei Schritte sind zur Auswahl der Designoption notwendig. Zunächst werden die Architekturen identifiziert, die eine Wertschöpfung im Rahmen der geplanten Anwendungen überhaupt möglich machen. Anschließend werden diese Architekturen gegenübergestellt und anhand der zuvor definierten und gewichteten Kriterien bewertet.

Eine Vorauswahl möglicher Designoptionen der Dateninfrastrukturdienste erfolgt auf Basis der Wertschöpfungsmöglichkeit. Dabei wird abgeschätzt, ob basierend auf den inhärenten Eigenschaften einer Designoption deren Einsatz zur Umsetzung der angedachten Anwendungen überhaupt möglich ist. Beispielsweise können rechtliche Anforderungen die Übertragung von Daten an einen zentralen oder vollständig verteilten Dienst ausschließen. Ein zentralisiertes beziehungsweise vollständig verteiltes Design wäre in diesem Falle im weiteren Auswahlprozess nicht mehr zu betrachten. Die in diesem Zusammenhang getroffene Vorselektion reduziert den Entscheidungsspielraum und damit die Komplexität der finalen Bewertung.

Zur finalen Bewertung der Gestaltungsoptionen werden diese entlang der Anforderungskriterien gegenübergestellt und evaluiert. Eine Evaluation sollte dabei in dem bereits in der Anforderungspriorisierung genutzten Evaluationssystem erfolgen, um Verständnisprobleme und damit den Aufwand für die beteiligten Stakeholder zu reduzieren. Auf Basis der Priorisierungsergebnisse wird abschließend die finale Rangfolge der Designoptionen gebildet. Die folgende Tabelle zeigt einen beispielhaften Entscheidungskatalog inklusive Evaluation. Eine erste Indikation, in welchem Umfang mögliche Designoptionen die Entscheidungskriterien für Dateninfrastrukturdienste erfüllen, ist dem im nächsten Abschnitt dargestellten Kriterienkatalog entnehmbar.

Kategorie	Entscheidungsfaktor	Beschreibung	Auslegung	Priorität	Zentralisiert	Verteilt
Wirtschaftlich	Kostenamortisation	...	Positiv	3	3	1
Organisatorisch	Datensouveränität	...	Positiv	2	0	1
Organisatorisch	Vertrauenswürdigkeit	...	Positiv	2	1	3
Technisch	Ressourcenintensität	...	Negativ	3	2	1
...
Gesamt					17	22

Tabelle 6: Beispielhafter Kriterienkatalog inklusive Evaluation

4 Kriterienkatalog für Dateninfrastrukturdienste

Dieser Kriterienkatalog listet relevante Entscheidungsfaktoren bei der Auswahl einer Designoption für Dateninfrastrukturen auf und bewertet sie. Der Kriterienkatalog liefert eine Arbeitshilfe in den Phasen der Anforderungsdefinition und des Grobdesigns von Dateninfrastrukturen. Im Folgenden sind die relevanten Entscheidungsfaktoren bei der Auswahl einer Dateninfrastruktur genauer beschrieben. Jeder Einflussfaktor ist durch ein individuelles Canvas dargestellt. Das Canvas beschreibt folgende Informationen:

- Kategorie des Entscheidungsfaktors: Vorrangige Motivation des Entscheidungsfaktors (wirtschaftlich, technisch, organisatorisch)
- Bezeichnung des Entscheidungsfaktors
- Beschreibung des Entscheidungsfaktors
- Relevanz des Entscheidungsfaktors: Beschreibung, für welche Anwendungen oder Rahmenbedingungen eine genauere Betrachtung des Entscheidungsfaktors üblicherweise notwendig ist
- Erfüllung des Entscheidungsfaktors durch die Dateninfrastruktur-Designoption: Beschreibung und Begründung, inwiefern der Entscheidungsfaktor durch die Gestaltungsoption der Dateninfrastruktur realisiert werden kann
- Wichtige Interdependenzen (falls vorhanden): Andere Entscheidungsfaktoren, die einen Einfluss auf den betrachteten Entscheidungsfaktor haben

Eine bearbeitbare Version dieses Kriterienkatalogs inklusive einer Vorlage zur Durchführung einer Nutwertanalyse zur Auswahl der Gestaltungsoption steht zusätzlich in Form eines Tabellendokuments zur Verfügung¹². Der Kriterienkatalog wurde basierend auf Literaturrecherchen erstellt und im Rahmen von Experteninterviews validiert, vertieft und angepasst.

Wirtschaftlich		Kostenamortisation					
Die Kosten einer Dateninfrastruktur beziehen sich auf die gesamten finanziellen Aufwendungen, die für den Aufbau, die Einrichtung, den Betrieb und die Wartung der Dateninfrastruktur erforderlich sind. Im Rahmen der Amortisationsrechnung wird geprüft, ob und zu welchem Zeitpunkt die Kosten durch die Gewinne und Abschreibungen refinanziert werden.							
Relevanz	Die Betrachtung der Kostenamortisation einer Dateninfrastruktur ist insbesondere relevant, wenn kommerzielle Anwendungen mit geringen Margen durch diese ermöglicht werden sollen.						
Erfüllung durch Dateninfrastruktur-Designoption							
Zentralisiert	+	Föderiert	o	Dezentralisiert	o	Verteilt	o
Zentralisierte Dateninfrastrukturen verursachen üblicherweise geringere Gesamtkosten als dezentralisierte, föderierte oder verteilte Systeme, in denen ein Netz aus mehreren Knoten aufgebaut werden muss.							
Wichtige Interdependenzen							
Operative Effizienz		↗		Kostenamortisation			
Komplexität		↘		Kostenamortisation			
Ressourcenintensität		↗		Kostenamortisation			

¹² Die digitale Version des Kriterienkatalogs ist online unter <https://doi.org/10.5281/zenodo.15053526> verfügbar.

Wirtschaftlich		Operative Effizienz					
Die operative Effizienz bezieht sich auf die Leistung, Geschwindigkeit und Effizienz, mit der Transaktionen von der Dateninfrastruktur ausgeführt werden können. Operative Effizienz wird beispielsweise durch Automatisierung gefördert.							
Relevanz	Operative Effizienz spielt insbesondere bei Data Sharing-Anwendungen eine Rolle, in denen Daten in kurzen Zeitabständen über verschiedene Parteien ausgetauscht werden müssen und sich kontinuierlich verändern.						
Erfüllung durch Dateninfrastruktur-Designoption							
Zentralisiert	+	Föderiert	○	Dezentralisiert	-	Verteilt	-
Zentralisierte Infrastrukturen reduzieren den technischen Overhead einer Datenreplikation und unterstützen die schnelle Automatisierung von Prozessen.							
Interdependenzen							
Skalierbarkeit			↗				Operative Effizienz
Komplexität			↗				Operative Effizienz
Orchestrierbarkeit			↗				Operative Effizienz

Wirtschaftlich		Nachhaltigkeit					
Die Nachhaltigkeit einer Dateninfrastruktur bezieht sich auf die Möglichkeit, diese langfristig und ohne die Beeinträchtigung wirtschaftlicher oder sozialer Aspekte nutzen zu können (Tan et al., 2024).							
Relevanz	Die wirtschaftliche und soziale Nachhaltigkeit ist insbesondere für Anwendungen mit sozialer und gesellschaftlicher Wirkung sowie für Dateninfrastrukturen mit langfristiger Lebensdauer relevant.						
Erfüllung durch Dateninfrastruktur-Designoption							
Zentralisiert	-	Föderiert	○	Dezentralisiert	○	Verteilt	+
Durch die Unabhängigkeit von einzelnen Akteuren kann eine verteilte Infrastruktur grundsätzlich langfristig unabhängig betrieben werden und vermeidet beispielsweise Lock-in-Effekte.							
Interdependenzen							
Portabilität			↗				Nachhaltigkeit
Fairness			↗				Nachhaltigkeit
Generativität			↗				Nachhaltigkeit
Komplexität			↘				Nachhaltigkeit

Organisatorisch		Vertrauenswürdigkeit					
Vertrauen bezieht sich auf die Bereitschaft einer Partei, sich den Handlungen einer anderen Partei auszusetzen, unabhängig von der Möglichkeit, diese andere Partei zu überwachen oder zu kontrollieren. Grundlage ist dabei, dass die andere Partei eine für den Vertrauensgeber wichtige Handlung ausführt (Mayer et al., 1995). Im Zusammenhang mit der gemeinsamen Nutzung von Daten betrifft das Vertrauen die Handlungen der Partner bei der gemeinsamen Nutzung von Daten, die Eigenschaften der Daten selbst und das Verhalten der Dateninfrastruktur.							
Relevanz	Vertrauenswürdigkeit ist für nahezu alle Anwendungen ein relevanter Faktor. Insbesondere spielt diese jedoch bei der Interaktion mit bislang unbekanntem Akteuren eine Rolle.						
Erfüllung durch Dateninfrastruktur-Designoption							
Zentralisiert	○	Föderiert	+	Dezentralisiert	+	Verteilt	+
Die Möglichkeit, selbst einen Teil des Systems bereitzustellen und unabhängig von einer zentralen Instanz zu sein, erhöht das Vertrauen in die Dateninfrastruktur.							
Interdependenzen							
Zuverlässigkeit			↗				Vertrauenswürdigkeit
Autonomie			↗				Vertrauenswürdigkeit
Auditierbarkeit			↗				Vertrauenswürdigkeit

Organisatorisch		Datensouveränität					
Datensouveränität beschreibt die Fähigkeit eines Datenbereitstellers, selbst über die Nutzung seiner Datenbestände (einschließlich der Metadaten (Bastiaansen et al., 2019)) während ihres gesamten Lebenszyklus und einschließlich der von Dritten durchgeführten Aktionen zu bestimmen (Hellmeier & Scherenberg, 2023; Scherenberg et al., 2024). Anforderungen an die Datensouveränität werden üblicherweise in Verträgen oder Richtlinien formuliert und können derzeit nur zum Teil technisch durchgesetzt werden.							
Relevanz	Datensouveränität ist insbesondere bei der Nutzung von hochsensiblen Daten für industrielle Anwendungen und andererseits bei der direkten Einbindung der Daten von natürlichen Personen notwendig.						
Erfüllung durch Dateninfrastruktur-Designoption							
Zentralisiert	-	Föderiert	+	Dezentralisiert	+	Verteilt	○
Dateninfrastrukturentwürfe, die eine physische oder virtuelle Zentralisierung von Daten vermeiden, erhöhen die Datensouveränität der beteiligten Akteure.							
Interdependenzen							
Rechtssicherheit			↗				Datensouveränität
Transparenz			↘				Datensouveränität
Autonomie			↗				Datensouveränität

Organisatorisch Transparenz

Transparenz beschreibt den Grad der Sichtbarkeit und Zugänglichkeit von Informationen über die Dateninfrastruktur, die Akteure und die stattfindenden Datentransaktionen. Dies umfasst Elemente wie den aktuellen Status sowie den Umfang, Ort und die Einheit der Leistungserbringung.

Relevanz Transparenz ist für Anwendungen wichtig, in denen die Aktionen der Dateninfrastruktur nachvollzogen werden müssen, um beispielsweise eine faire Zusammenarbeit zu ermöglichen oder eine finanzielle Kompensation zu schaffen. Jedoch kann in einigen Fällen eine hohe Transparenz hinsichtlich der Daten oder der Transaktionen auch unerwünscht sein.

Erfüllung durch Dateninfrastruktur-Designoption

Zentralisiert	○	Föderiert	○	Dezentralisiert	○	Verteilt	+
---------------	---	-----------	---	-----------------	---	----------	---

In verteilten Dateninfrastrukturen werden die erfolgten Transaktionen für alle Akteure, die einen Knoten betreiben, sichtbar.

Interdependenzen

Datenintegrität und Datenverifizierbarkeit	↗	Transparenz
--	---	-------------

Organisatorisch Auditierbarkeit

Die Auditierbarkeit bezieht sich auf die Fähigkeit der Dateninfrastruktur, von einer dritten Partei auditert (d.h., inspiziert und überprüft) zu werden.

Relevanz Die Auditierbarkeit einer Dateninfrastruktur ist insbesondere in Bereichen mit hohen regulatorischen Anforderungen von Relevanz.

Erfüllung durch Dateninfrastruktur-Designoption

Zentralisiert	+	Föderiert	○	Dezentralisiert	-	Verteilt	-
---------------	---	-----------	---	-----------------	---	----------	---

Der Aufwand, um eine Dateninfrastruktur zu auditieren, ist bei zentralisierten Architekturen am geringsten. In Architekturen mit mehreren Knoten muss potenziell jeder einzelne Knoten auditiert werden.

Interdependenzen

Autonomie	↘	Auditierbarkeit
Komplexität	↘	Auditierbarkeit

Organisatorisch Autonomie

Autonomie bezieht sich auf das Ausmaß, in dem ein Akteur Dateninfrastrukturdienste nutzen kann, ohne von einem anderen spezifischen Akteur abhängig zu sein.

Relevanz Hohe Bedarfe nach Autonomie sind insbesondere in Industriebereichen mit „The winner takes it all“-Effekten zu erwarten, in denen Infrastrukturbereitstellende auch ein Eigeninteresse an der Realisierung von intelligenten Diensten besitzen.

Erfüllung durch Dateninfrastruktur-Designoption

Zentralisiert	-	Föderiert	+	Dezentralisiert	+	Verteilt	+
---------------	---	-----------	---	-----------------	---	----------	---

In föderierten, dezentralisierten und verteilten Architekturen können Dateninfrastrukturdienste ohne Abhängigkeit von einem zentralen Akteur realisiert werden.

Interdependenzen

Portabilität	↗	Autonomie
Komplexität	↗	Autonomie

Organisatorisch Komplexität

Komplexität ist eine Eigenschaft, die von natürlichen Personen wahrgenommen wird. Sie bezieht sich auf die Verständlichkeit einer Technologie und die Einfachheit, mit der sie entwickelt, implementiert, gewartet und genutzt werden kann. Je höher die Komplexität, desto größer das zur Umsetzung der Lösung benötigte Knowhow und die benötigte Zeit zur ersten Erreichung einer produktiven Nutzung.

Relevanz Eine geringe Komplexität und damit einhergehende Technologieakzeptanz ist insbesondere für Anwendungen relevant, in welchen kleinere und mittlere Unternehmen (KMU) oder natürliche Personen über die Dateninfrastruktur miteinander interagieren. Zudem ermöglicht eine geringe Komplexität die Realisierung von Quick Wins.

Erfüllung durch Dateninfrastruktur-Designoption

Zentralisiert	+	Föderiert	○	Dezentralisiert	+	Verteilt	○
---------------	---	-----------	---	-----------------	---	----------	---

In diesem Sinne entspricht das „+“ einer guten (geringen) Komplexität. Zentralisierte Dateninfrastrukturen sind üblicherweise einfach zu implementieren und zu warten. Das Gleiche gilt für dezentralisierte Dateninfrastrukturen im Sinne mehrerer zentralisierter Dienste.

Organisatorisch Fairness

Fairness beschreibt den Umfang, in dem die Dateninfrastruktur ein sogenanntes "level playing field" schafft, indem sie die Bedürfnisse der nutzenden Parteien ausgleicht und Offenheit für weitere Dienstangebote gewährleistet.

Relevanz Eine faire Möglichkeit zur Nutzung der Dateninfrastruktur ist insbesondere in solchen Szenarien möglich, in denen konkurrierende Serviceangebote durch dieselbe Dateninfrastruktur gestützt werden sollen.

Erfüllung durch Dateninfrastruktur-Designoption

Zentralisiert	○	Föderiert	+	Dezentralisiert	+	Verteilt	+
---------------	---	-----------	---	-----------------	---	----------	---

Eine faire Nutzungsmöglichkeit der Dateninfrastruktur wird durch die Bereitstellung eigener Knoten der Dateninfrastruktur gewährleistet, die eine Ausrichtung an den individuellen Bedürfnissen eines Akteurs ermöglichen.

Interdependenzen

Komplexität	↗	Fairness
Externe Interoperabilität	↗	Fairness
Portabilität	↗	Fairness

Organisatorisch Vertraulichkeit

Die Vertraulichkeit einer Dateninfrastruktur umfasst Mechanismen, die den Schutz der Privatsphäre fördern, um den Bedürfnissen der Datenanbietenden sowie der betroffenen Personen sowie und den gesetzlichen Anforderungen gerecht zu werden.

Relevanz Die Vertraulichkeit einer Dateninfrastruktur spielt insbesondere für den Transfer von personenbezogenen Daten oder weiteren für die Geschäftsmodelle der Akteure besonders sensiblen Daten eine wichtige Rolle.

Erfüllung durch Dateninfrastruktur-Designoption

Zentralisiert	-	Föderiert	+	Dezentralisiert	+	Verteilt	-
---------------	---	-----------	---	-----------------	---	----------	---

Die Vertraulichkeit wird insbesondere durch Architekturen gewahrt, in denen Daten nur mit ausgewählten Parteien geteilt werden und eine lokale Verarbeitung von Daten stattfinden kann.

Organisatorisch Rechtssicherheit

Die Rechtssicherheit bezieht sich auf die Erfüllung der geltenden Gesetzgebung. Dazu gehören die allgemeine Datengesetzgebung wie die Datenschutzgrundverordnung, der Data Act oder der AI-Act sowie darüberhinausgehende bereichsspezifische regulatorische Anforderungen, z.B. im Bezug auf Datensicherheit oder Datenlokalisierung.

Relevanz Die Einhaltung relevanter rechtlicher Regularien ist für jegliche Dateninfrastruktur relevant.

Erfüllung durch Dateninfrastruktur-Designoption

Zentralisiert	○	Föderiert	○	Dezentralisiert	○	Verteilt	○
---------------	---	-----------	---	-----------------	---	----------	---

Allgemein kann keine Architektur als generell rechtssicherer bewertet werden. Grundsätzlich ist es empfehlenswert, den Einzelfall zu prüfen.

Organisatorisch Strategischer Fit

Unter der Charakteristik "Strategischer Fit" wird das Maß verstanden, in welchem die Dateninfrastruktur mit den für den individuellen Akteur relevanten Strategien harmonisiert. Dazu gehören etwa die Organisationsstrategie, die IT-Strategie oder die Datenstrategie.

Relevanz Die Notwendigkeit einer Bewertung des strategischen Fits muss für die individuelle Organisation beurteilt werden.

Erfüllung durch Dateninfrastruktur-Designoption

Zentralisiert	○	Föderiert	○	Dezentralisiert	○	Verteilt	○
---------------	---	-----------	---	-----------------	---	----------	---

Eine allgemeine Beurteilung des strategischen Fits ist nicht möglich, da dieser von den individuellen Strategien der Akteure abhängt.

Organisatorisch Interne Interoperabilität

Interne Interoperabilität beschreibt die Fähigkeit einzelner Dateninfrastrukturdienste, untereinander Informationen auszutauschen und die ausgetauschten Informationen gegenseitig zu nutzen. Aus Sicht der Dateninfrastruktur bezieht sich Interoperabilität hauptsächlich auf organisatorische, semantische und technische Interoperabilität (Firdausy et al., 2022).

Relevanz Interne Interoperabilität zwischen Dateninfrastrukturdiensten ist für Anwendungen relevant, in denen Dateninfrastrukturdienste in räumlicher Verteilung benötigt werden.

Erfüllung durch Dateninfrastruktur-Designoption

Zentralisiert		Föderiert	+	Dezentralisiert	-	Verteilt	○
---------------	--	-----------	---	-----------------	---	----------	---

Da ein föderiertes System per Definition auf Interoperabilität und Informationsaustausch basiert, ist hier die interne Operabilität generell am größten. Existiert nur ein zentraler Dienst, so ist keine Interoperabilität zwischen Diensten gleichen Typs notwendig.

Organisatorisch Externe Interoperabilität

Externe Interoperabilität beschreibt die Fähigkeit einer Dateninfrastruktur, sich in bestehende und genutzte Unternehmensdateninfrastrukturen, Plattformen und andere bestehende Dateninfrastrukturen (beispielsweise in verbundenen Sektoren) auf technischer und prozessualer Ebene zu integrieren, um eine möglichst nahtlose Erfahrung zu bieten. In externer Sicht ist Interoperabilität hinsichtlich rechtlicher, organisatorischer, semantischer und technischer Aspekte erforderlich.

Relevanz Eine hohe Interoperabilität mit externen Systemen ist insbesondere dann erforderlich, wenn nahtlose Prozesse zwischen den Systemen eines Akteurs und den Dateninfrastrukturdiensten gewährleistet werden sollen oder wenn vielgenutzte Plattformen existieren, deren Ablösung durch die Dateninfrastruktur mit hohen Aufwänden verbunden ist.

Erfüllung durch Dateninfrastruktur-Designoption

Zentralisiert	-	Föderiert	+	Dezentralisiert	o	Verteilt	-
---------------	---	-----------	---	-----------------	---	----------	---

Durch ihre Flexibilität und Anpassungsfähigkeit der Dienste kann durch föderierte Dateninfrastrukturen potenziell die höchste Interoperabilität mit externen Diensten erreicht werden.

Technisch Portabilität

Portabilität ist die Fähigkeit, Daten oder Softwaredienste problemlos von einem System auf ein anderes zu übertragen.

Relevanz Die Portabilität eines Dateninfrastrukturdienstes ist üblicherweise von Relevanz, wenn mit dessen Nutzen höhere langfristige Kosten einhergehen oder die Gefahr besteht, das System nicht weiter betreiben zu können.

Erfüllung durch Dateninfrastruktur-Designoption

Zentralisiert	-	Föderiert	+	Dezentralisiert	+	Verteilt	o
---------------	---	-----------	---	-----------------	---	----------	---

Föderierte und dezentralisierte Architekturen unterstützen im Allgemeinen den Umzug der Dienste von einem System auf ein anderes System am besten.

Technisch Skalierbarkeit

Die Skalierbarkeit einer Dateninfrastruktur bezieht sich auf die Fähigkeit, potenziell wachsende Datenmengen, gestiegene Transaktionszahlen oder höhere Akteurszahlen zu unterstützen, ohne die Servicequalität einschränken zu müssen.

Relevanz Die Skalierbarkeit einer Dateninfrastruktur ist insbesondere bei Anwendungsfällen mit einer großen Anzahl von Beteiligten oder einer hohen Frequenz von Datentransaktionen relevant.

Erfüllung durch Dateninfrastruktur-Designoption

Zentralisiert	o	Föderiert	+	Dezentralisiert	+	Verteilt	-
---------------	---	-----------	---	-----------------	---	----------	---

Zentralisierte Systeme stellen einen Flaschenhals bei der Bearbeitung von Anfragen dar. Verteilte Systeme sind durch ihre Konsenslogiken oftmals in der Transaktionszahl beschränkt.

Interdependenzen

Kostenamortisation	↘	Skalierbarkeit
--------------------	---	----------------

Technisch		Ressourcenintensität					
Die Ressourcenintensität einer Dateninfrastruktur umfasst beispielsweise die Nutzung von Speicher- und Rechenressourcen, die wiederum den Energieverbrauch der Lösung beeinflussen.							
Relevanz	Die Ressourcenintensität einer Dateninfrastruktur ist insbesondere für nachhaltigkeitsorientierte Anwendungen ein wichtiges Kriterium. Ebenso sollte dieses Kriterium bei stark skalierenden Dateninfrastrukturen betrachtet werden.						
Erfüllung durch Dateninfrastruktur-Designoption							
Zentralisiert	+	Föderiert	○	Dezentralisiert	○	Verteilt	-
Da der Ressourcenverbrauch mit der zunehmenden Anzahl redundanter Dienste ansteigt, besitzt das zentralisierte Dateninfrastrukturdesign die geringste Ressourcenintensität.							

Technisch		Generativität					
Generativität beschreibt die Eigenschaft einer Dateninfrastruktur, durch einzelne Organisationen anpassbar oder erweiterbar zu sein, um weitere Mehrwertdienste oder Anwendungsfälle zu ermöglichen.							
Relevanz	Die Generativität einer Dateninfrastruktur ist vor allem relevant, wenn der Umfang der unterstützten Anwendungsfälle zum Zeitpunkt des Designs nicht abgesehen werden kann oder noch nicht vollständig definiert ist.						
Erfüllung durch Dateninfrastruktur-Designoption							
Zentralisiert	-	Föderiert	+	Dezentralisiert	+	Verteilt	-
Insbesondere föderierte und dezentralisierte Dateninfrastrukturen ermöglichen die Anpassung der individuellen Dienste, solange definierte Mindeststandards eingehalten werden.							

Technisch		Orchestrierbarkeit					
Orchestrierbarkeit beschreibt die Fähigkeit, eine Dateninfrastruktur global zu koordinieren und zu optimieren.							
Relevanz	Die globale Orchestrierbarkeit einer Dateninfrastruktur ist dann besonders sinnvoll, wenn eine Anwendung oder ein System über mehrere geographische Regionen hinweg operiert und eine koordinierte und effiziente Verarbeitung und Verwaltung der Daten erforderlich ist.						
Erfüllung durch Dateninfrastruktur-Designoption							
Zentralisiert	○	Föderiert	○	Dezentralisiert	-	Verteilt	+
Eine verteilte Architektur bietet die besten Voraussetzungen für eine globale Orchestrierung, da sie eine virtuell zentralisierte Steuerung mit verteilter und redundanter Datenhaltung verbindet.							

Technisch		Datenintegrität und Datenverifizierbarkeit					
Datenintegrität bezieht sich auf die Fähigkeit einer Dateninfrastruktur, die Richtigkeit und Vollständigkeit der Daten über ihren gesamten Lebenszyklus zu gewährleisten (Bastiaansen et al., 2019). Verifizierbarkeit bezieht sich auf die Möglichkeit, die Richtigkeit und Authentizität der Daten selbst prüfen zu können.							
Relevanz	Integrität und Verifizierbarkeit von Daten sind insbesondere für Anwendungen relevant, die eine datenbasierte Steuerung umsetzen oder eine Aufteilung von finanziellen Mitteln basierend auf zugelieferten Daten durchführen.						
Erfüllung durch Dateninfrastruktur-Designoption							
Zentralisiert	+	Föderiert	○	Dezentralisiert	○	Verteilt	+
Die Sicherstellung der Datenintegrität ist in zentralisierten Dateninfrastrukturen durch zentralisierte Kontrollinstanzen am einfachsten umzusetzen. Eine Verifizierung durch Außenstehende ist jedoch schwierig. Hierzu bietet sich die Nutzung einer verteilten Infrastruktur an.							

Technisch		Zuverlässigkeit					
Zuverlässigkeit bezeichnet die Fähigkeit der Dateninfrastruktur, jederzeit wie vorhergesehen, d.h. ohne Unterbrechungen oder Ausfälle, zu funktionieren.							
Relevanz	Die zuverlässige Funktion einer Dateninfrastruktur ist für nahezu alle Anwendungen unabdingbar. Insbesondere relevant ist dies für Anwendungen, die einen Datenaustausch in Echtzeit benötigen und bei denen Verzögerungen in der Datenübertragung einen Ausfall des Systems herbeiführen könnten.						
Erfüllung durch Dateninfrastruktur-Designoption							
Zentralisiert	-	Föderiert	○	Dezentralisiert	○	Verteilt	+
Fallen in föderierten oder dezentralisierten Dateninfrastrukturen einzelne Knoten aus, können die weiteren Knoten das System weiter betreiben. Fallen in verteilten Systemen einzelne Knoten aus, ist das Gesamtabbild weiterhin über andere Knoten verfügbar.							
Interdependenzen							
Skalierbarkeit		↗		Zuverlässigkeit			

Technisch		Wiederherstellung					
Wiederherstellung bezieht sich auf die Fähigkeit, den Zugang zu Daten und Systemen, z. B. im Falle eines Hardwareausfalls oder eines Bedienungsfehlers, wiederzuerlangen.							
Relevanz	Die Wiederherstellbarkeit eines Teilsystems der Dateninfrastruktur kann insbesondere dann ein relevantes Kriterium darstellen, wenn eine große Angst vor Datenverlusten vorherrscht.						
Erfüllung durch Dateninfrastruktur-Designoption							
Zentralisiert	-	Föderiert	○	Dezentralisiert	-	Verteilt	+
Einzelne Teilsysteme lassen sich bei verteilter Dateninfrastruktur wiederherstellen und mit den gemeinsam geteilten Daten populieren.							
Interdependenzen							
Zuverlässigkeit		↗		Wiederherstellung			

Technisch		Informationssicherheit und Cybersicherheit					
Sicherheit bezeichnet generell den Schutz von Vermögenswerten vor den verschiedenen Bedrohungen, die von bestimmten inhärenten Schwachstellen ausgehen. Informationssicherheit bezieht sich dabei auf Informationen als Asset, während Cybersicherheit auch den Schutz von Personen und weiteren über den Cyberspace erreichbaren Vermögenswerten (bspw. Maschinen) betrachtet (Solms & van Niekerk, 2013).							
Relevanz	Besondere Relevanz haben Informations- und Cybersicherheit für Bereiche der kritischen Infrastruktur aber auch für Sektoren wie die industrielle Produktion, in denen Cyberangriffe potenziell hohe wirtschaftliche Schäden verursachen können.						
Erfüllung durch Dateninfrastruktur-Designoption							
Zentralisiert	<input type="radio"/>	Föderiert	<input type="radio"/>	Dezentralisiert	<input type="radio"/>	Verteilt	<input type="radio"/>
Eine generelle Aussage zu Informations- oder Cybersicherheit auf Basis der Architekturoptionen ist nicht möglich.							

5 Ausblick: die nächsten Schritte

Bei der Auslegung von Dateninfrastrukturdiensten muss eine fundierte Entscheidung getroffen werden, die auf den tatsächlichen Anforderungen der Geschäftsmodelle basiert, die notwendigen gesellschaftlichen Rahmenbedingungen berücksichtigt und einen Ausgleich zwischen den Interessen der verschiedenen am Data Sharing beteiligten Akteure schafft. Diese Entscheidungshilfe unterstützt die informierte Gestaltung der Systemarchitektur durch einen strukturierten Gestaltungsprozess. Im Anschluss folgen nun das Detaildesign, die Umsetzung sowie der Betrieb der Dateninfrastruktur. Zur Unterstützung des detaillierten Entwurfs der Dateninfrastruktur bietet sich ein Rückgriff auf bestehende Referenzarchitekturen verschiedener Initiativen an:

Vor dem Hintergrund einer zentralisierten Gestaltung der Dateninfrastruktur erläutern die Referenzarchitekturen für Big Data von ISO (ISO, 2020) und NIST (NIST, 2019) Rollen, Aktivitäten, Funktionen und Architekturkomponenten, um Big Data-Architekturen umzusetzen. Zum Detaildesign von föderierten Dateninfrastrukturen können die Architekturdokumente der IDSA (International Data Spaces Association, 2022) und von Gaia-X (Gaia-X European Association for Data and Cloud AISBL, 2024) als Referenz herangezogen werden. Zudem liefert die technische Spezifikation von SIMPL eine Übersicht existierender zentraler und dezentraler technischer Komponenten zur Realisierung funktionaler Anforderungen von Dateninfrastrukturen (Sovereign-X, 2024).

Für DLT-basierte Systeme wird in ISO 23257 eine Referenzarchitektur definiert (ISO, 2022). Aufgrund der begrenzten Interoperabilität einzelner DLT zueinander und auf Basis ihrer unterschiedlichen Eigenschaften, beispielsweise hinsichtlich des genutzten Konsensprotokolls, müssen zur konkreten Umsetzung einzelne DLT-Frameworks betrachtet werden. Im Bereich des Data Sharing sind die Frameworks Hyperledger Fabric (Hyperledger, 2023) oder Ocean Protocol (Ocean Protocol Foundation Ltd., 2024) besonders populär.

Für die Implementierung und den Betrieb der Dateninfrastrukturen kommen neben der Umsetzung durch Cloud Services Provider vermehrt auch dedizierte Unternehmen, die sich auf die Bereitstellung von Dateninfrastrukturdiensten fokussieren, in Betracht. Beim Betrieb von föderierten oder verteilten Infrastrukturen ist es möglich, mehr als einen Betreibenden zu wählen oder den einzelnen Akteuren eine freie Wahl hinsichtlich des Betriebs ihrer individuellen Dateninfrastrukturdienste zu überlassen.

Insgesamt zeigt die Auswahl der Systemarchitektur von Dateninfrastrukturen beispielhaft die Komplexität der Fragestellungen, mit denen sich Akteure in der Datenwirtschaft beschäftigen müssen. Entsprechend sollten Unternehmen verstärkt Kompetenzen in den Bereichen Dateninfrastrukturen und Datenökosysteme aufbauen, um die Potenziale von Anwendungen auf Basis geteilter Daten realisieren zu können. Wertschöpfungspotenziale bieten sich durch verschiedene Rollen und Aufgaben wie beispielsweise die Anwendungsentwicklung, die Systemintegration, die Datenbereitstellung oder den Infrastrukturbetrieb. Eine Kompetenz im Bereich von Dateninfrastrukturen und Datenökosystemen ermöglicht zudem eine frühzeitige Beteiligung in den Entwurfsphasen von Dateninfrastrukturen und den damit verbundenen Vorteil, die Dateninfrastruktur anhand der Bedürfnisse des Unternehmens mitgestalten zu können. Auch im Hinblick auf die vermehrte regulatorische Pflicht zur Datenbereitstellung, beispielsweise im Rahmen des Data Act (Straub & Bogenstahl, 2024), stellen Data Sharing-Kompetenzen einen wichtiger Faktor für die Kontinuität des Geschäftsbetriebs dar.

Referenzen

- Bader, S., Belyaev, A., Clauer, E., Diedrich, C., Fritz, R., Hasler, J., Karnouskos, S. & Schnicke, F. (2023). *Decentralized-Registries-Taxonomy-of-decentralized-registries-and-an-architectural-overview*. <https://doi.org/10.13140/RG.2.2.34793.26723>
- Bastiaansen, H., Dalmolen, S., Kollenstart, M. & Punter, M. (2019). Infrastructural Sovereignty over Agreement and Transaction Data ('Metadata') in an Open Network-Model for Multilateral Sharing of Sensitive Data. *Fortieth International Conference on Information Systems*. https://aisel.aisnet.org/icis2019/economics_is/economics_is/23/
- Demchenko, Y., Laat, C. de & Membrey, P. (2014). Defining architecture components of the Big Data Ecosystem. In W. W. Smari (Hrsg.), *2014 International Conference on Collaboration Technologies and Systems (CTS 2014): Minneapolis, Minnesota, USA, 19 - 23 May 2014 ; [including symposia and workshops]* (S. 104–112). IEEE. <https://doi.org/10.1109/CTS.2014.6867550>
- Dodds, L. & Wells, P. (2019). Data infrastructure. In T. Davies (Hrsg.), *The State of Open Data: Histories and Horizons*. African Minds. <https://idrc-crdi.ca/en/book/state-open-data-histories-and-horizons>
- DS4SSCC. (2023). *The Data Cooperation Canvas*. <https://www.datacooperationcanvas.eu/local/canvas/uploads/The%20Data%20Cooperation%20Canvas%20-%20Template.pdf>
- Firdausy, D. R., Alencar Silva, P. de, van Sinderen, M. & Iacob, M.-E. (2022). Towards a Reference Enterprise Architecture to enforce Digital Sovereignty in International Data Spaces. In *2022 IEEE 24th Conference on Business Informatics (CBI)* (S. 117–125). IEEE. <https://doi.org/10.1109/CBI54897.2022.00020>
- Gaia-X European Association for Data and Cloud AISBL (Hrsg.). (2024). *Gaia-X Architecture Document - 24.04 Release*. <https://docs.gaia-x.eu/technical-committee/architecture-document/24.04/>
- Gane, C. & Sarson, T. (1979). *Structured systems analysis: tools and techniques*. Prentice-Hall software series. Prentice-Hall.
- Hellmeier, M. & Scherenberg, F. von (2023). A DELIMITATION OF DATA SOVEREIGNTY FROM DIGITAL AND TECHNOLOGICAL SOVEREIGNTY. *Thirty-first European Conference on Information Systems (ECIS 2023)*. https://aisel.aisnet.org/ecis2023_rp/306/
- Hyperledger (Hrsg.). (2023). *A Blockchain Platform for the Enterprise*. <https://hyperledger-fabric.readthedocs.io/en/release-2.5/>
- International Data Spaces Association (Hrsg.). (2022). *IDS-RAM 4*. <https://docs.internationaldataspaces.org/ids-knowledgebase/ids-ram-4>
- ISO (2019). *ISO 15704:2019: Enterprise modelling and architecture — Requirements for enterprise-referencing architectures and methodologies*.
- ISO (2020). *ISO/IEC 20547-3:2020: Information technology — Big data reference architecture — Part 3: Reference architecture*.

- ISO (2022). *ISO 23257:2022: Blockchain and distributed ledger technologies — Reference architecture*.
- Mayer, R. C., Davis, J. H. & Schoorman, D. F. (1995). An Integrative Model of Organizational Trust. *Academy of Management Review*. <https://doi.org/10.2307/258792>
- NIST. (2019). *NIST Big Data Interoperability Framework*. <https://doi.org/10.6028/NIST.SP.1500-6r2>
- Object Management Group. (2015). *Business Motivation Model: Version 1.3*. <https://www.omg.org/spec/BMM/1.3/>
- Ocean Protocol Foundation Ltd. (Hrsg.). (2024). *Architecture Overview*. <https://docs.oceanprotocol.com/developers/architecture>
- Osterwalder, A., Pigneur, Y. & Clark, T. (2010). *Business model generation: A handbook for visionaries, game changers, and challengers*. Wiley.
- Otto, B. & Burmann, A. (2021). Europäische Dateninfrastrukturen. *Informatik Spektrum*. Vorab-Onlinepublikation. <https://doi.org/10.1007/s00287-021-01386-4>
- Otto, B., Hompel, M. ten & Wrobel, S. (Hrsg.). (2022). *Designing Data Spaces: The Ecosystem Approach to Competitive Advantage* (1st ed. 2022). Springer International Publishing. <https://doi.org/10.1007/978-3-030-93975-5>
- Scherenberg, F. von, Hellmeier, M. & Otto, B. (2024). Data Sovereignty in Information Systems. *Electronic Markets*, 34(1). <https://doi.org/10.1007/s12525-024-00693-4>
- Schleimer, A. M., Jahnke, N. & Otto, B. (2023). Architecture Design Options for Federated Data Spaces. *Proceedings of the 56th Hawaii International Conference on System Sciences*, 3643–3652. <https://scholarspace.manoa.hawaii.edu/bitstreams/941a67b1-d2e4-4394-8f3f-11b8b7d370e9/download>
- Solms, R. von & van Niekerk, J. (2013). From information security to cyber security. *Computers & Security*, 38, 97–102. <https://doi.org/10.1016/j.cose.2013.04.004>
- Sovereign-X. (2024). *Functional and Technical Architecture Specifications*. <https://simpl-programme.ec.europa.eu/system/files/2024-12/Simpl-Open%20Functional%20and%20Technical%20Architecture%20Specifications.pdf>
- Straub, S. & Bogenstahl, C. (2024). *Die aktuelle EU-Gesetzgebung im Bereich Digitalisierung und Datenwirtschaft*. https://www.iit-berlin.de/wp-content/uploads/2024/12/KI_Leitfaden_EU_Recht-2024.pdf
- Tan, K. L., Chi, C.-H. & Lam, K.-Y. (2024). Survey on Digital Sovereignty and Identity: From Digitization to Digitalization. *ACM Computing Surveys*, 56(3), 1–36. <https://doi.org/10.1145/3616400>
- Weill, P., van den Berg, E., Birnbaum, J. & Planta, M. de. (2024). *What's Next: Top Performers Are Becoming Real-Time Businesses*. https://c isr.mit.edu/publication/2024_0801_Real-TimeBusiness_WeillvanderBergBirnbaumdePlanta?utm_source=pressrelease&utm_medium=pr&utm_campaign=realtimebusiness#fn_1