Bundesministerium für Bildung und Forschung

Finanziert von der Europäischen Union
NextGenerationEU

April 2024

Wissenschaftliche Begleitung und Vernetzung der Projekte zur Entwicklung und praktischen Erprobung von Datentreuhandmodellen in den Bereichen Forschung und Wirtschaft

Arbeitspaket 1.1 Bericht zur Bestandsaufnahme

www.technopolis-group.com

April 2024

Wissenschaftliche Begleitung und Vernetzung der Projekte zur Entwicklung und praktischen Erprobung von Datentreuhandmodellen in den Bereichen Forschung und Wirtschaft

Arbeitspaket 1.1 Bestandsaufnahme

Technopolis Group: Stephan Kreutzer, Prof. Dr. Thomas Heimer, Heike Nachtigall, Lisa Pschorn, Fabian Waiblinger
Fraunhofer ISI: Prof. Dr. Knut Blind, Dr. Nicholas Martin, Dr. Djerdj Horvat
Law & Innovation: Prof. Dr. Max von Grafenstein, Marcus Schweinberg
GRI GmbH, RWTH Aachen: Prof. Dr. Rita Streblow, Junsong Du, Joel Schölzel

Die Studie wird im Auftrag des Bundesministeriums für Bildung und Forschung (ko-finanziert durch das Programm „NextGenerationEU“ der Europäischen Union) durchgeführt.

Inhaltsverzeichnis

Abbildungen

Tabellen

1 Einführung: Einordnung in die Begleitforschung und methodisches Vorgehen

Daten gelten heute als der wesentliche Treiber für die Wertschöpfung, da sie neue Erkenntnisse, neue Anwendungen oder Prozesse schaffen können. Um diese Wirkung entfalten zu können, müssen Daten aus verschiedenen Bereichen zusammengeführt werden. Das Teilen und Nutzen von Daten für datengetriebene Forschung, Innovation und Wertschöpfung hilft so bei der Schaffung von Lösungen für gesellschaftliche Herausforderungen wie die Entwicklung nachhaltiger Mobilitätskonzepte, die Erforschung komplexer Klimasysteme oder die Bekämpfung seltener Krankheiten. Daten erlangen somit erst durch das Teilen innerhalb einer Anwendungsdomäne ihren Wert, der sich durch das Teilen von Daten zwischen Akteuren aus unterschiedlichen Anwendungsdomänen in Forschung und Wirtschaft weiter steigern kann.

Aus innovationspolitischer Sicht werden Daten nach unserer **Arbeitshypothese** aktuell sowohl innerhalb als auch zwischen Anwendungsdomänen nicht ausreichend zwischen Akteuren in Forschung und Wirtschaft geteilt. Gründe hierfür sind unter anderem zu hohe Transaktionskosten aufgrund fehlender technischer Voraussetzungen und zu hoher rechtlicher Hürden, mangelnde Anreize, fehlendes Vertrauen und nicht zuletzt Koordinationsprobleme unter den Beteiligten, die entsprechenden Voraussetzungen gemeinsam zu schaffen. Man kann spieltheoretisch auch von einem **Dilemma kollektiven Handelns** sprechen.

Zur Förderung des Teilens von Daten gibt es verschiedene **Lösungsansätze**. Die geeignetsten Instrumente unterscheiden sich je nach Anwendungsfall und Datenökosystem. Ein möglicher Lösungsansatz wird in der Literatur unter dem Begriff **‚Datentreuhandmodelle'** beschrieben (DT). Das Bundesministerium für Bildung und Forschung (BMBF) fördert deshalb 20 **Pilotprojekte**, die in verschiedenen Sektoren, aber auch sektorübergreifend Datentreuhandmodelle als mögliches Instrument zur Förderung der Nutzung von Daten für Forschung und Innovation entwickeln.

Das BMBF hat ein Konsortium mit der **wissenschaftlichen Begleitung der Umsetzung der geförderten Pilotprojekte** beauftragt. Die Begleitforschung ergänzt die Erprobung von Datentreuhandmodellen durch eine Analyse der wissenschaftlichen Literatur, durch die übergeordnete Analyse von in den Pilotprojekten auftretenden Herausforderungen und entwickelten Lösungsansätzen sowie durch die Ausarbeitung von Bausteinen zielführender Datentreuhandmodelle. Dabei konzentriert sich die Begleitforschung auf **vier Querschnittsthemen** mit jeweils unterschiedlicher fachlicher Perspektive und Fragestellungen: Technische Infrastruktur, rechtliche Rahmenbedingungen, Geschäftsmodelle sowie Akzeptanz, Skalierung und Transfer von DT.

Der **vorliegende Bericht** präsentiert die Ergebnisse aus dem Arbeitspaket 1.1 – **Bestandsaufnahme** in der Begleitforschung. Hierfür wurde eine Auswahl der **wissenschaftlichen Literatur** zu Datentreuhandmodellen (im Folgenden: DT) ausgewertet. Die Ergebnisse werden in **Abschnitt 2** präsentiert. Zunächst werden relevante, in der Literatur verwendete Konzepte in unser Begriffsverständnis eingeordnet, welches sich an der Definition des BMBF orientiert. Anschließend werden Aussagen aus der ausgewerteten Literatur zu zentralen Fragestellungen in den vier Querschnittsthemen präsentiert. Das Kapitel stellt einen Zwischenstand dar und wird im Laufe der Begleitforschung weiterentwickelt. Außerdem wurden insgesamt sechs Anwendungsbeispiele von DT oder vergleichbaren Lösungsansätzen zur Förderung des Datenteilens aus dem In- und Ausland analysiert (**Use Cases**). Die Ergebnisse finden sich in **Abschnitt 3.** Im **Anhang** findet sich das Literaturverzeichnis.

2 Literaturanalyse: Erste Erkenntnisse

2.1 Begriffsverständnis

Ausgangspunkt der begrifflichen Eingrenzung bildet die im Auftrag zur Begleitforschung verwendete Definition eines DT. Diese definiert DT als: „Neutraler **Intermediär**, der einen **vertrauensvollen** und fairen **Ausgleich der Interessen** der beteiligten **Akteur*innen** – Datengebende sowie Datennutzende– ermöglicht, ggf. neue Vertrauensbeziehungen anbahnt, den **technischen und organisatorischen Zugang** zu qualitativ hochwertigen Daten unter Wahrung des **Datenschutzes** sowie **Interoperabilität** garantiert“ (BMBF, 2021). DT können dabei

verschiedene Ausprägungen haben und sowohl ‚passiv' agieren und Daten lediglich ‚lesen' als auch ‚aktiv' Daten in einen Datenraum ‚einschreiben' bzw. diese bearbeiten. DT können also einseitig oder zweiseitig Datenströme organisieren.

In der Literatur wird zwischen den folgenden **Data Governance-Mechanismen** unterschieden, die weitgehend in die BMBF-Definition eines Datentreuhandmodells fallen und in der englischsprachigen Literatur auch Untermengen des Begriffs „**Data Stewardship**" bilden:

- **Data Cooperatives**: Datengenossenschaften, in denen die Mitglieder demokratisch die Kontrolle und Entscheidungsfindung über ihre Daten ausüben (ODI, 2019).
- **Data Trusts**: Eine rechtliche Struktur, welche die Verantwortung und Verpflichtungen für die Verwaltung von Daten für einen vereinbarten Zweck gegenüber einer Gruppe von Begünstigten wahrnimmt (ODI, 2019; IPPI, 2022). Die Anforderung an einen Datentreuhänder (DT) ist häufig vergleichbar mit einem Datatrustee eines Data Trust.
- **Data Commons**: Anwendung von Grundsätzen und Erkenntnissen aus der Bewirtschaftung gemeinsamer Ressourcen auf Daten (ODI, 2019). Die Anforderung an einen Datentreuhänder (DT) ist häufig vergleichbar mit einem Verwalter eines Data Commons.
- **Data Spaces:** Dezentralisiertes Datenökosystem (bzw. Dateninfrastruktur), das auf gemeinsam vereinbarten Technologien, Werten, Standards oder Schnittstellen basiert und einen effektiven und vertrauenswürdigen Austausch von Daten zwischen den Ökosystem-Teilnehmern ermöglicht. Ziel ist die Schaffung mehrseitiger Märkte, die Steigerung bzw. Verbesserung der Datenverfügbarkeit und bessere Datenzugriffsmöglichkeiten der Teilnehmer. Bausteine eines Datenraums umfassen Bedingungen und Mechanismen in Verbindung mit Datenangeboten, wie für die Preisgestaltung und Vertragsprozesse sowie die Veröffentlichung und das Auffinden von Datenangeboten (Otto et al., 2022). Innerhalb des Ökosystems werden verschiedene Rollen definiert, die sich in die drei Gruppen (i) Governance, (ii) Teilnehmende, und (iii) unterstützende Dienstleistungen einteilen lassen. Die **Governance-Entität von Datenräumen** kann auch Funktionen wie die Herstellung von Rechtssicherheit etwa über ein Identitätsmanagement der Teilnehmenden oder Zertifizierungen übernehmen und fällt somit in unsere breit angesetzte Definition eines Datentreuhänders (DT). Dies trifft im Kern auch auf die Rolle des „**Föderator**" im Kontext von Gaia-X zu (vgl. dazu Kraemer et al., 2023). Datenräume können offen oder geschlossen sein (in verschiedenen Abstufungen), der Zugang kann diskriminierend oder nicht-diskriminierend erfolgen.
- **Personal Information Management Systems** (PIMS): Systeme, die Individuen mehr Kontrolle über seine/ihre persönlichen Daten geben. PIMS ermöglichen es Einzelpersonen, ihre persönlichen Daten in sicheren Speichersystemen zu verwalten und sie nach eigenem Ermessen mit anderen zu teilen. Anbieter von Online-Diensten und Werbetreibende müssen dann mit PIMS interagieren, wenn sie die diese Daten verarbeiten wollen.[1] Bei PIMS handelt es sich um Geschäftsmodelle, die im Auftrag von Verbraucher*innen gegenüber Dritten agieren (Schneider, 2022).

Datentreuhandmodelle lassen sich anhand der Literatur entlang verschiedener Dimensionen unterscheiden, z.B. der Art der Datenspeicherung (zentral / dezentral), Art der Datennutzung, Art der Datenmonetarisierung oder Art der Rechtsform (Arlinghaus et al., 2021; Lindner & Straub, 2023; Stevens & Boden, 2022). Zudem ist es hilfreich, zwischen B2B (Business-to-Business)-Intermediären für den Datentransfer zwischen Unternehmen[2] und C2B (Consumer-to-Business)-Intermediären zu unterscheiden (Schneider, 2022; RfII, 2021).

Die Datenethikkommission folgt einem engeren Begriffsverständnis, das lediglich bestimmte C2B-Modelle, sogenannte Privacy Management Tools sowie PIMS umfasst (Specht Riemenschneider; Kerber, 2022). Bei Datenmanagement- und Datentreuhandsystemen wird zwischen zwei Modellen unterschieden: zum einen (i) „technische Dashboards" mit Einwilligungsmanagement, zum anderen (ii) umfassende Dienstleistungen der

[1] https://edps.europa.eu/data-protection/our-work/subjects/personal-information-management-system_en.

[2] Oder auch zwischen Unternehmen und Behörden (B2G– Business-to-Government).

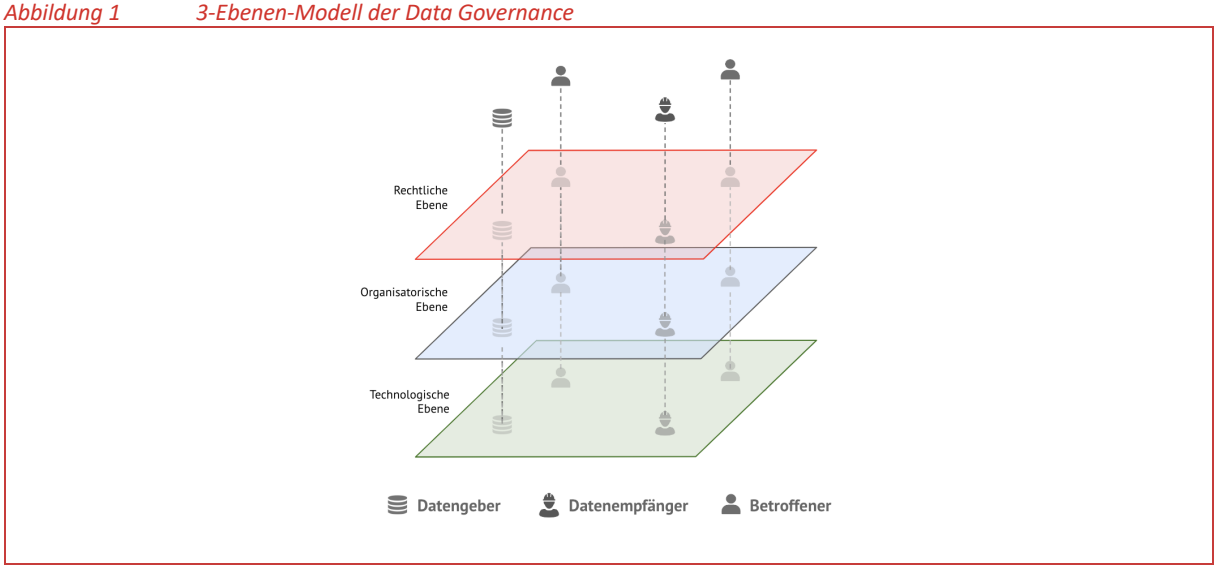
Daten- und Einwilligungsverwaltung (RfII, 2021). Entsprechend können im Auftrag der Datengebenden neben der Verwaltung von Zugriffsrechten auch Funktionen wie die Speicherung, Aufbereitung, Pseudonymisierung, Veredelung sowie die Weitergabe von aggregierten Daten oder Auswertungen, oder auch die Verhandlung über Datenzugriffsrechte sowie die Durchsetzung von Auskunft-, Widerruf-, und Löschansprüche wahrgenommen werden (RfII, 2021; Schneider, 2022). Laut Rfii ist das Echo aus Verbraucherschutzsicht zu diesem Geschäftsmodell durchwachsen (RfII, 2021).

Daneben umfassen C2B-Modelle auch Treuhandstellen für persönliche Daten in Bereichen wie Gesundheit und Mobilität. Im B2B-Bereich finden sich auch Initiativen von Unternehmen, die ein „Participatory Model of Data Stewardship“ entwickeln (Element AI, Nesta, 2019).

2.2 Theoretischer Rahmen für Datentreuhandmodelle: Data Governance in einem 3-Ebenen-Modell

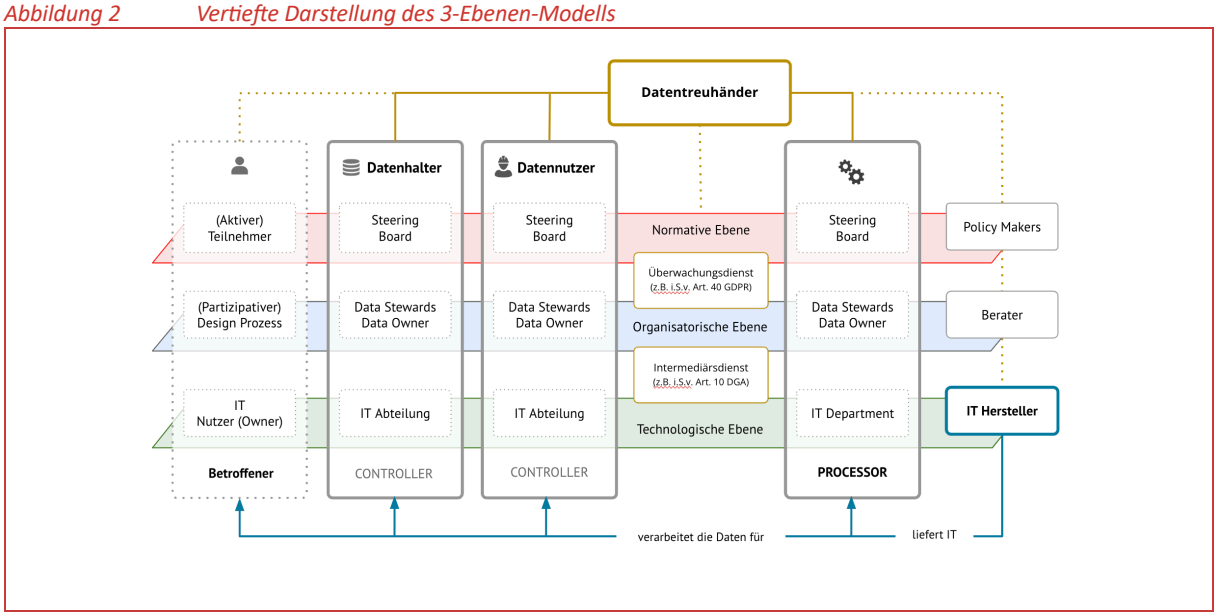
Für die weitere Ausdifferenzierung von Datentreuhandmodellen folgen wir dem **Data Governance-Framework** von v. Grafenstein (2021) als theoretischem Rahmen, der analytisch zwischen drei grundsätzlichen Data Governance-Ebenen unterscheidet: Der normativen, der organisatorischen und technologischen Ebene. Die Herausforderung für erfolgreiche Datentreuhandmodelle besteht nach unserer Arbeitshypothese darin, die Beiträge der auf den unterschiedlichen Ebenen handelnden Akteure so zu koordinieren, dass aus Sicht jedes Akteurs, der mit dem Teilen der Daten erzielte Mehrwert die verbundenen Risiken und Kosten deutlich übersteigt. „Mehrwert“ ist dabei nicht ausschließlich monetär zu verstehen. Weitere Motive zum Datenteilen, bzw. zur Teilnahme an DT-Modellen, sind u.a. Altruismus, PR und Branding, oder Auslagerung von F&E Aktivitäten und sogar Compliance Anforderungen (vgl. QT 4). Eine zusätzliche Herausforderung erfährt diese Koordinationsleistung dadurch, dass sich der Wert und die Risiken nicht statisch etwa zum Zeitpunkt des Teilens der Daten, sondern primär dynamisch nach dem jeweils verfolgten, üblicherweise erst später als erfolgversprechend entdeckten Nutzungszweck bestimmen lassen.

Abbildung 1 3-Ebenen-Modell der Data Governance



Quelle: Eigene Darstellung

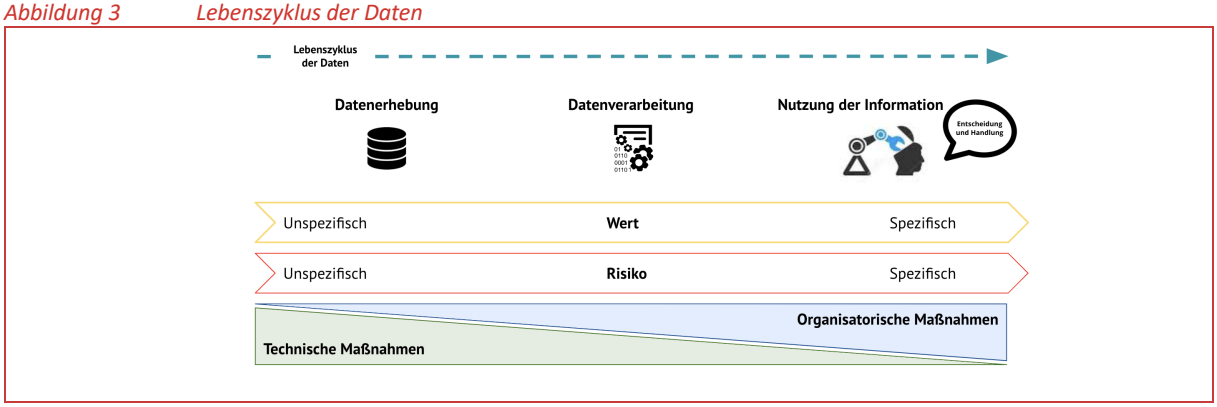
Abbildung 2 Vertiefte Darstellung des 3-Ebenen-Modells



Quelle: Eigene Darstellung

Die in den folgenden Abschnitten dargestellten Querschnittsbereiche analysieren diese Herausforderungen daher in ihren wechselseitigen Abhängigkeiten anhand des typischen **Data Life Cycles** von der Erhebung zum Teilen der Daten bis zur Nutzung der aus ihnen gewonnenen Informationen.

Abbildung 3 Lebenszyklus der Daten



Quelle: Eigene Darstellung

Im Folgenden fassen wir zentrale Erkenntnisse der vorhandenen Literatur zu den Querschnitts-Fragen der Begleitforschung zusammen und zeigen etwaige Lücken im Erkenntnisstand auf. Im weiteren Verlauf der Begleitforschung wird die Literaturanalyse fortgesetzt.

2.3 Technische Infrastruktur und Datensicherheit

In der (insbesondere internationalen) Literatur sind, wie bereits eingeführt, verschiedene Konzepte zu finden, die bisher nicht eindeutig voneinander abgegrenzt wurden. Trotz der unterschiedlichen technischen Strukturen sind die **technischen Anforderungen** der verschiedenen Konzepte sehr ähnlich. Grundsätzlich sind drei Hauptanforderungen zu identifizieren (IDS, 2021):

1. Datenübertragung und Interoperabilität sicherstellen,
2. Datensicherheit und -souveränität sicherstellen,
3. Datenwertschöpfung ermöglichen durch Qualitätssicherung und Verwaltung der Metadaten.

Um diese Anforderungen zu erfüllen, müssen unterschiedliche **technische Bausteine** eingesetzt werden.

Die **Datenhaltung** bildet den Grundbaustein und ist daher der größte Einflussfaktor, der sowohl die Architektur als auch die Funktionalität der technischen Infrastrukturen beeinflusst. Je nach dem Grad der Zentralisierung gibt es zwei Hauptstrukturen der Datenhaltung: zentralisierte (AZT Automotive, 2019; ODI Wildlife, 2019; Young, 2019; Lomotey, 2022) und dezentrale Speicherung (Mills, 2019; ODI Food Waste, 2019; Grossman, 2018).

Bei der **zentralen Datenspeicherung** werden die Daten gesammelt und zentral gespeichert und verwaltet. In der Regel wird für diese Struktur eine zentrale Datenbank eingesetzt, in die Datenanbieter ihre Daten hochladen können. Statische Daten, wie z.B. bestehende Wildtierdaten (ODI Wildlife, 2019), werden passiv gesammelt. In diesem Fall wird den Datengebenden ein Portal zur Verfügung gestellt, über das sie ihre Daten in die Datenbank hochladen können. Bei dynamischen Daten, wie z.B. Betriebsdaten autonomer Fahrzeuge, werden die Daten selbstständig an die Datenbank gesendet. Mit dem Abonnement- und Benachrichtigungsmechanismus werden die Daten beim Auftreten bestimmter Ereignisse an eine zentrale Datenbank gesendet (AZT Automotive, 2019).

Bei der **dezentralen Datenspeicherung** werden die Daten über die lokale Speicherumgebung jedes Datenanbieters verteilt, während die Zugriffsmethoden auf die Daten zentral gespeichert werden. Allerdings ist die Implementierung sehr unterschiedlich. Der größte Unterschied liegt in der Art und Weise, wie die Daten für die Nutzenden bereitgestellt werden. Grossman et al. (2018) weisen darauf hin, dass die Bereitstellung von APIs für den Datenzugriff notwendig ist, um einen dezentralen Datenaustausch zu ermöglichen, d.h. der Datenanbietende muss die Daten in Form von Ressourcen konsolidieren und APIs für den Zugriff auf die Ressourcen sowie die entsprechende Dokumentation bereitstellen. Unter dieser Voraussetzung haben DT zwei Hauptverantwortungen. Zum einen erfassen DTs die Metadaten der auszutauschenden Daten, um die Datenerkennung und -abfrage zu erleichtern. Zum anderen erstellen und behalten sie die Persistent Identifier (PID) der einzelnen Daten, um den Datenzugriff zu erleichtern und die Mehrdeutigkeit zu vermeiden (Grossman et al., 2016).

Jedoch sind die oben geforderten APIs für viele bestehende Datensätze meistens nicht vorhanden. Aus diesem Grund liegt eine weitere Aufgabe der DTs darin, diese Infrastruktur nachzurüsten. Ein konkretes Beispiel stellt die Schnittstelle Data eXchange Controller (DXC) eines Softwareunternehmens Databroker dar. Mit Hilfe dessen können die Datengebenden ihre lokalen Daten mit APIs zur Verfügung stellen. Den Zugriff zum API bekommen die Datennutzenden über die DTs, in diesem Fall die Databroker.

Die zwei Arten der Datenhaltung weisen ihre eigenen Vorteile in Bezug auf **Interoperabilität** und **Vertrauen** auf. Bei einer zentralen Datenhaltung ist die Interoperabilität leichter zu gewährleisten. Da die gemeinsam zu nutzenden Daten zentral in einer Datenbank gespeichert werden, können durch die Verwendung eines standardisierten Datenmodells die Unterschiede zwischen verschiedenen Datenanbietern deutlich überbrückt werden (AZT Automotive, 2019; i4Trust, 2021). Dies führt dazu, dass alle in der zentralen Datenbank gespeicherten Daten die gleiche Datenstruktur aufweisen. Infolgedessen können auch unterschiedliche Datennutzende die standardisierten Daten nutzen.

Die zentrale Datenhaltung zeigt Vorteile im Sinne von Interoperabilität, während die dezentrale Datenhaltung mehr Vertrauen bei einem Datenaustausch schaffen kann. Erstens steht der Datenaustausch immer unter der Kontrolle des Datengebenden. Jede Datennutzung muss auf das lokale System des Datenanbietenden zugreifen. Dabei hat der Datengebende die Möglichkeit, den Datenfluss eigenständig zu kontrollieren. Zweitens gewinnt die dezentrale Datenhaltung mehr Resilienz gegenüber potenziellen Angriffen, denn eine Datenverletzung auf einem einzelnen Knoten führt nicht zum Verlust von anderen Daten (Lindner, 2023).

Abgesehen von den Unterschieden der Datenhaltung gestaltet sich die **Verwaltung der Datenzugriffsrechte** bei allen Datenaustauschmodellen recht einheitlich. Im Allgemeinen wird dies durch die Zusammenarbeit von zwei technischen Bausteinen erreicht: die Access-Control-Komponente (AC) und die Identity-Management-Komponente (IDM) (Lomotey, 2022; IDS, 2021).

Die AC ist für die Autorisierung zuständig, die den Zugriff auf die einzelnen Daten beschränken kann. In einem Anwendungsfall der Logistikbranche darf ein Logistikunternehmen beispielsweise nur auf die der Lieferungsaufträge zugreifen, an denen es direkt oder indirekt beteiligt ist. Nach einer Authentifizierung des Besuchers wird nur der Zugriff gestattet, der den Berechtigungsanforderungen entspricht (i4share, 2022).

Die Authentifizierung erfolgt dann über die IDM, die die Identität der Datennutzenden sowie ihrer Zugriffsrechte prüfen kann. In der Regel müssen sich alle Akteure beim IDM registrieren und erhalten einen Global-Unique-Identifier (GUID) als ihre Identität (IDS, 2021). Der IDM wird dem AC bei jedem Zugriffsversuch mitteilen, ob der Datennutzende auf die entsprechenden Daten zugreifen darf.

Neben der Kontrolle der Zugriffsrechte besteht noch eine weitere Herausforderung, die **Datensouveränität** zu gewährleisten, d.h. sicherzustellen, dass die Datennutzenden die Daten nach Erhalt nicht missbrauchen. Diese Herausforderung ist besonders wichtig für die Anwendungsfälle mit sensiblen Daten (z.B. personenbezogene Gesundheitsdaten). In diesem Zusammenhang versuchen Zrenner et al. (2019), die sogenannte Policy-Enforcement-Komponente (PEC) auf den Automobilsektor anzuwenden. Die Datennutzung wird durch die PEC eingeschränkt. Je nachdem, wo die PEC eingesetzt wird, können unterschiedliche Regulierungsniveaus erreicht werden. Darüber hinaus weisen Zrenner et al. (2019) darauf hin, dass die Art der Datenspeicherung (zentral oder dezentral), wie oben beschrieben, auch ein weiterer wichtiger Hebel für die Datensouveränität ist. Zudem ist die Prüfung und Kontrolle der Datennutzung auch ein wichtiger Aspekt für die Datensouveränität. Zu diesem Zweck kann die Blockchain-Technologie die Transparenz und Nachvollziehbarkeit der Datennutzung erhöhen und damit die Datensouveränität verstärken (Lomotey et al., 2022). Allerdings weisen die derzeitigen technischen Methoden für die Datensouveränität noch viele Einschränkungen auf (Lomotey et al., 2022; Zhang, 2021).

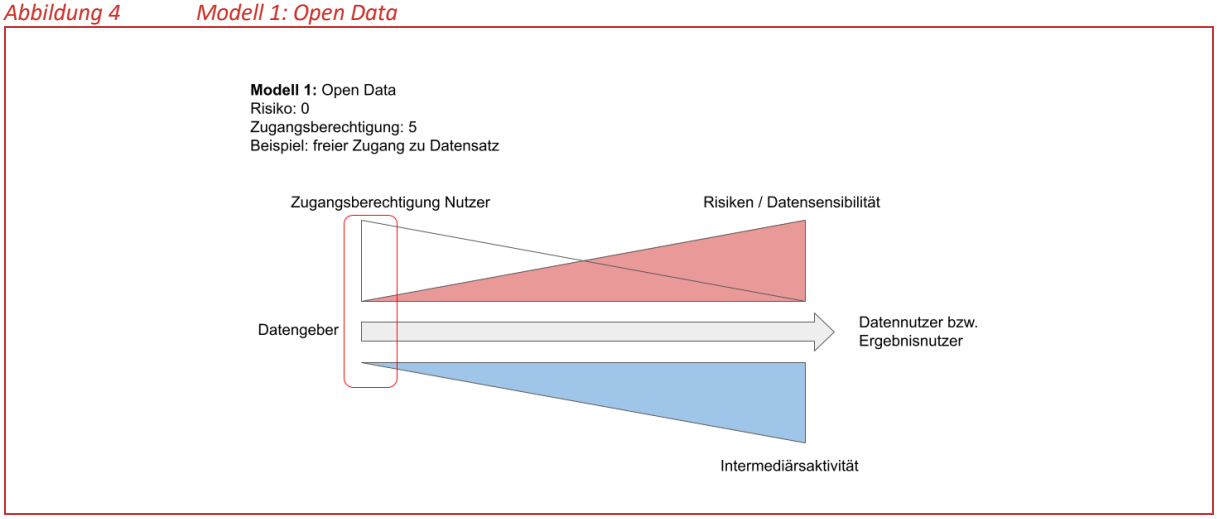
Zusammenfassend lässt sich feststellen, dass die Art und Weise der Datenspeicherung der wichtigste Faktor für die Wahl der technischen Infrastruktur des DT ist. Zurzeit ist die zentrale Datenhaltung noch die vorherrschende Methode. Dezentralisierte Speichermodelle müssen noch weiterentwickelt und erforscht werden. Darüber hinaus stehen für die Interoperabilität und die Datenzugriffskontrolle bereits zahlreiche und relativ ausgereifte technische Bausteine zur Verfügung. Schließlich lässt sich folgender **Forschungsbedarf** feststellen: die technischen Maßnahmen zum Schutz der sensiblen Daten, zur Kontrolle der Datennutzung, und zur Gewährleistung der Datensouveränität.

2.4 Rechtliche Rahmenbedingungen und Ausgestaltung der Datentreuhandmodelle

Unsere Arbeitshypothese geht von der in Abschnitt 2.2 angesprochenen Problematik des **Wert-Risiko-Dilemmas des Datengebenden** aus. Intermediäre treten auf beiden Seiten dieser Kosten-Nutzen-Rechnung auf. Zum einen besteht das Potential, dass sie vor allem durch eine spezialisierte Organisationsstruktur helfen, die Wertschöpfung zu konkretisieren, zu erhöhen oder zu realisieren. Zum anderen können sie helfen, die Risiken zu minimieren. Schließlich kann eine Skalierung prozessualer Intermediärstätigkeiten zur Kostensenkung im Bereich des sicheren Datenteilungsprozesses führen.

Wir schlagen vor, Datentreuhandmodelle ausgehend vom Wert-Risiko-Dilemma in **drei Modelle** zu unterteilen, um eine genauere Vergleichbarkeit zu gewährleisten. Die Modelle unterscheiden sich dabei in der Qualität der Zugangsberechtigung des Nutzers auf die Daten. Die Zugangsberechtigung ist am höchsten, wo die (Compliance-) Risiken des Datengebenden am niedrigsten sind, vgl. **Modell 1** (siehe Abbildung 4). Einer Intermediärstätigkeit bedarf es hier nur ggf. bezüglich organisatorischer Elemente, dagegen weniger bis gar nicht in der Stellung als Sicherheitsgarant, vorausgesetzt die Daten sind schon anonymisiert bzw. an ihnen bestehen keine anderweitigen Schutzrechte.

Abbildung 4 Modell 1: Open Data

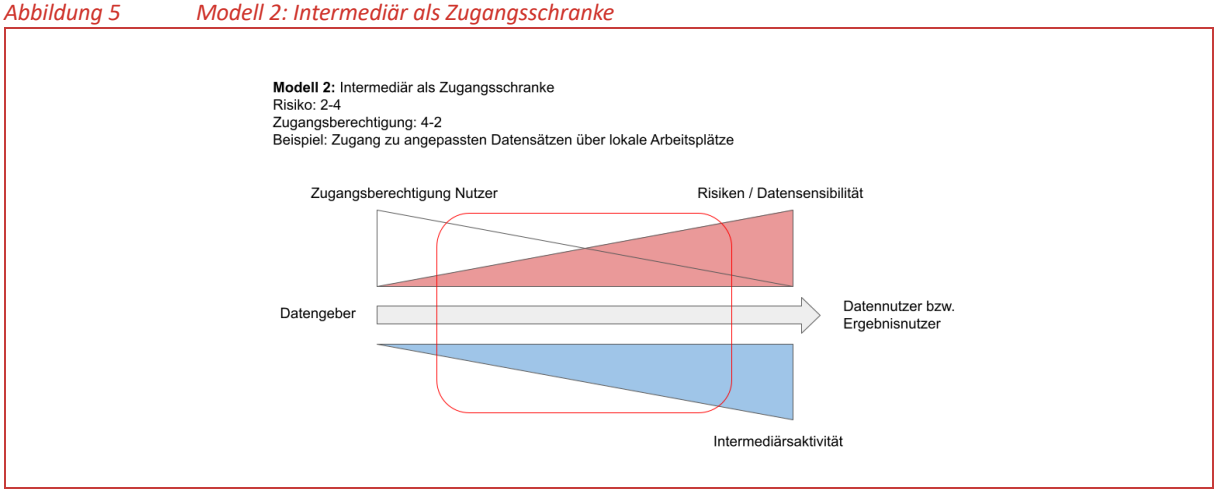


Quelle: Eigene Darstellung

In den Use Cases, die dem **Modell 2** zuzuordnen sind, steigen die (Compliance-) Risiken des Datengebenden an. Hierbei kann es sich um Daten handeln, denen durch rechtliche Einordnung besonderer Schutzcharakter zukommt, z.B. Geschäftsgeheimnisse, personenbezogene Daten, etc. Indirekt proportional zum Anstieg der Risiken kann typischerweise die Zugangsberechtigung des Datennutzenden eingeschränkt werden. Der Datengebende verlangt bzw. benötigt hier gesteigerte Schutzmechanismen, die ggf. durch den Intermediär sichergestellt werden. Hierbei lassen sich je grundsätzlich nach technischer und rechtlicher Einschränkung des Zugriffs auf die Daten verschiedene Sicherheitsstufen unterscheiden:

- Stufe 1: Formale Prüfung der Nutzungsberechtigung
- Stufe 2 (zusätzlich zu Stufe 1): Juristische Instrumente, z.B. Nutzungsvereinbarungen
- Stufe 3 (zusätzlich zu Stufe 1 und 2): Technische Instrumente, z.B. Zugang nur über lokale Arbeitsplätze

Abbildung 5 Modell 2: Intermediär als Zugangsschranke



Quelle: Eigene Darstellung

Zuletzt sehen wir als besonderes Modell mit niedrigster Zugangsberechtigung das **Modell 3**. Hierbei hat der Nutzer gerade keinen Zugriff auf die Daten. Vielmehr kann er allein Anfragen an den Datengebenden stellen. Der Intermediär kann dabei auf verschiedene Weise den Datenverarbeitungsprozess leiten bzw. überwachen. Er prüft die für das Vorliegen der zur Beantwortung der Anfrage notwendigen Daten bzw. Datenqualität. Unter Verwendung dieser Daten modelliert er ein Ergebnis passend zur Anfrage des Nutzers oder gibt dem Datengebenden unterstützende Anweisungen, damit dieser die Anfrage selbst beantworten kann und überprüft

lediglich das Ergebnis. Schlussendlich eröffnet er dem Nutzer das Ergebnis des Prozesses. Der Nutzer hat allein Zugriff auf das Ergebnis und wird deshalb als “Ergebnisnutzer” eingeordnet.

Entlang dieser Systematisierung werden wir in der Begleitforschung die Pilotprojekte danach untersuchen, auf welche Weise sie den Zugang zu Daten rechtlich, technisch und organisatorisch herstellen bzw. unterstützen. Schwerpunkt wird bei dieser Untersuchung sein, wie die untersuchten Datentreuhandmodelle auf diese Weise die Einhaltung von auf das Teilen der Daten anwendbaren gesetzlichen Vorschriften (allen voran das Datenschutzrecht) sicherstellen oder zumindest unterstützen. Ein wichtiger Baustein wird dabei wiederum die Analyse von konkreten vertraglichen Vereinbarungen über Datenzugriffsrechte sein.

2.5 Geschäfts- und Betriebsmodellentwicklung

Ziel eines Datentreuhandmodells ist zum einen die Gewährleistung von Kontrollmöglichkeiten, Partizipation, Selbstbestimmung und Teilhabe der Datengebenden gemäß den Datenschutzbestimmungen, zum anderen eine erhöhte wirtschaftliche Datenverwertbarkeit zur Förderung von Innovation und Wertschöpfung (Schneider, 2022). Der Wert von Datentreuhandmodellen (*in Quelle: Data Stewardship*) ergibt sich in Kontexten, wenn die Erhebung, der Zugang und die Nutzung von Daten nicht nur für die einzelnen Datengebenden, sondern auch für die breitere Gesellschaft mit Risiken und Vorteilen verbunden sind (IPPI, 2022). Aus den beiden Perspektiven kann sich ein Spannungsverhältnis ergeben, das in den Geschäfts- und Betriebsmodellen berücksichtigt werden muss.

DT können in Datenökosystemen eine zentrale **Funktion** zur Lösung des skizzierten Spannungsverhältnisses einnehmen, indem sie Datennutzende und Datengebende neutral zusammenführen. Als unabhängige Instanz soll ein DT den Ausgleich unterschiedlicher - und oft widersprüchlicher - Anreize zur Datenbereitstellung und -verarbeitung sowie gemeinsamer Datennutzung ermöglichen, während die Interessen und das Recht auf informationelle Selbstbestimmung aller Stakeholder gewahrt werden (ODI, 2019; Kühling et al., 2020; Blankertz & Specht-Riemenschneider, 2021). DT können so die Kosten für die Verwaltung und die gemeinsame Nutzung von Daten senken und durch partizipative Entscheidungsprozesse Anreize zur Datennutzung und -weitergabe setzen. Durch eine breitere Teilhabe an den Zielen und Erträgen der wirtschaftlichen Datenverwertung kann dies wiederum zu neuer **Wertschöpfung**, z.B. zur Entwicklung neuer Technologien oder einer verbesserten Entscheidungsbasis führen (ODI, 2019; Blankertz & Specht-Riemenschneider, 2021).

Besonders in Konstellationen mit starker Machtasymmetrie oder geringer Konkurrenz, z.B. marktbeherrschende Stellung globaler Plattformen vs. Markteinsteigern, Asymmetrien zwischen Datengebenden (z.B. Verbraucher*innen) und Datennutzenden (z.B. Forschende) können DT von besonderer Relevanz sein (RfII, 2020; Blankertz & Specht-Riemenschneider, 2021). Das **Risiko von Monopolisierungstendenzen** in Datenmärkten findet sich zum einen im Kontext datenschutzrechtlicher Einwilligung in Bezug auf die Verarbeitung personenbezogener Daten gegenüber großen Onlineplattformen, mobilen Apps und digitalen Dienstleistungen (Element AI, Nesta, 2019; Specht-Riemenschneider & Kerber, 2022). Letztere zeichnen sich sowohl durch Marktmacht (aufgrund von Netzwerk- und Skaleneffekten) als auch durch Informationsasymmetrien aus, die wiederum (unter anderem) durch eine Informationsüberlastung der Nutzer*innen entstehen. Beide Tatbestände des Markversagens verstärken sich gegenseitig. Hier können z.B. PIMS dazu beitragen, dass Nutzer*innen die Verarbeitung ihrer Daten besser kontrollieren können und wirken so dem Informationsüberlastungsproblem entgegen (Specht-Riemenschneider & Kerber, 2022). Zum anderen finden sich Monopolisierungsrisiken auch im B2B Umfeld. Im Automobilsektor beispielsweise hemmt die „Gatekeeper- Position der Autohersteller“ Wettbewerb und Innovationen. Durch die neutrale Instanz einer Datentreuhand könnte hier eine breitere Nutzung von Mobilitätsdaten (z.B. Sensorik-generierte Daten) erreicht und einer monopolistischen Kontrolle durch einzelne Autohersteller verhindert werden (Specht-Riemenschneider & Kerber, 2022).

Durch eine **Monopolisierung der Datenmärkte** werden Anreize und Möglichkeiten für datengetriebene Geschäftsmodelle massiv eingeschränkt, indem Bedingungen für Datenzugang, Angebote und Preismodelle durch den Monopolisten festgelegt werden (Otto et al., 2022). Neue Governance Modelle wie Datentreuhänder können dieses Machtungleichgewicht teilweise beseitigen und einen offeneren Datenzugang unter Datenschutzvoraussetzungen wie auch die Entwicklung eines fairen Markts fördern (Element AI, Nesta, 2019). Dafür muss vermieden werden, durch Lock-in Effekte oder Datensilos eine monopolähnliche „Supertreuhand“ entstehen zu lassen (Schneider, 2022; Otto et al., 2022)

Die **Finanzierung eines DT** kann durch folgende Kanäle erfolgen (ODI, 2019; Blankertz et al., 2020; Element AI, Nesta, 2019): (a) Öffentliche Finanzierung (b) Finanzierung durch Datennutzende auf Basis von Lizensierung oder

Gebühren für Datennutzung oder -dienstleistungen, oder (c) Finanzierung durch Interessengruppen oder nicht-staatliche Organisationen (z.B. philanthropische Spenden, Vereinsmitgliedsgebühren). Für eine öffentliche Finanzierung müsste der Staat eine Begründung des staatlichen Eingriffs vorlegen. Eine Umlegung der Kosten einer DT auf das Kollektiv wird kritisch beurteilt (Blankertz & Specht-Riemenschneider, 2021). Jedoch lassen sich bislang noch keine tragfähigen privatwirtschaftlichen Finanzierungsmodelle am Markt erkennen (Schneider, 2022).

Auch die **Bepreisung von Daten** stellt der Literatur zufolge, insbesondere im C2B-Kontext, noch eine große Herausforderung dar. So zeichnen sich bislang keine verlässlichen ökonomischen Modelle zur Abschätzung vom Wert bestimmter Daten ab, entsprechend mangelt es an Datenbewertungsstandards und -instrumenten (Kühling et al., 2020). Weiterhin sei der Wert von Daten in hohem Maße subjektiv (Otto et al., 2022). Grundsätzlich wird empfohlen, Geschäftsmodelle nicht auf die Bepreisung der Daten selbst zu stützen, sondern auf Gebühren. Hier lässt sich zwischen Subskriptionsmodellen, Fixpreismodellen, Gebühren pro Nutzung („pay-per-use“), Paketpreismodellen, Mitgliedschaften, Transaktionsgebühren und Gebühren für Dienstleistungen unterscheiden (Lindner & Straub, 2023; Otto et al., 2022). Eine vom Datenvolumen abhängige Bepreisung wird kritisch gesehen, da diese Anreize schafft ein erhöhtes Datenvolumen zu „verkaufen" (Blankertz & Specht-Riemenschneider, 2021). Auf welche Dienstleistungen sich konkret Gebühren erheben lassen hängt in erster Linie von den wahrgenommenen Funktionen der DT ab.

Funktionen bzw. Geschäftsmodellkomponenten eines DT können folgende Aktivitäten umfassen, wobei mit der Einführung des Data Governance Act der Europäischen Union Grenzen gesetzt werden: (1) Datenverwaltung, z.B. Speicherung, Zugriffsrechte/Datenschutz, Einwilligung, Ermächtigung und Widerrufe sowie (2) Datenaufbereitung und -auswertung und (3) angrenzende Dienstleistungen, z.B. Anonymisierung, Pseudonymisierung, Verschlüsselung, Verknüpfung/Mapping, Veredelung, Datenqualitätsprüfung, Schulungen, Beratungsleistungen, Onboarding-Prozesse für neue Akteure, Angebot diverser Soft- und Hardwaretools und Bereitstellung von IT- Infrastruktur. (Arlinghaus et al., 2021, Blankertz et al., 2020; ODI, 2019; Lindner & Straub, 2023). In einem Datenraum, wie oben definiert, übernimmt die Governance Ebene bzw. der Föderator Funktionen wie z.B. das Identitätsmanagement, Zertifizierung und die Orchestrierung des Ökosystems/Datenraums im Allgemeinen (Otto et al., 2022).

Letztlich ist die Wahl eines geeigneten Finanzierungsmodells stark von Faktoren wie dem Zweck des DT, dem Typ und der Anreize von Datenbereitstellenden, sowie der Arten der Nutzung und der Nutzenden abhängig. So schränken beispielweise die Bedingungen, unter denen Datengebende ihre Daten des DT bereitstellen, die möglichen kommerziellen Dienstleistungen des DT ein (ODI, 2019).

Unter dem Data Governance Act (DGA) darf ein DT keine unternehmerischen Interessen verfolgen, das heißt, es dürfen keine wirtschaftlichen Gewinne durch die Nutzung der Daten erzielt werden. Entsprechend ist lediglich eine Refinanzierung durch die „Verwaltung“ der Daten / des Datenökosystems / des Datenraums möglich (RfII, 2021). **Neutralitätsanforderungen** setzen ferner die Abwesenheit von Abhängigkeiten oder direkten Anbindungen an Wertschöpfungsketten (vertikale Integration) voraus (Blankertz & Specht-Riemenschneider, 2021; Blankertz et al., 2020). Neben den Anforderungen des DGA werden auch potenzielle Interessenskonflikte des DT und die Sicherstellung von Datenschutz als Argumente gegen wirtschaftliche Abhängigkeit und kommerzielles Eigeninteresse an der Datenverwertung angeführt und von der breiten Mehrheit in der Literatur unterstützt (Kühling et al., 2020; Lind, Suckfüll, 2013; Stevens & Boden, 2022).

Auf der anderen Seite schränken die Neutralitätsanforderungen den Spielraum von Geschäftsmodellansätzen in Bezug auf Datenauswertung oder Daten bestehender Geschäftsbereiche stark ein. Ferner wird eine „Überregulierung“ durch den DGA kritisiert, dessen Anforderungen laut Autor*innen zum Teil über die der DSGVO hinausgehen (Stevens & Boden, 2022). Auch wird das Risiko einer Unternutzung der DT betont, im Fall eines zu passiven DT. Entsprechend sei eine anwendungsabhängige Prüfung von Neutralitätsbedingungen sowie eine Ausgestaltung in Abhängigkeit der individuellen Zielsetzung sinnvoll (Blankertz & Specht-Riemenschneider, 2021; Blankertz et al., 2020). Im Kontext von Forschungsdatenzentren ergibt sich beispielsweise ein potenzieller Widerspruch zwischen Neutralitätsanspruch und Forschungstätigkeit (RfII, 2021).

DT können grundsätzlich durch staatliche, nicht-profitorientierte, gemeinnützige, genossenschaftliche oder privatwirtschaftliche Einrichtungen **betrieben** werden (Arlinghaus et al., 2021, Lindner, 2021). Die Einschränkung der möglichen wahrgenommenen Funktionen im Rahmen des Data Governance Act schränken wiederum eine profitorientierte, privatwirtschaftliche Lösung jedoch erheblich ein (Schneider, 2022; Element AI, Nesta, 2019).

Hier ist den Autor*innen zufolge eine Anschubfinanzierung durch den Staat oder gemeinnützige Organisationen nötig. Nach Etablierung könne eine Kostendeckung der laufenden Ausgaben durch die Erhebung von Gebühren an Datennutzende erfolgen (Schneider, 2022; Element AI, Nesta, 2019).

Neben der Bepreisung von Daten und Dienstleistungen im Rahmen eines DT-Finanzierungsmodells findet sich in der Literatur auch eine Diskussion zu monetären **Anreizen** bzw. Vergütung von datengebenden Individuen (B2C Kontext). Hier herrscht ein breiter Konsens darüber, dass durch eine solche Vergütung konträre soziale Effekte zu erwarten sind. Ein solches Modell fördere die soziale Ungleichheit, indem Datenschutz und Privatsphäre an die finanziellen Möglichkeiten der Datengebenden gekoppelt werden und Anreize für eine erhöhte Datenbereitstellung gesetzt werden (Schneider, 2022; DSSC, 2022). Ferner sei eine solche Datenbepreisung nicht sinnvoll, da die Daten eines einzelnen Individuums ohnehin keinen großen Mehrwert haben (DSSC, 2022). Eine freiwillige und auf sozialem Mehrwert basierende Bereitstellung von Daten durch Individuen (z.B. im Kontext von Gesundheitsdaten und -forschung) sei somit vorzuziehen (Element AI, Nesta, 2019).

Das aktuelle Anbieterumfeld staatlicher DT besteht hauptsächlich aus Treuhandstellen von Hochschulen oder anderweitigen Einrichtungen, deren Aktivitäten auf eine sichere Verwahrung und Bereitstellung von Forschungsdaten (oftmals aus der Medizin, z.B. Biobanken, Krebsregister) abzielen (Arlinghaus et al., 2021). Der Kund*innenkreis umfasst weitestgehend den Hochschulsektor sowie die Pharmaindustrie. Universitäre DT handeln zwar weitgehend autonom, sind jedoch organisatorisch und finanziell nicht unabhängig (Arlinghaus et al., 2021). Das Feld der nicht-staatlichen DT lässt sich in non-profit Organisationen (z.B. Zusammenschluss verschiedener Akteur*innen in einem Verein) und for-profit Organisationen unterteilen. Letztere sind z.B. kommerzielle PIMS, untern denen sich jedoch bislang auch noch keine etablierten Konzepte für eine nachhaltige Finanzierung abzeichnen und auch die Frage der Neutralität und der konträren Anreizsetzung, wie oben beschrieben, scheinen noch offen (Langford et al., 2020). Ferner befinden sich diverse Datenräume für die Industrie in verschiedenen Sektoren im Aufbau, insb. auf EU-Ebene (Schneider, 2022).

Im weiteren Verlauf der Begleitforschung sollen Anforderungen der Ausgestaltung von Datentreuhandmodellen weiter präzisiert werden. Die gesichtete Literatur deutet bereits darauf hin, dass es je nach Zweck des DT, den Bedarfen und den Erfordernissen sowie den Interessen der Akteure unterschiedlicher Ausgestaltungen bedarf (Element AI, Nesta, 2019).[3] Das größte übergeordnete Hemmnis bei der Etablierung von Treuhandmodellen scheint in einem nachhaltigen Finanzierungskonzept zu liegen, welches zudem den Neutralitätsanforderungen des Data Governance Act (DGA) gerecht werden muss.

2.6 Akzeptanz, Skalierung und Transfer

Die **Akzeptanzfrage** stellt sich aus zwei Perspektiven: Welche Faktoren entscheiden, ob Datengebende (Privatpersonen; Firmen, Sonstige) ihre Daten teilen; welche Aspekte, ob (potentielle) Datennutzende die Daten auch verwenden würden. Da die Forschung zu DT noch jung ist, wird zur Beantwortung dieser Fragen auch die breitere Literatur zu Data Sharing herangezogen.

Privatpersonen als Datengebende. Umfragen weisen darauf hin, dass viele Menschen in Deutschland und anderen Ländern grundsätzlich bereit sind, auch hochsensible Daten (z.B. Gesundheitsdaten) für „gute Zwecke“ wie medizinische Forschung oder Dekarbonisierung, zu teilen (Köngeter et al. 2022; Mohr & Cloos 2022; Meijer & Potjer 2018; ODI, 2019, Acharya & Mekker, 2022). Allgemein scheint die Bereitschaft etwas höher, Daten mit der Universitäts- als der Industrieforschung zu teilen. Es besteht aber durchaus auch die Bereitschaft, Daten mit der Industrie zu teilen. Persönlich direkt Betroffene (z.B. Krebspatienten) oder ideell motivierte (z.B. ökologisch Sensibilisierte) teilen eher als die Gesamtbevölkerung (Köngeter et al., 2022; Mohr & Cloos, 2022). Normative

[3] Während bei Datentreuhändern im Gesundheitssektor vor allem gewährleistet sein sollte, dass Daten nicht für Dritte außerhalb der autorisierten Ärzteschaft zugänglich sind und ausschließlich Forschungszwecken im Sinne des Gemeinwohlinteresse dienen, wird im Onlinesektor vor allem eine Lösung benötigt, die Nutzer*innen eine bessere Kontrolle über ihre Daten ermöglicht. Im Automobilsektor wiederum hemmt die „Gatekeeper- Position der Autohersteller“ Wettbewerb und Innovationen.

Erwägungen scheinen für Privatpersonen eine wichtige Motivation zum Teilen zu sein; Hoffnung auf eigene Vorteile aber auch (Meijer & Potjer, 2018; Köngeter et al., 2022; ODI, 2019; Choi, 2022). Es gibt Hinweise, dass Individuen eher bereit sind, Daten zu teilen, wenn dies als solidarische Pflicht der Allgemeinheit gegenüber verstanden wird, und andere ihre Daten ebenfalls teilen (Köngeter et al., 2022; Choi, 2022). Die wichtigsten Hemmnisse, die Privatpersonen vom Datenteilen abhalten, sind Datenschutzbedenken und sonstige Sorgen vor Missbräuchen durch Firmen oder den Staat, sowie der zum Teilen nötige Aufwand (Blankertz, 2020; Köngeter et al., 2022; Acharya & Mekker, 2022; Meijer & Potjer, 2018; Choi et al., 2022). Je größer der nötige zeitliche, kognitive oder materielle Aufwand und je komplexer der Prozess, desto weniger teilen Menschen ihre Daten. Dabei „schlägt" Aufwand mitunter Datenschutz: werden Probanden in Experimenten gebeten, zwischen unterschiedlichen Modalitäten der Datenfreigabe zu wählen, sind Modalitäten, die ihnen weniger oder sogar keine Kontrolle über die weitere Verwertung ihrer Daten geben, ihnen dafür aber auch weniger Aufwand bereiten, wesentlich populärer als Modalitäten, die höhere Kontrolle/Datenschutz gewähren aber mehr Aufwand erfordern (Köngeter et al., 2022). Zumindest unter Krebspatienten scheint auch die Wichtigkeit, die die Befragten dem Datenschutz beimessen, zu sinken, wenn sie auf gesamtgesellschaftliche positive Wirkungen von schwächerem Datenschutz, wie die Erleichterung der Forschung, hingewiesen werden (Köngeter et al., 2022). Offensichtlich ist Datenschutz für viele ein gesellschaftliches Ziel neben anderen, aber kein alles überragender Wert.

Gleichzeitig scheint die allgemeine Reputation der als Datentreuhänder fungierenden Stelle eine wichtige Rolle für die Akzeptanz zu spielen: So indiziert das Scheitern von Googles Toronto Sidewalk Labs-Projekt (Scassa, 2020), dass Akteure mit schlechtem Ruf als „Datenkrake" es ungleich schwerer haben, Vertrauen zu erwirtschaften, ungeachtet der rechtlichen oder technischen Details der Datennutzung. Umgekehrt zeigt der Fall der UK Biobank (s. Use Case-Fallstudie), dass Menschen auch hochsensible Daten teilen werden, wenn ihnen der Zweck einleuchtet und der Datentreuhänder vertrauenswürdig erscheint.

Diese Ergebnisse sind konsistent mit dem belastbar attestierten Privacy Paradox. Sie legen nahe, dass es sich bei den in der Bevölkerung verbreiteten Sorgen bzgl. „Datenschutz" oft um eher diffuse Ängste handelt, die nur bedingt von objektiven Fragen des Datenschutzes und der Data Governance abhängen. Entsprechend dürften sie auch nur bedingt durch technische oder rechtliche „Kniffe" aufgelöst werden können. Belastbare Datensicherheit ist zwar eine notwendige aber wahrscheinlich nicht ausreichende Bedingung für Vertrauensaufbau. Da Vertrauens- und Akzeptanzaufbau, gerade seitens Privatpersonen, psychosoziale Prozesse sind, dürften deliberative und partizipative Verfahren, bei denen die Datengebenden bzw. ihre Repräsentanten bei der Gestaltung des Datentreuhänders (Governance, Ziele) mitwirken, von großer Wichtigkeit für seine Akzeptanz sein.

Deliberative und partizipative Prozesse ermöglichen es auch, kontinuierliche Kommunikation mit (potenziellen) Datengebenden zu unterhalten. Das ist wichtig, da aktive Marketing-Kampagnen nötig sind, um Data Sharing zu stimulieren (Meijer & Potjer 2018, Paprica et al. 2020) und da die Erwartungen von privaten Datengebenden oft über belastbaren Datenschutz hinausgehen. Viele wollen über Ergebnisse, die mit ihren Daten erzielt werden, informiert werden (Köngeter et al., 2022; ODI, 2019). Das Gefühl, zu etwas Größerem beizutragen, ist teilweise ein wichtiger Motivator für Data Sharing (Meijer & Potjer, 2018).

Unternehmen als Datengebende. Nur wenige deutsche Unternehmen sind heute bereit, ihre Daten zu teilen (Fraunhofer 2022, BDI 2021, RfII 2021). Es gibt keine Hinweise, dass die Bereitschaft zum Datenteilen in anderen Ländern grundsätzlich größer ist.

Firmen sehen vom Data Sharing ab, weil sie bislang meist wenig direkten Nutzen davon erwarten (bzw. weil Geschäftsmodelle fehlen, die Datenteilen belohnen würden); sie aber Risiken daraus erwachsen sehen. Zu diesen zählen Sorgen um Datensicherheit, Abfluss von Geschäftsgeheimnissen, Compliance-Verletzungen aufgrund unklarer Rechtslagen sowie die Gefahr von Reputationsschäden für die Firma oder die Branche, falls Daten „falsch" interpretiert werden (RffII, 2020, CfDEI, 2021; Brown et al., 2022; FPF, 2017; Blankertz, 2020; Blankertz/Specht-Riemenschneider, 2021). Weitere Hürden sind Aufwand, fehlende Ressourcen und technische Infrastruktur sowie fehlende Data Literacy/Data Skills (RfII, 2020; ODI, 2019; CfDEI, 2021). Es gibt Anzeichen, dass digital-oder daten-affine Firmen und Firmen mit starker Eigenforschung, eher bereit sind, Daten mit Treuhändern oder der Wissenschaft zu teilen (FPF, 2017). Möglicherweise fühlen sich diese Firmen eher im Stande, Risiken abzuschätzen oder Aufwände gering zu halten.

Umgekehrt teilen Unternehmen Daten primär dann, wenn sie darin Vorteile erblicken – u.a. geldliche, durch die direkte Monetarisierung von Daten (selten), Zugang zu den Daten Anderer („pay to play"), Auslagerung von F&E (an die Wissenschaft) oder von Compliance-Verantwortung (an Datentreuhänder), Management von Stakeholder-Beziehungen und PR (FPF, 2017; CfDEI, 2021; ADRF, 2018; ODI, 2019) – und glauben, Risiken und Aufwand kontrollieren zu können. In der Praxis findet Data Sharing bisher meist nur nach sehr umfangreichen und aufwendigen, oft persönlichem, Vertrauensaufbau und Aushandlungsprozessen statt (FPF, 2017; ADRF, 2018).

Akzeptanz durch Datennutzende. Auch für Datennutzende ist die Verarbeitung von durch Dritte bereitgestellte Daten mit Risiken behaftet. Daten und Metadaten können inkorrekt oder mit Biases behaftet sein; „kleine", nicht immer leicht erkennbare Fehler (z.B. inkonsistente Einheiten) können erhebliche Folgen haben (Gal & Rubinstein, 2019). Forschung zur Nutzung geteilter Daten hat sich bislang primär mit Data Sharing in wissenschaftlichen Communities befasst. Wesentliche Faktoren, die die Literatur herausgearbeitet hat, sind Vertrauen in die Qualität und die Aufbereitung der Daten einschließlich Metadaten und Provenance, sowie die Bedienbarkeit der technischen Infrastruktur und Interfaces. Zwar existieren mittlerweile erste Standards und Zertifikate für Datenrepositorien; zumindest unter Wissenschaftler*innen spielen die (oft persönliche oder community-basierte und informelle) Reputation der Personen und Repositorien, die Daten bereitstellen, weiterhin eine sehr große Rolle für die Vertrauensbildung (vgl. Yoon & Lee, 2019 und die darin zitierte Literatur).

Standards und Zertifizierung. Es existieren bereits verschiedene Standards, die für Datentreuhänder relevant sind. Dazu zählen „allgemeine" Standards, wie die für Datensicherheit (z.B. ISO 27001), sowie spezialisiertere, wie die für Datenrepositorien (ISO 16363) und Digitale Archive (DIN 316442) (Yoon & Lee, 2019) oder Daten Provenance (ISO/IEC AWI 5181). Auch für Biobanken gibt es einen ersten Standard (ISO 20387:2018). Über diese hinaus sind uns aber keine weiteren dezidierten formalen Standardisierungsaktivitäten für Datentreuhänder bekannt. Gleichwohl lieferte zumindest eines unserer Interviews Hinweise auf erste Bottom-up getriebene und informelle, de-facto Standardisierungsaktivitäten durch die Datentreuhänder selbst.

Auf Standards können Zertifizierungen aufsetzen. Martin & Pasquale (2019) nennen sieben Dimensionen eines Datentreuhänders, die zertifiziert werden sollten: seine Governance, Finanzierung, Datenzugangsregeln, Datensicherheit, Ethik und Nachhaltigkeit, und Datenqualität. Soweit aus der Literatur ersichtlich, gibt es bislang jedoch keine Institutionen, die sämtliche dieser Dimensionen zertifizieren. Es sind uns auch keine speziell auf Datentreuhänder gemünzte Zertifizierungen bekannt.

Mittelfristig dürfte die Entwicklung von Standards und Zertifikaten eine wichtige Rolle für die Skalierung von Datentreuhändern und ihre Akzeptanz insbesondere auf Seite von Datennutzenden spielen. Hier ist vor allem der Bedarf von Standards und Gütesiegeln für Daten, ihre Aufbereitung und Kuratierung sowie Metadaten und Provenance zu nennen (RfII, 2020; Stalla-Bourdillon, 2021; Gal & Rubinfeld, 2019). Ist die Qualität von Daten nicht belastbar attestiert oder fehlen Metadaten und Kontextinformationen, können sie oft nur begrenzt genutzt werden. Fehlende Daten-, Metadaten- und Kuratierungsstandards verkomplizieren zudem die Zusammenführung verschiedener Datensets erheblich (Stalla-Bourdillon, 2021; Gal & Rubinfeld 2019). Dabei müssen Standards in Teilen wahrscheinlich sektorspezifisch sein (RfII, 2021).

Inwiefern Standards und Zertifizierungen für die Akzeptanz von Datentreuhändern seitens der Datengebenden zum jetzigen Zeitpunkt eine wichtige Rolle spielen können, ist jedoch weniger klar. Datentreuhänder sind eine sehr neue Institution, deren Aufbau regelmäßig komplizierte fallspezifische Fragen aufwirft. Ihr Aufbau braucht daher regelmäßig einen deliberativen Prozess mit sehr umfangreichen Abstimmungen und Verhandlungen unter den beteiligten Akteuren. So werden gemeinsame Verständnisse, Zielsetzungen und Akzeptanz/Vertrauen geschaffen. Insofern einschlägige Standards und Zertifizierungen selbst erst noch erarbeitet und danach von Akteuren akzeptiert und umgesetzt werden müssen (was ebenfalls ein längerfristiger Prozess sein dürfte), ist nicht davon auszugehen, dass sie kurzfristig und quasi „automatisch" Akzeptanz schaffen würden. Zum jetzigen Zeitpunkt dürfte eher die gemeinsame Erarbeitung von Standards und Zertifizierungen durch Akteure akzeptanzstiftend wirken, wobei das akzeptanz-/vertrauensgenerierende Moment vorerst mehr im „Weg" zum Standard als in diesem selbst liegen dürfte.

Gaia-X und PETs. Die Gaia-X Initiative wird in der gesichteten Datentreuhänder-Literatur noch wenig besprochen (aber siehe die Aufsätze in Otto et al. 2022). Indem Gaia-X eine Referenzarchitektur und Dienste bereitstellt, kann sie den Aufbau von Datentreuhändern unterstützen. Zudem bietet die Initiative Raum, um die Aushandlungsprozesse, die für den Aufbau von Datentreuhändern essentiell sind, zu forcieren. Als föderierte

Struktur kann sie schließlich den Zusammenschluss und die Interoperabilität zwischen verschiedenen sich etablierenden Datentreuhändern ermöglichen.

Privacy-Enhancing Technologies (PETs) ist ein Oberbegriff für ein breiteres Spektrum von Technologien und Verfahren, die Datenschutz und Datensicherheit bei der Datenverarbeitung erhöhen sollen (ENISA, 2016). Von besonderer Relevanz für Datentreuhänder sind solche PETs, die Datenauswertung unter Einhaltung hoher Datenschutz- und Datensicherheitsstandards ermöglichen, z.B. Secure Multiparty Computation und föderiertes Lernen (Datenauswertung ohne eigenen Zugriff auf die Daten), Homomorphic Encryption (Auswertung verschlüsselter Daten) und Trusted Execution Environments (Verarbeitung in besonders sicheren Umgebungen (Royal Society, 2019). Neben der geschützten Verarbeitung personenbezogener Daten ermöglichen sie insbesondere auch die sichere Auswertung von Industriedaten, und können so eine Lösung für den Schutz von Geschäftsgeheimnissen bei gleichzeitigem Teilen von Daten darstellen. Sie werden von verschiedenen Autoren als wichtige Komponente für die Akzeptanz von Data Sharing seitens der Datengebenden betrachtet (RfII, 2020; CfDEI, 2020; Stalla-Bourdillon, 2021).

3 Use Cases

3.1 Einführung

Im Folgenden präsentieren wir sechs Use Cases von Anwendungsfällen von Datentreuhandmodellen und ähnlichen Ansätzen aus dem In- und Ausland, die nicht durch das BMBF im Pilotverfahren gefördert sind. Der Fokus liegt hierbei auf derzeit auftretenden Herausforderungen bei der Entwicklung von Datentreuhandmodellen. Bei der Auswahl wurde darauf geachtet, eine möglichst große Bandbreite an Anwendungsfeldern sowie Governance-Modellen, Geschäftsmodellen, Rechtsformen sowie technischen Voraussetzungen zu berücksichtigen. Die Analyse der Use Cases fließt in die weitere Beforschung der vier Querschnittsthemen ein. Erkenntnisse aus den Use Cases wurden außerdem zur Ein- und Abgrenzung des Untersuchungsgegenstands genutzt.

3.2 Use Case 1: Catena-X

Tabelle 1 Steckbrief: Catena-X

Steckbrief	
Verantwortliche Organisation(en)	133 Unternehmen der Automobilindustrie unter dem Dach des Catena-X Automotive Network e.V.
Typ der Datenbereitsteller	Aktive, dezentrale Bereitstellung
Funktionen des Datentreuhänders	Es handelt sich um ein offenes und kollaborativ angelegtes Datenökosystem bzw. die Bereitstellung einer Infrastruktur
Sektor	Automobilindustrie
Gründung/Projektbeginn	2020
Aktueller Entwicklungsstand	Das BMWK-geförderte Forschungsprojekt läuft bis 2024 (Projektlaufzeit: 01.08.2021 – 31.07.2024) (BMWK, o.J.). Erste Entwicklungsergebnisse von Catena-X werden 2023 durch eine erste Betreibergesellschaft (Cofinity-X) an den Markt gebracht. Weitere Betreibergesellschaften sollen folgen.
Rechtsform	e.V.

3.2.1 Hintergrund

Catena-X versteht sich als offenes und kollaborativ angelegtes Datenökosystem. Ziel des laufenden Projekts ist die „Bereitstellung einer Umgebung für den Aufbau, Betrieb und die kollaborative Nutzung durchgängiger Datenketten entlang der gesamten automobilen Wertschöpfungskette“ (Catena-X, 2023a). Aktuell erhält das

Catena-X Forschungskonsortium eine Projektförderung des BMWK. Ziel der Förderung ist die Entwicklung der rechtlichen, operativen und technischen Basis für Services und das digitale Ökosystem. Zudem werden zehn industrierelevante Use Cases erarbeitet, sukzessive erprobt und nutzerorientiert weiterentwickelt (BMWK, o.J.). Abbildung 6 fasst die 10 Anwendungsfälle zusammen (Catena-X, 2023b).

Im Catena-X-Ökosystem sind alle Akteure in durchgängigen Wertschöpfungsketten verbunden. Alle Partner agieren auf Augenhöhe, während die Souveränität über die eigenen Daten gewährleistet wird, sodass Lock-in-Effekte (d.h. ein erschwerter Wechsel zwischen Anbietern und Infrastrukturen) nicht auftreten können. Das Projekt verfolgt ausdrücklich nicht das Ziel der Datenmonetarisierung. Vielmehr sei das Ziel, Daten zusammenzuführen, um neue Wertschöpfungsketten zu schaffen sowie bestehende Prozesse zu optimieren, um so gemeinsam einen Wert durch Anwendungsfälle zu generieren (Bitkom, 2021). Hierbei fungiert Catena-X als „Trust Anchor" für einen freien Marktzugang (SAP, 2021).

Das Catena-X System besteht aus verschiedenen Einheiten: (i) dem **Verein** „Catena-X Automotive Network e.V.", (ii) einer **Entwicklungsumgebung**, die aktuell aus den 18 Partnern des Vorhabens besteht, sowie (iii) einem dezentralen **Ökosystem für Betreiberangebote**. Der Verein ist zuständig für die neutrale Governance, für die Standardisierung und Zertifizierung sowie für die Steuerung des Ökosystems und umfasst direkte Industriepartner wie beispielsweise Hersteller (OEMs), Zulieferer und Recycler sowie indirekte Partner wie Service Provider und Beratungspartner wie z.B. Forschungsinstitute. Im Rahmen von Arbeitsgruppen können Mitglieder das Catena-X-Ökosystem aktiv mitgestalten. Aufgabe der Entwicklungsumgebung ist die Erarbeitung technischer Lösungen, die durch den Verein wiederum zertifiziert werden können. In der Betreiberumgebung werden kommerziellen Dienste und Geschäftsanwendungen von unterschiedlichen Anbietern betrieben und somit entwickelte Lösungen an den Markt gebracht bzw. Nutzern über Marktplätze zur Verfügung gestellt (Catena-X, 2022a). Der Verein soll sich langfristig über Mitgliedsbeiträge finanzieren, die nach Umsätzen gestaffelt sind und so auch für kleinere Unternehmen finanzierbar sind. Die Betreibergesellschaften erheben für ihre Services wiederum Nutzungsgebühren.

Abbildung 6 10 Anwendungsfälle von Catena-X

1. **Nachhaltigkeit**: Standards und Methoden zur Einsparung von CO_2
2. **Rückverfolgbarkeit**: Durchgängige Datenketten über den gesamten Lebenszyklus
3. **Kreislaufwirtschaft:** Daten zur maximalen Nutzung von Ressourcen
4. **Manufacturing as a Service**: Fertigungsunterstützung aus dem Verbund
5. **Business Partner Data Management:** Harmonisierung identischer Daten unterschiedlicher Quellen
6. **Qualitätsmanagement:** Datenbasiert entlang der Wertschöpfungskette
7. **Modulare Produktion:** Datenbasierte Industriestandards zur Optimierung industrieller Fertigung
8. Online-Steuerung und Simulation: Effizienz im Materialfluss
9. **Bedarfs- und Kapazitätsmanagement:** Sicherer Austausch realer Daten bei partnerschaftlicher Kollaboration
10. **Digital Behavior Twins**: Daten- und modellzentrierte Entwicklungs- und Betriebsunterstützung

3.2.2 Konzeption, Herausforderungen, Lösungsansätze

Die Veröffentlichung von Standards durch den Verband zielt darauf ab, Interoperabilität, Datenhoheit und Sicherheit für alle Teilnehmenden im Datenraum zu ermöglichen. Um in dem Ökosystem arbeiten zu können, müssen sich Teilnehmende an die veröffentlichten Standards halten. Die Catena-X-Standards basieren u.a. auf den Konzepten und Prinzipien von Gaia-X bzw. der International Data Space Association (IDSA), sowie auf Industriestandards und Best Practices, erweitert werden diese jedoch auch durch die Einbeziehung von

automobilspezifischen Anforderungen und Anwendungsfällen. Die Zertifizierung von Ökosystemteilnehmern und Softwarekomponenten schafft Transparenz und Vertrauen in das Ökosystem. Eine Zertifizierung einer Softwarekomponente bestätigt beispielweise, dass diese interoperabel, datenschutzkonform und sicher im Ökosystem einsetzbar ist. Durch die Schaffung einer Architektur, von Standards, IDs, Wertschöpfungsstrukturen und semantischen Modelle (u.v.m.) werden technische Grundvoraussetzungen geschaffen, auf denen in der Folge Anwendungen basieren können (Catena-X, 2022a).

Die Architektur des Catena-X-Datenraums umfasst drei Kernbereiche (Catena-X, 2023c):

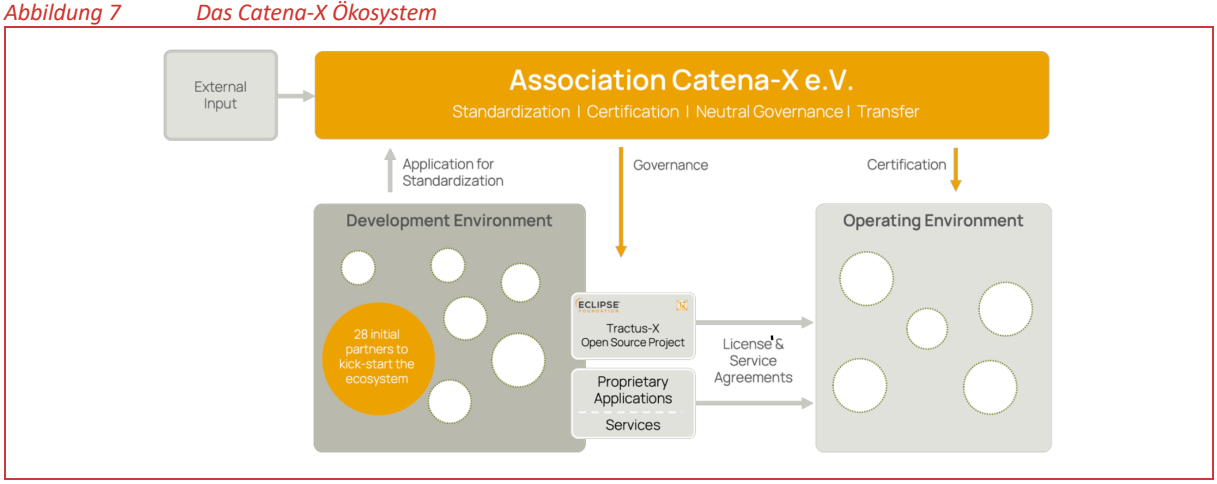
1. Kerndienste stellen die Basisfunktionen des Catena-X-Datenraums sicher (z.B. den Marktplatz oder das Digital Twin-Register)
2. Dienste für Datenanbieter und -anwender (wahlweise als Managed Service über einen Enablement Service Provider) zur Bereitstellung und Verarbeitung von Daten innerhalb des Catena-X Datenraums
3. Geschäftsanwendungen für spezifische Unternehmenslösungen

Die Situation der Branche vor Catena-X zeichnete sich durch Kooperationen aus, die sich maximal auf eine Lieferkette bezogen und zumeist einen Peer-to-Peer-Charakter aufwiesen. Ferner wies die Branche eine Vielzahl proprietärer Ökosysteme und Standards auf, der Datenaustausch hing von Vermittlern ab. Das Ökosystem von Catena-X soll nun die gesamte Wertschöpfungskette abbilden, ein offenes Multi-Vendor-Ökosystem schaffen und den Akteuren Souveränität über ihre eigenen Daten ermöglichen (Catena-X, 2022b).

Catena-X soll im Kern 5 **Mehrwerte** für die beteiligten Akteure schaffen: (i) schnellerer Prozess zur digitalen „Reife“ und zur Wertschöpfung durch einen für KMU geeigneten, ganzheitlichen Onboarding-Service, (ii) eine „Community of Trustees“, ein Branchendatenraum mit gemeinsamen Werten, Grundsätzen und neutraler Führung, (iii) Bibliothek „digitaler Zwillinge“ mit umfassenden Datensätze über alle Partner hinweg, die z.B. durch eine gemeinsame Semantik vereint sind, (iv) aktive Datenketten, d.h., verbundene digitale Zwillinge und Mitglieder, einschließlich Gebäude- und Behördendienste, sowie (v) Software Development Kits und Dienste für ein offenes Ökosystem: Förderung der Zusammenarbeit und Nutzung, Nährboden für Innovationen Catena-X (2023d).

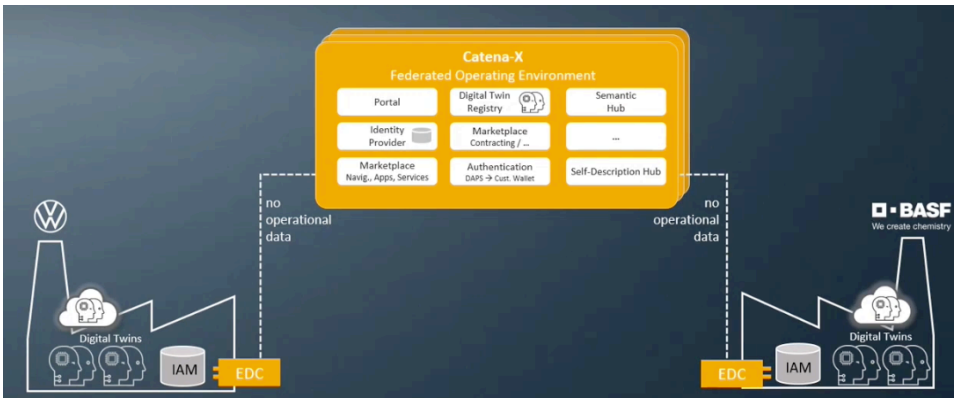
Im Anwendungsfall Nachhaltigkeit ermöglicht Catena-X beispielsweise einen standardisierten Datenaustausch innerhalb der Lieferkette zur Berechnung von CO_2-Emissionen oder zur Prüfung der Vorgaben des Lieferkettengesetzes. Standardisierte Verfahren dieser Prozesse zur Datenübermittlung ermöglichen den Unternehmen in diesem Kontext hohe Effizienzgewinne. Ein wesentlicher Faktor für den zukünftigen Erfolg von Catena-X ist eine industrieübergreifende Teilnahme am Ökosystem.

Abbildung 7 Das Catena-X Ökosystem



Quelle: Catena-X (2022a).

Abbildung 8 Betriebsumgebung von Catena-X



Quelle: Catena-X (2023d).

3.3 Use Case 2: EuroDaT

Tabelle 2 Steckbrief: EuroDaT

Steckbrief	**EuroDaT**
Verantwortliche Organisation(en)	Das EuroDaT Konsortium besteht aus Atos, d-fine, Deloitte, Deutsches Forschungszentrum für Künstliche Intelligenz, Goethe-Universität Frankfurt, Hessisches Ministerium für Wirtschaft, Energie, Verkehr und Wohnen, Lexemo, TechQuartier, T-Systems, Universität des Saarlandes und dem Zentrum Verantwortungsbewusste Digitalisierung (ZEVEDI). Das EuroDaT Konsortium wird von d-fine als Konsortialführer koordiniert.
Typ der Datenbereitsteller	Aktive Datenbereitstellung
Funktionen des Datentreuhänders	Bereitstellung der Infrastruktur, Vertrauensbildung als neutraler Akteur, technische Überprüfung der Algorithmen, Pflege der IT
Sektor	Finanzwirtschaft
Gründung/Projektbeginn	Beginn 2022
Aktueller Entwicklungsstand	Variiert je nach Anwendungsfall, technische Infrastruktur ist entwickelt, Geschäftsmodell ist noch in Bearbeitung
Rechtsform	Hessische Ministerium für Wirtschaft, Energie, Verkehr und Wohnen hat Trägerschaft des Datentreuhänders übernommen

3.3.1 Hintergrund

Das Projekt EuroDaT geht auf die Initiative des Landes Hessens zurück, die schon seit 2018 am Aufbau eines Finanzdaten-Clusters gemeinsam mit Akteuren aus Politik und Verwaltung, hessischen Universitäten sowie Unternehmen am Finanzplatz Frankfurt arbeitet. Die ursprüngliche Motivation war der Aufbau von Big Data im Finanzsektor, um die Entwicklung von KI-Anwendungen zu ermöglichen.

Das Projekt ist in vier Teilprojekte bzw. Anwendungsfälle unterteilt:

1. Erkennung von Finanzkriminalität (Geldwäschebekämpfung),
2. Bewertung von Finanzunternehmen nach Umwelt- und Sozialkriterien
3. Haushaltsbarometer für private Finanzdaten
4. Bereitstellung von Mikrodaten für wirtschaftswissenschaftliche Forschung

Neben den vier Anwendungsfällen existiert ein zusätzliches Teilprojekt mit einer Sonderrolle, welches die Aufgabe hat sicherzustellen, dass langfristig eine eigenständige handlungs- und funktionsfähige Organisation aus EuroDaT

entsteht. Diese soll es ermöglichen, dass in der Zukunft neben den vier Anwendungsfällen noch weitere aktuell noch unbekannte Anwendungsbereiche bearbeitet werden können.

Die Bundesregierung unterstützt EuroDaT seit 2022. Das voraussichtliche Förderende ist 2024. Das bereitstehende Budget umfasst 20 Mio. Euro. Das EuroDaT-Konsortium besteht aus privaten Unternehmen, Forschungsinstituten, Universitäten und Landesministerien.

3.3.2 Konzeption, Herausforderungen, Lösungsansätze

Der Datentreuhänder stellt in EuroDaT ausschließlich die Infrastruktur zur Verfügung und sorgt für deren Betrieb. Dazu gehört die Pflege der Software, der Bau zusätzlicher Schnittstellen, die technische Prüfung der Algorithmen sowie die Fehlerbehandlung. Da insbesondere bei sensiblen Finanzdaten die Vertrauensbildung zwischen Datengebenden und Datennutzenden zentral für die Funktionsfähigkeit ist, soll der Datentreuhänder neutral sein und kein eigenes Geschäftsinteresse haben. Daher hat das hessische Ministerium die Rolle als Hauptgesellschafter inne. Dies soll Neutralität nach außen signalisieren und damit Vertrauen vermitteln. Mittelfristig sollen zusätzliche Gesellschafter aus Forschung und der Industrie gewonnen werden. Für den nachhaltigen Betrieb des Datentreuhänders soll eine Gebühr pro Datenanfrage erhoben werden, um die Kosten und eine Marge für die Risikoabdeckung zu erwirtschaften.

Zukünftig sollen externe Service Provider die Algorithmen der beauftragten Datenauswertung inhaltlich prüfen. Da bisher keine Service Provider beteiligt sind, übernimmt dies aktuell der Datentreuhänder. Perspektivisch soll EuroDaT offen für verschiedene Service Provider sein, um so Wettbewerb zu erzeugen.

Der Startpunkt des Ablaufs einer Datenverarbeitung im EuroDaT-Ökosystem ist ein Vertragsverhältnis zwischen dem Datengebende, dem Ergebnisnutzer und eventuell einem Service Provider. Der Vertrag umfasst die Definition der Daten, der darauf anwendbaren Analysen (Filterung, Algorithmen) sowie Umfang und Zeitpunkt der Auswertungen. Die Umsetzung des Vertrags durch die Datenauswertung findet automatisiert statt. Der verschlüsselte Roh-Datensatz des Datengebendes wird in dem geschützten Bereich von EuroDaT entschlüsselt, zum Auswertungszweck gefiltert und erst danach der gekapselten Applikation des Datendienstleisters zur Verfügung gestellt. Der Ergebnisnutzer erhält zum Schluss die definierte Auswertung. Grundidee von EuroDaT ist, dass zu keinem Zeitpunkt des Zyklus jemand Zugriff auf den *gesamten* unverschlüsselten Datensatz hat. Durch die gekapselt zur Verfügung gestellten Filter- und Auswertungsalgorithmen hat auch der Datentreuhänder zu keinem Zeitpunkt direkten Zugriff auf die unverschlüsselten Daten. Die IP der Auswertungsalgorithmen bleiben ebenfalls geschützt. Technisch basiert die Infrastruktur auf GAIA-X. Ferner wir auf GAIA-X-Zertifizierungen zurückgegriffen.

Das Design von EuroDaT ist so ausgestaltet, dass Monopolisierungstendenzen verhindert werden sollen. Neben dem Wettbewerb der Serviceprovider sollen Daten diskriminierungsfrei denjenigen zur Verfügung gestellt werden, denen die Datengebende Zugriff gewähren. Weitere Einschränkungen sind nicht vorgesehen.

Als **zentrale Herausforderung** für EuroDaT hat sich die Anreizstruktur für die Datengebende herausgestellt. Dies ist beispielsweise im Anwendungsfall der Finanzkriminalität und der Datenbereitstellung durch einzelne Banken nicht abschließend geklärt, insbesondere da eine Datenbereitstellung in diesem Kontext auch mit Risiken des Datenschutzes einhergeht. Die Problematik ist bei den verschiedenen Anwendungsfällen unterschiedlich stark ausgeprägt.

Das Design von EuroDaT ist offen gegenüber verschiedenen Datenformaten. Die Vision des Projekts ist es daher, dass EuroDaT perspektivisch auch in anderen Industrien verwendet werden kann (EuroDaT, 2022).

3.4 Use Case 3: Evarest

Tabelle 3 Steckbrief: Evarest

Steckbrief	**EVAREST**
Verantwortliche Organisation(en)	DFKI (Konsortialführer), Lebensmittelunternehmen und Forschungsinstitutionen
Typ der Datenbereitsteller	Semi-automatisierte Datenerzeugung
Funktionen des Datentreuhänders	Schaffung eines lokalen und globalen Datenökosystems:

	• Entwicklung und Verwertung von Datenprodukten als Wirtschaftsgut im Ökosystem der Lebensmittelproduktion • Offene, technische Datenplattform über Unternehmensgrenzen hinweg mit ökonomischen und rechtlichen Nutzungskonzepten • Multi-sided Daten-Plattform für den abgesicherten, rechtskonformen Handel von Daten und Algorithmen • Erschließung neuer Geschäftsmodelle für Lebensmittelproduzenten • Selbstbeschreibung von Datenprodukten und Services
Sektor	Lebensmittelindustrie, Agrarindustrie
Projektbeginn/Förderlaufzeit	01.01.2019-30.04.2022
Aktueller Entwicklungsstand	Für die Skalierung des Modells fehlt es noch an großen Industriepartnern
Rechtsform	privatrechtlich

3.4.1 Hintergrund

Mit 171 Mrd. Euro Umsatz und ca. 580.000 Beschäftigten nimmt die Lebensmittelindustrie volkswirtschaftlich wie auch gesellschaftlich eine strategisch wichtige Rolle in der deutschen Wirtschaft ein. Bisher werden die produzierten Daten nicht umfassend genutzt. Das Projekt wurde vom Bundesministerium für Wirtschaft und Energie gefördert. Das Konsortium bestand aus dem Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI) als Konsortialführer, der Agrarmarkt Informations-Gesellschaft mbH (AMI), der Chocoladefabriken Lindt & Sprüngli GmbH, dem Forschungsinstitut für Rationalisierung e.V. (FIR) an der RWTH Aachen, der Software AG sowie der Universität des Saarlandes. Darüber hinaus unterstützen die BVE e.V. und das DIL e.V. das Projekt als Assoziierte.

3.4.2 Konzeption Herausforderung und Lösungsansätze

Für nicht-kritische Daten bzw. größere Datenmengen wird ein cloud-basierter Ansatz gewählt (vgl. oben unter Punkt 1.4 Modell 2), wohingegen für kritische Daten mit besonderem Schutzwert ein sog. Broker-Modell erarbeitet wurde. Der „Broker“ genannte Intermediär verfolgt dabei mehrere Funktionen. Vordergründig tritt er als Datenverarbeitungshilfe auf. Die Datennutzenden im Broker-Modell sind als reine Ergebnisnutzer einzuordnen (vgl. oben unter Punkt 1.4 Modell 3). Beim Upload der Datensätze durch den Datengebende in das System werden halbautomatisiert Datenbeschreibungen zum jeweiligen Datensatz generiert (Katalogcharakter: Metadaten; keine Samples). Je nach Verarbeitungsanfrage des Ergebnisnutzers gibt der Intermediär Vorschläge für passende Datensätze und passende Algorithmen zur Lösung der Anfrage des Ergebnisnutzers vor. Die Modellierung kann dann durch den Datengebende selbst unter Hilfe des Intermediärs oder durch den Intermediär stattfinden. Der Ergebnisnutzer hat zu keinem Zeitpunkt Zugriff auf die Rohdaten.

Der Broker überprüft die Angemessenheit des Preises für einen Datensatz / Verarbeitungsvorgang und garantiert für die Qualität der Daten zur Erreichung des geforderten Ergebnisses. Grundlage für seine Einordnung bildet ein “Broker-Framework”, d.h. ein Vertrag, in dem Datengebende und Ergebnisnutzer die Nutzungsrechte und Verarbeitungserfolge definieren.

Der Broker tritt zur Lösung des Wert-Risiko-Dilemmas jedoch nicht nur als Sicherungs- und Verarbeitungsmodul auf, sondern setzt ggf. schon auf Seite der Datennutzenden und auf Seite der Datengebende gewisse Incentives für die Datenteilung. Einerseits müssen Datengebende vom Mehrwert der Teilung überzeugt werden (finanzieller Anreiz, Service-Anreiz; Anreiz der Verbesserung des Ökosystems). Andererseits müssen auch Datennutzende über den Mehrwert geschult werden, um auch auf Nachfrageseite Aktivität zu generieren.

Für die Skalierung des Modells fehlt es noch an großen Industriepartnern.

3.5 Use Case 4: Forschungsdatenzentren

Tabelle 4 Steckbrief: Forschungsdatenzentren

Steckbrief	**Forschungsdatenzentren (FDZ) der statischen Ämter des Bundes und der Länder**
Verantwortliche Organisation(en)	statistische Ämter des Bundes und der Länder
Typ der Datenbereitsteller	Aktive Datenbereitstellung
Funktionen des Datentreuhänders	In Erfüllung gesetzlicher Verpflichtungen stellen die statistischen Ämter des Bundes und der Länder über die Forschungsdatenzentren eine technische, organisatorische und rechtliche Infrastruktur bereit, um den Zugang der Wissenschaft zu den Mikrodaten der amtlichen Statistik durch die Einrichtung unterschiedlicher Datennutzungswege zu erleichtern (siehe oben bei Punkt 1.4 Modell 2 und 3).
Sektor	Forschung
Projektbeginn	FDZ des Bundes seit 2001, FDZ der Länder seit 2002
Aktueller Entwicklungsstand	Technologische, organisatorische und rechtliche Infrastruktur ist voll entwickelt und in Betrieb
Rechtsform	öffentlich-rechtlich

3.5.1 Hintergrund

Im Jahr 1999 wurde im Auftrag des BMBF ein Gutachten der „Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik“ (KVI) erstellt. Dessen zentrale Empfehlung war die Einrichtung von Forschungsdatenzentren bei den öffentlichen Datenproduzenten, um den Zugang der Wissenschaft zu den Mikrodaten der amtlichen Statistik durch die Einrichtung unterschiedlicher Datennutzungswege zu erleichtern. Auf dieser Grundlage folgte eine Förderung durch das BMBF zum Aufbau zunächst des FDZ des Bundes im Jahr 2001 und darauf aufbauend im Jahr 2002 der FDZ der Länder. Die FDZ werden durch den Wissenschaftlichen Beraterkreis (WBK) unterstützt. Die Organisation bzw. Koordination mit den jeweils anderen Forschungsdatenzentren erfolgt im Rat für Sozial- und Wirtschaftsdaten (RatSWD).

3.5.2 Konzeption Herausforderung und Lösungsansätze

Die FDZ stellen den Zugang zu den Daten über abgestufte Zugangswege bereit. Die Daten stehen dabei in jeder Ausgestaltung des Use Cases nur dem begrenzten Kreis der nach dem BStatG Anspruchsberechtigten zur Verfügung (vgl. besonderes Privileg für Wissenschaft nach § 16 Abs. 6 BStatG – vgl. oben unter Punkt 1.4 Modell 2 in seiner Stufe 1). Unter Umständen steht ein Zugriff auf die Public Use Files einem erweiterten Kreis an Anspruchsberechtigten zur Verfügung (vgl. Oben unter Punkt 1.4 Modell 1).

In Bezug auf die Sicherungsinstrumente verfolgen die FDZ einen Dreiklang der folgenden Elemente: Anonymisierung, technisch-organisatorische Zugangsbeschränkungen und rechtliche Verpflichtungen. Die FDZ unterscheiden dabei zwischen Onsite-Nutzung (Kontrollierte Datenfernverarbeitung und Gastarbeitsplätze vor Ort) und Offsite-Nutzung (Scientific Use Files und Public Use Files).

Bei der Offsite-Nutzung unterscheidet das FDZ zwischen Public Use Files (faktisch anonymisierte Datensätze, höchste Form der Informationseinbuße; vgl. Oben unter Punkt 1.4 Modell 2 - Stufe 1) und Scientific Use Files (faktisch anonymisierte Datensätze, geringeres Maß an Informationseinbuße; vgl. oben unter Punkt 1.4 Modell 2 - Stufe 2). Die Ergebnisse der Datenverarbeitungen bei den jeweiligen Nutzern werden nicht überprüft, vielmehr verschiebt sich die Sicherung auf das vorgelagerte Element der Anonymisierung. Zusätzlich hierzu greifen rechtliche Schutzmechanismen, indem Nutzende Nutzungsvereinbarungen mit weniger starken (PUF) oder stärkeren Verpflichtungen (SUF) zur Geheimhaltung, Löschung und zeitlichen Begrenzung der Nutzung abschließen müssen.

Die Onsite-Nutzung eröffnet dem Datennutzenden dabei die höchste Datenqualität und Zugriff auf formal anonymisierte Mikrodaten. Nutzende müssen vor Ort die Datenverarbeitung vor- nehmen (ggf. neuerdings auch durch einen gesicherten Remote-Zugang). Die Verarbeitung wird durch Logfiles nachvollzogen und

stichprobenartig geprüft. Das Ergebnis der Verarbeitung wird durch Angestellte des FDZ nach menschlichem Ermessen auf die Gefahr von Rückschlüssen (insb. Reidentifizierung) durch Zusammenschau anderer Verarbeitungsvorgänge und Veröffentlichungen geprüft (vgl. oben unter Punkt 1.4 Modell 2 - Stufe 3). Die Überprüfung ist sehr aufwändig; bisher wurden keine automatisiert.

3.6 Use Case 5: i4Trust

Tabelle 5 Steckbrief: i4Trust

Steckbrief	**i4Trust**
Verantwortliche Organisation(en)	Die i4Trust-Initiative ist ein von der Europäischen Union (EU) finanziertes Projekt zur Entwicklung einer vertrauenswürdigen und sicheren Rahmenstruktur für den Datenaustausch zwischen verschiedenen Organisationen in unterschiedlichen Branchen. I4Trust besteht aus zwei Hauptorganisationen, FIWARE Foundation und iShare Foundation.
Typ der Datenbereitsteller	Zentrale und dezentrale Datenbereitstellung
Funktionen des Datentreuhänders	Bereitstellung der IT-Infrastruktur, Verwaltung der Identitäten von Teilnehmenden, Unterstützung zur Bereitstellung der standardisierten Schnittstelle,
Sektor	Hauptsächlich in Logistik, Energie, Landwirtschaft, und Smart-City
Gründung/Projektbeginn	2020
Aktueller Entwicklungsstand	Die grundliegende Rahmenstruktur sowie die technischen Bausteine sind verfügbar. Die erste Versuchsphase mit 13 Teilprojekten hat begonnen, in denen die Funktionalität der Rahmenstruktur in verschiedenen Szenarien getestet wird.
Rechtsform	Foundation

3.6.1 Hintergrund

Die i4Trust ist gefördert im Rahmen des Forschungs- und Innovationsprojektes, Horizon 2020, der Europäischen Union. Das Ziel ist eine vertrauenswürdige und sichere Rahmenstruktur für den Datenaustausch zwischen verschiedenen Organisationen in unterschiedlichen Sektoren zu entwickeln. FIWARE Foundation und iShare Foundation sind die Kernmitglieder von i4Trust. Dabei liefert FIWARE die wichtigsten technischen Bausteine für i4Trust, z.B. die standardisierte API (Application Programming interface) und Datenmodelle. Und iShare hingegen stellt die Lösungen auf der organisatorischen und strukturellen Ebene bereit.

Zwei Hauptgründe haben zur Auswahl von i4Trust als einer der Use Cases geführt: :i4Trust bietet eine Vielzahl von Open-Source-Tools und -Ressourcen, um Organisationen bei der Implementierung des Frameworks zu unterstützen. Dazu gehören Bausteine zur Implementierung verschiedener Funktionen, eine Reihe von Leitlinien und bewährten Verfahren sowie ein Anwendungsfall für MVPs. Die meisten der bereitgestellten Technologiemodule sind in einem neuen Sandbox-Stil verpackt und können durch die einfache Installation der entsprechenden Orchestrierungssoftware problemlos eingesetzt werden. Dies ermöglicht es uns, das i4Trust-Framework auf sehr intuitive Weise zu analysieren und validieren. i4Trust hat eine Reihe von Erfolgsgeschichten im Jahr 2022 veröffentlicht mit einer Umsetzung in13 konkreten Projekten. Derzeit ist nur eine kurze

Fallbeschreibung für jeden Fall verfügbar. Ein ausführlicherer Bericht wird jedoch in Kürze veröffentlicht werden. Diese praktischen Beispiele helfen uns, das i4Trust-Framework aus der Perspektive verschiedener Bereiche zu analysieren.

Ergänzend zu einem Interview haben wir i4Trust auf Basis verfügbarer Informationen und Unterlagen, wie z.B. auf GitHub[4], analysiert.

3.6.2 Konzeption, Herausforderungen, Lösungsansätze

Die i4Trust Rahmenstruktur wird von vielen technischen Bausteinen unterstützt. Je nach ihrer Funktionalität können die Bausteine in drei Kategorien unterteilt werden:

1. Interoperabilität von Daten
2. Datensouveränität und -vertrauen
3. Datenwertschöpfung

Die wichtigsten Bausteine zur Gewährleistung der Dateninteroperabilität sind zwei technische Standards:

NGSI-LD ist ein API-Standard für das Management von Kontextinformationen. i4Trust bietet mehrere Open-Source-Kontextbroker, die auf diesem Standard basieren: Orion-LD, Scorpio und Stellio. Die Verwendung dieser Kontextbroker ist nicht obligatorisch. Wichtig ist es nur, dass die Kommunikation innerhalb des Data Spaces mit dem NGSI-LD-Standard erfolgt.

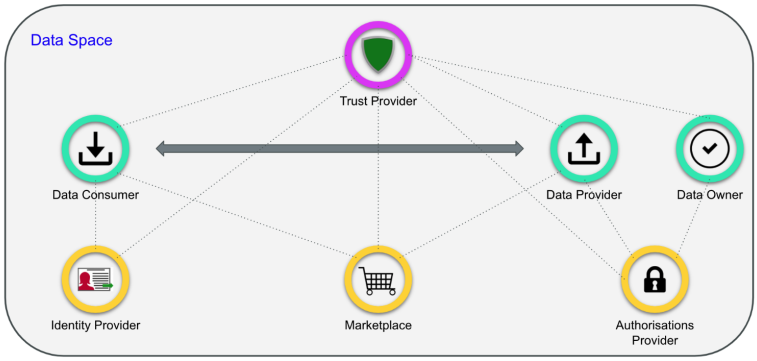
Um unterschiedliche Interpretation der gleichen Informationen zu vermeiden, wird Smart Data Models von i4Trust eingeführt. Dies ist die Sammlung vieler Open-Source Datenmodelle, die gängige Anwendungsszenarien im Bereich des IoT abdecken. Für die nicht abgedeckten Anwendungen wird ein detaillierter Prozess zur Erstellung und Erweiterung des Datenmodells zur Verfügung gestellt.

Um Datensouveränität und -vertrauen zu gewährleisten, stützt sich der i4Trust auf zwei technische Bausteine, die Identity Management Komponente (IM) und die Access Control Komponente (AC). Die Funktion des IMs besteht darin, ein umfassendes und sicheres Identitätsmanagementsystem bereitzustellen, das Benutzenden je nach ihrer Identität unterschiedliche Zugriffsrechte gewähren kann. Die Zugriffskontrolle beschränkt den Zugang zu verschiedenen Anwendungen, Diensten und Ressourcen auf Grundlage der von IM bereitgestellten Informationen über die Benutzerrechte.

Zu guter Letzt unterstützt i4Trust die Schaffung eines *Marktplatzes*. Der i4Trust Marketplace basiert auf dem FIWARE Business API Ecosystem (BAE), das die Veröffentlichung von Daten oder Diensten ermöglicht und die Erstellung und Überwachung von Prozessen im Zusammenhang mit intelligenten Verträgen (die die Rechte und Pflichten für die Nutzung von Daten und Diensten klar beschreiben) unterstützt. Mit Hilfe des Marktplatzes werden Funktionen für die gemeinsame Nutzung von Daten, die Verwaltung des Datenzugriffs und die Verrechnung von Leistungen in den Datenraum integriert, was die Koordinierung zwischen den verschiedenen Beteiligten erheblich vereinfacht.

[4] https://github.com/i4Trust

Abbildung 9 Architektur der i4Trust-Rahmenstruktur



Quelle: i4Trust (2022).

Abbildung 9 stellt die Architektur der i4Trust-Rahmenstruktur dar. Davon sind drei Komponenten die Kernbausteine:

1. Standardisierte Datenaustausch-API: garantiert die Interoperabilität des Datenaustauschs
2. Die AC: sichert den Zugang der Daten
3. Die IDM: garantiert die Vertrauenswürdigkeit des Datenaustauschs

Mit diesen Komponenten kann ein MVP des Data Spaces aufgebaut werden. Natürlich gibt es noch weitere Bausteine, die den Datenaustausch optimieren können, wie z.B. ein Marktplatz für Geschäftsvorgänge oder Distributed Ledger Technology (DLT), um den Datenaustausch vertrauenswürdiger zu machen. Allerdings fehlt i4Trust dafür entweder eine detaillierte Beschreibung für die Funktionalität und Verwendung von diesen Bausteinen oder sie werden nicht in realen Anwendungsfällen eingesetzt. Darüber hinaus hat i4Trust auch keine Lösung für den Schutz von sensiblen Daten gegeben. Dafür konnten wir keine technischen Bausteine bzw. Einleitung für die Anonymisierung oder Pseudonymisierung der Daten finden. An dieser Stelle können wir folgende Urteile über i4Trust fällen. Die Stärke von i4Trust ist, dass es standardisierte technische Open-Source Bausteine anbietet. Mit Hilfe dieser Bausteine lässt sich schnell eine funktionsfähige Umgebung zum Datenaustausch einrichten. Gleichzeitig bietet i4Trust auch praktische Anleitungen zur Schaffung von Vertrauen. Es wird eine dritte Partei als Vertrauensperson eingeführt, die die Verwaltung des Datenaustauschs koordinieren soll. Allerdings weist i4Trust in zwei Aspekten noch Lücken auf. Erstens legt i4Trust nicht fest, wie die Interessen der verschiedenen Beteiligten ausgeglichen werden sollen, z. B. wie die dritte Partei, die als Vertrauensanbieter fungiert, Gewinne erzielen soll. Zweitens ist i4Trust aufgrund des fehlenden Schutzes sensibler Daten für den Einsatz in bestimmten Bereichen wie z.B. dem Austausch von Gesundheitsdaten nicht geeignet. Die technischen Bausteine zur Unterstützung dieser beiden Anforderungen wären daher sehr wertvoll für die weitere Forschung.

3.7 Use Case 6: UK Biobank

Tabelle 6 Steckbrief: UK Biobank

Steckbrief	
Verantwortliche Organisation(en)	UK Biobank Limited
Typ der Datenbereitsteller	Privatpersonen, aktiv
Funktionen des Datentreuhänders	Erhebung, zentrale Speicherung, Aufbereitung und Bereitstellung der Daten; Zugangsmanagement; Betrieb, Pflege, Erweiterung der Infrastruktur

Sektor	Medizin
Gründung/Projektbeginn	Aufbau 2006-12; Daten werden seit 2012 genutzt
Aktueller Entwicklungsstand	In Betrieb. Daten und Infrastruktur werden aktiv genutzt. Etwa 35.000 Forscher*innen weltweit sind als Nutzer registriert; ~3100 Forschungsprojekte wurden seit 2012 mit UK Biobank Daten durchgeführt, >3200 Veröffentlichungen. Datensätze werden regelmäßig erweitert.
Rechtsform	Eingetragene Stiftung und Kapitalgesellschaft mit begrenzter Haftung

3.7.1 Hintergrund

UK Biobank wurde 2006 als gemeinsame non-profit Initiative von Wellcome Trust, UK Medical Research Council, Department of Health, der Schottischen Regierung und der Northwest Regional Development Agency gegründet. Ziel war es, mit UK Biobank einen der weltweit größten und qualitativ hochwertigsten medizinischen Forschungsdatensätze zu schaffen und diesen über regelmäßige Neuerhebungen mit den gleichen Datenspendern laufend zu erweitern, um so prospektive Studien und ein langfristiges Tracking über 40-50 Jahre zu ermöglichen.

UK Biobank erhält eine Grundfinanzierung von GBP 10 Millionen im Jahr (bislang insges. GBP 180M) von den genannten Gründern und weiteren staatliche Stellen und Stiftungen. Weitere GBP 5-6M werden durch Zugangsgebühren generiert; sie werden primär für das laufende Zugangsmanagement aufgewendet. Schließlich bekommt UK Biobank Industriespenden von durchschnittlich 90 Mio. Pfund im Jahr, um spezifische Erweiterungen des Datensets vorzunehmen (z.B. die Genomsequenzierung [Kosten: GBP 200M]).

3.7.2 Konzeption, Herausforderungen, Lösungsansätze

Daten und Akzeptanz seitens der Datengebenden. UK Biobank enthält die vollständigen Gesundheitsdaten von 500.000 in Großbritannien lebenden Datengebende (Privatpersonen). Bei allen wurden ihren elektronischen Gesundheitsakte, Blut- und Stuhlproben, Genotyp einschließl. Exom- und Genomsequenzierung, Knochendichte, Lungenfunktion, >30 Biomarker, Angaben zu Ernährung, Arbeitsgeschichte, psychischer und kognitiver Verfassung u.a.m., erfasst. Bei jeweils 100.000 wurden MRI, DEXA und Ultrasound Scans gemacht, Speichelproben entnommen oder körperliche Bewegung über 7 Tage gemonitort. Die Daten werden regelmäßig neu erhoben.

Für die Erhebung wurde eine repräsentative Stichprobe von 5 Millionen Personen aus dem Gesamtpatientenregister des NHS gezogen und angeschrieben. Die Rücklaufquote war ~10%, mit ausgewogener regionaler und ethnischer Verteilung aber einem Bias zu Bessergebildeten. Die Daten sind somit nicht vollkommen repräsentativ. Aufgrund der großen Zahl der Datengebenden erlauben sie dennoch *belastbar generalisierbare Rückschlüsse* auf alle wesentlichen Bevölkerungsgruppen in Großbritannien. Laut dem UK Biobank Interviewpartner ist das wichtiger als Repräsentativität, da diese aufgrund demographischen Wandels ohnehin nicht konstant ist.

Akzeptanz seitens der Datengebenden scheint hoch. Die Teilnahme ist freiwillig und pro-bono (Datengebende bekommen keine Belohnung oder Gegenleistung); die Zahl der „Aussteiger*innen“ sehr niedrig trotz des relativ zeitaufwendigen Teilnahmeprozesses.
Zugang und Akzeptanz seitens der Datennutzenden. UK Biobank Daten stehen allen Forschenden der Welt zur Verfügung, mit gleichen Zugangsbedingungen. Es wird nicht zwischen Industrie- und universitärer Forschung unterschieden; Firmen, die die Erstellung eines neuen Datensatzes finanziert haben (z.B. Genomsequenzierung), bekommen zu diesem jedoch 9 Monate Exklusivzugang.

Um Zugang zu bekommen, müssen Forschende nachweisen, (i) dass sie bei einem Institut oder einer Firma angestellt sind, welche (ii) in der Forschung tätig ist und (iii) sie selbst aktiv als Wissenschaftler*innen arbeiten. Falls in den letzten 3 Jahren Beschwerden (*complaint*) gegen den oder die Forschende erhoben wurden, müssen diese detailliert werden. Schließlich muss die geplante Forschung, benötigte Daten und etwaige durch die Arbeit

neugenerierte Daten/Variablen beschrieben werden. Die Bearbeitungszeit bis zur Datenfreigabe liegt bei ~17 Wochen (Stand 2022). Die Zugangsgebühr (für drei Jahre) liegt bei GBP 3000, 6000 oder 9000 je nach Datensatz, ermäßigt auf GBP 500 für Studierende und Forschende aus Entwicklungsländern.

Alle abgeleiteten Variablen und verwendeten Algorithmen müssen UK Biobank nach Abschluss der Forschung zur Verfügung gestellt werden und werden umgehend veröffentlicht.
UK Biobank wurde in den ersten Jahren über Fachveranstaltungen beworben. Die Zahl der Datennutzenden wuchs sehr rasch, nachdem die *American Society of Human Genetics* prominent über UK Biobank berichtete. Heute sind ~90% der Nutzenden universitäre Forschende, 10% Industrie. Alle der 35 weltweit wichtigsten Pharmafirmen nutzen UK Biobank; der Interviewpartner berichtete aber, dass Firmen manchmal vor der Nutzung aus (unbegründeter) Sorge, sie müssten IP o.ä. abtreten, oder es doch „einen Haken geben müsse", zurückschrecken.

Governance. Der Vorstand, Aufsichtsgremien/Beiräte (*Scientific Advisory, Strategy, Ethics, Access*) und *Expert Working Groups* bestehen v.a. aus renommierten Wissenschaftler*innen sowie einer kleinen Zahl von Repräsentant*innen der *Core Funder* und der Zivilgesellschaft.

Zugangsbedingungen werden hauptsächlich vertraglich geregelt. Die Durchsetzung findet primär rechtlich statt; weniger durch Technikgestaltung. UK Biobank übernimmt keine Verantwortung für den Schutz von IP oder Geschäftsgeheimnissen der Nutzenden und ihrer Institutionen.

Technische Infrastruktur. Die Daten sind zentral abgespeichert (Amazon Cloud). Nutzende greifen auf sie über eine Plattform der Firma DataNexus zu und können sie *in situ* über die Plattform auswerten, mittels bereitgestellter Analysetools oder indem sie eigene Scripts hochladen. Viele Daten können auch heruntergeladen werden (Ausnahme: genetische Daten). In den nächsten Jahren soll die Download-Option aber beendet werden.

Standards und Zertifizierung. UK Biobank ist nach ISO 9001:2015 (Qualitätsmanagement) und ISO 27001:2013 (Informationssicherheit) zertifiziert und gem. UK Human Tissue Act 2004 lizensiert. Daten werden nach ICD 10 (WHO) codiert. Dieser Datenstandard wird, so der Interviewpartner, auch von anderen führenden Biobanken verwendet. Die ISO Biobank Standards 20387:2018 und 21899:2020 sind bekannt; eine entsprechende Zertifizierung wird aber *nicht* angestrebt.
UK Biobank hat die eigenen Prozesse, v.a. der Datenerhebung, genau standardisiert. Diese Vorgaben (*Protocols*) sind veröffentlicht; andere Biobanken verwenden sie. Laut Interview-partner findet unter führenden Biobanken ein begrenztes Maß informeller, Bottom-up-getriebener de facto Standardisierung statt. UK Biobank hat dabei eine gewisse Beispielfunktion inne. Z.B. wird ICD 10 allgemein als Datenstandard verwendet. Es sei wünschenswert, die Entwicklung gemeinsamer Prinzipien stärker zu forcieren; noch gäbe es hier aber wenig Bewegung.

Dennoch seien Abweichungen unvermeidlich, etwa aufgrund unterschiedlicher Rechtslagen. Während die *interne* Standardisierung der Prozesse innerhalb der jeweiligen Biobank äußerst wichtig sei (z.B. bei der Datenerhebung), sei eine *völlige* Standardisierung über alle Biobanken hinweg jedoch unnötig. Abweichungen z.B. bei Datenstandards seien tolerabel, solange die Daten/Einheiten unterschiedlicher Biobanken leicht aufeinander gemappt werden könnten.

Ökosystem und Gaia-X. Zu diesem Zeitpunkt ist UK Biobank nicht Teil eines größeren Datenökosystems. Perspektivisch könnte ein breiteres Ökosystem bestehend aus föderierten Bio-Datenbank entstehen. Der Interviewpartner hielt dies für zumindest mittelfristig wahrscheinlich, sowie sich die nötigen Technologien entwickeln, um Daten *remote* auszuwerten und über die Grenzen unterschiedlicher Datenbanken hinweg zu kombinieren, ohne die Daten allerdings direkt zwischen einzeln Biobanken hin- und herzuschieben (d.h. ohne sie aus der Kontrolle der einzelnen Biobank zu geben). Die Gaia-X Initiative sei UK Biobank bekannt, man beteilige sich aber nicht direkt.

Erfolgsfaktoren. 5 Erfolgsfaktoren lassen sich bei UK Biobank identifizieren: (1) Pilotierung und strenge Standardisierung der eigenen Prozesse, insbesondere der Datenerhebung. (2) Leistungsfähige IT-Systeme, die auf die Verarbeitung von medizinischen Daten und klinischen Studien zugeschnitten sind. (3) Einfache und für alle gleiche Zugangsregeln. (4) Zentralisiert verfügbare Gesundheitsdaten (elektronische Patientenakten). (5) Berichte/Bewerbung durch reichweitenstarke Akteure in den Fachcommunities (z.B. American Society for Human Genomics).

Anhang A: Literaturverzeichnis

ADRF Network Working Group Participants (2018). Data Sharing Governance and Management. ADRF Network Working Group Reports. 2. https://repository.upenn.edu/admindata_reports/2.

Acharya, S., Mekker, M. (2022). Measuring Data Sharing intention and its association with the acceptance of connected vehicles. Transportation research part F: traffic psychology and behaviour. 89: 423-436.

Arlinghaus, T., Kus, K., Kajüter, P., & Teuteberg, F. (2021). Datentreuhandstellen gestalten: Status quo und Perspektiven für Geschäftsmodelle. HMD Praxis der Wirtschaftsinformatik, 58(3), 565-579.

AZT Automotive (2019). Investigating Accidents Involving Highly Automated Vehicles: Concept of a Data Trustee and Data Model for Future Homologation. https://trid.trb.org/view/1755981.

Bitkom (2021). Catena-X: Eine Dateninfrastruktur für die automobile Wertschöpfungskette. Vortrag im Rahmen der Digital Mobility Conference am 24. November 2021, mit Oliver Ganser, Programmleitung Catena-X, BMW. URL: https://www.youtube.com/watch?v=Df4-qpLfw8E (letzter Abruf am 20.3.23).

Blankertz, A. (2020). Designing Data Trusts Why We Need to Test Consumer Data Trusts Now. Stiftung Neue Verantwortung.

Blankertz, A., Braunmühl, P. V., Kuzev, P., Richter, F., Richter, H., & Schallbruch, M. (2020). Datentreuhandmodelle-Themenpapier.

Blankertz, A., & Specht-Riemenschneider, D.L. (2021). Wie eine Regulierung für Datentreuhänder aussehen sollte.

Brown, C., Regan, A., van der Burg, S. (2022). Farming futures: Perspectives of Irish agricultural stakeholders on Data Sharing and data governance. Agriculture and Human Values. https://doi.org/10.1007/s10460-022-10357-8are.

BMBF (2021). Richtlinie zur Förderung von Projekten zur Entwicklung und praktischen Erprobing von Datentreuhandmodellen in den Bereichen Forschung und Wirtschaft, Bundesanzeiger vom 08.01.2021 https://www.bmbf.de/bmbf/shareddocs/bekanntmachungen/de/2021/01/3292_bekanntmachung.

Bundesministerium für Wirtschaft und Klimaschutz (BMWK) (o.J.). Projektsteckbriefe. URL: https://www.bmwk.de/Redaktion/DE/Downloads/P-R/projektsteckbriefe.pdf?__blob=publicationFile&v=6 (letzter Abruf am 20.3.23).

Bundesverband der Deutschen Industrie BDI (2021). Datenwirtschaft in Deutschland - Wo stehen die Unternehmen in der Datennutzung und was sind ihre größten Hemmnisse? https://www.iwkoeln.de/studien/klaus-heiner-roehl-lennart-bolwin-wo-stehen-die-unternehmen-in-der-datennutzung-und-was-sind-ihre-groessten-hemmnisse.html

Catena-X (2022a). Catena-X Operating Model Whitepaper. Release V2 - 21.11.2022. URL: https://catena-x.net/fileadmin/user_upload/Publikationen_und_WhitePaper_des_Vereins/CX_Operating_Model_Whitepaper_02_12_22.pdf (letzter Abruf am 20.3.23).

Catena-X (2022b). The first open and collaborative data ecosystem. URL: https://catena-x.net/fileadmin/user_upload/Vereinsdokumente/Catena-X_UEbersicht.pdf (letzter Abruf am 05.12.2022).

Catena-X (2023a). Catena-X: Das erste offene und kollaborativ angelegte Datenökosystem. URL: https://catena-x.net/de/ueber-uns (letzter Abruf am 20.3.23).

Catena-X (2023b). Mehrwerte. URL: https://catena-x.net/de/mehrwerte (letzter Abruf am 20.3.23).

Catena-X (2023c). Catena-X einführen & umsetzen. Catena-X betreiben. URL: https://catena-x.net/de/tractus-x/catena-x-einfuehren-umsetzen/sie-wollen-zertifizierter-betreiber-werden (letzter Abruf am 20.3.23).

Catena-X (2023d). Vision & Ziele. Die Version von Catena-X. URL: https://catena-x.net/de/tractus-x/vision-ziele (letzter Abruf am 20.3.23).

Centre for Data Ethics and Innovation CfDE (2021). Unlocking the value of data: Exploring the role of data intermediaries. https://www.gov.uk/government/publications/unlocking-the-value-of-data-exploring-the-role-of-data-intermediaries/unlocking-the-value-of-data-exploring-the-role-of-data-intermediaries.

Choi, W., Chang, S.-H., Yang, Y-.S., Jung, S., Lee, S.-J., Chun, J.-W., Kim, D.-J., Lee, W., Choi, Y.I. (2022). Study of the factors influencing the use of MyData platform based on personal health record Data Sharing system. BMC Medical Informatics and Decision Making 22(182).

Databroker (besucht am 03.2023). Documentation – Data Exchange Controller (DXC). https://www.databroker.global/documentation/DataExchangeController.

Data Spaces Support Centre (DSSC) (2022). Starter Kit for Data Space Designers. Interim Version, December 2022.

Element AI, Nesta (2019). Data Trusts. A new tool for data governance. URL: https://hello.elementai.com/rs/024-OAQ-547/images/Data_Trusts_EN_201914.pdf (zuletzt abgerufen am 14.03.23).

European Union Agency for Cybersecurity ENISA (2016). Readiness Analysis for the Adoption and Evolution of Privacy Enhancing Technologies. https://doi.org/10.2824/614444.

EuroDaT (2022). Datensouveränität in vernetzten Ökosystemen. EuroDaT im Überblick. URL: https://www.eurodat.org/fileadmin/user_upload/Broschuere_Datensouveraenitaet_in_vernetzten_OEkosystemen_06-2022.pdf (zuletzt abgerufen am 21.03.2023).

Fraunhofer (2022). Anreizsysteme und Ökonomie des Data Sharings: Handlungsfelder des unternehmensübergreifenden Datenaustausches und Status quo der deutschen Wirtschaft. https://ieds-projekt.de/wp-content/uploads/2022/03/IEDS-Whitepaper-1.pdf.

Future of Privacy Forum (2017). Understanding Corporate Data Sharing Decisions: Practices, Challenges, and Opportunities for Sharing Corporate Data with Researchers. Washington DC.

Gal, M.S., Rubinfeld, D.L. (2019). Data Standardization. New York University Law Review 94(737), 737-770.

Grafenstein, M. v. (2021). Kurzpapier: Framework zur Erfassung „erfolgreicher“ Data Governance-Modelle. HIIG Discussion Paper Series 2021-11. 7 pages. https://doi.org/10.5281/6327345.

Grossman et al. (2016). A Case for Data Commons: Toward Data Science as a Service. https://doi.org/10.1109/MCSE.2016.92.

Grossman (2018). A Proposed End-To-End Principle for Data Commons. https://medium.com/@rgrossman1/a-proposed-end-to-end-principle-for-data-commons-5872f2fa8a47.

I4Trust (2022). i4Trust Building Blocks (Version 3.0). https://i4trust.org/wp-content/uploads/CollMi-i4Trust-Impact-Story.pdf.

International Data Spaces Association (IDS) (2021). Design Principles for Data Spaces. https://doi.org/10.5281/zenodo.5244997.

Isreal Public Policy Institute (IPPI) (2022): What is a data stewardship, and how could it address questions of power imbalance in the data economy? URL: https://www.ippi.org.il/what-is-data-stewardship-and-how-could-it-address-questions-of-power-imbalance-in-the-data-economy/ (zuletzt abgerufen am 14.03.23).

Köngeter, A., Schickhardt, C., Jungkunz, M., Bergbold, S., Mehlis, K., Winkler, E. (2022). Patients' Willingness to Provide Their Clinical Data for Research Purposes and Acceptance of Different Consent Models: Findings From a Representative Survey of Patients With Cancer. Journal of Medical Internet Research, 24(8).

Kühling, J., Sackmann, F., & Schneider, H. (2020). Datenschutzrechtliche Dimensionen Datentreuhänder, Kurzexpertise im Auftrag des Bundesministeriums für Arbeit und Soziales. IZA Research Report No. 104, 2020, abrufbar unter https://ftp. iza. org/report_pdfs/iza_report_104. pdf (zit.: Kühling/Sackmann/Schneider).

Kraemer, P., Niebel, C., Reiberg, A. (2023). Gaia-X und Geschäftsmodelle: Typen und Beispiele. White Paper 1/2023, Februar 2023.

Langford, J., Poikola, A., Janssen, W., Lähteenoja, V., & Rikken, M. (2020). Understanding MyData Operators. MyData Global.

Lind, Hans-Günter, Suckfüll, Hans (2013). Die Initiative zu einer Deutschen Daten Treuhand (Dedate) als Ultima Ratio der persönlichen Digitalen Datenwirtschaft (PDD). Ansätze und Strukturen für eine gezielte Verwertung persönlicher digitaler Daten unter Berücksichtigung aller Interessensgruppen – Dateneigentümer, Wirtschaft und Staat.

Lindner, M.; Straub, S. (2023). Datentreuhänderschaft. Status Quo und Entwicklungsperspektiven. Kurzstudie im Auftrag des Bundesministeriums für Wirtschaft und Klimaschutz von der Begleitforschung zum Technologieprogramm - „Smarte Datenwirtschaft".

Lomotey et al. (2022). Data Trusts as a Service: Providing a platform for multi-party Data Sharing. https://doi.org/10.1016/j.jjimei.2022.100075.

Martin, S., Pasquale, W. (2019). Exploring Data Trust Certifications. Oxford Insights. https://www.theodi.org/wp-content/uploads/2019/04/Report_-Exploring-Data-Trust-Certification.pdf .

Meijer, A., Potjer, S., (2018). Citizen-generated open data: An explorative analysis of 25 cases. Government Information Quarterly 35, 613-621

Mills et al. (2019). Who Owns the Future? Data Trusts, Data Commons, and the Future of Data Ownership. https://doi.org/10.2139/ssrn.3437936

Mohr, S., Cloos, J. (2022). Acceptance of Data Sharing in Smartphone Apps from Key Industries of the Digital Transformation: A Representative Population Survey for Germany. Technology Forecasting and Social Change 176

Open Data Institute (2019). Data trusts: lessons from three pilots. URL: https://www.theodi.org/article/odi-data-trusts-report/.

ODI Wildlife (2019). ODI Report: Exploring the potential for data trusts to help tackle the illegal wildlife trade. https://docs.google.com/document/d/1HIlG6oVq71tk2Gy2mpI1-IYNbBKboLHIg9goplUCzI8/edit?usp=sharing&usp=embed_facebook

ODI Foodwaste (2019). ODI report: Exploring the potential of data trusts in reducing food waste. https://docs.google.com/document/d/1v9O3exRdZFu6h-xqo11Ej3Ul4cyePfRCYYBDrpUoNBk/edit?usp=sharing&usp=embed_facebook

Otto, B., ten Hompel, M., Wrobel, S. (2022). Designing Data Spaces. The Ecosystem Approach to Competitive Advantage. URL: https://link.springer.com/content/pdf/10.1007/978-3-030-93975-5.pdf (zuletzt abgerufen am 14.03.23).

Paprica, P.A., Sutherland, E., Smith, A., Brudno, M., Cartagena, R.G., Crichlow, M., Courtney, B.K., Loken, C., McGrail, K.M., Ryan, A., Schull, M.J., Thorogood, A., Virtanen, C., Yang, K., (2020). Essential requirements for establishing and operating data trusts: practical guidance co-developed by representatives from fifteen Canadian organizations and initiatives. International Journal of Population Data Science. 5(1)(31)

Rat für Informationsinfrastrukturen (RfII) (2020) Stellungnahme des Rates für Informationsinfrastrukturen (RfII) Datentreuhandstellen gestalten: Zu Erfahrungen der Wissenschaft. https://d-nb.info/1209282283/34. Zugegriffen: 4. Nov. 2020.

Rat für Informationsinfrastrukturen (RfII) (2021). Datentreuhänder: Potenziale, Erwartungen, Umsetzung Workshop der AG Datentreuhänderschaft des RfII am 25. September 2020 Zusammenfassender Workshop-Bericht.

Royal Society (2019). Protecting privacy in practice The current use, development and limits of Privacy Enhancing Technologies in data analysis. https://royalsociety.org/-/media/policy/projects/privacy-enhancing-technologies/Protecting-privacy-in-practice.pdf

SAP (2022). CATENA-X: Industry Network & Technology for a Data Driven Value Chain | SAP Community Call. URL: https://www.youtube.com/watch?v=FlvTmWZrix4, Präsentationsfolien: https://dam.sap.com/mac/app/p/pdf/asset/preview/jxDPqRH?ltr=a&rc=10 (letzter Abruf am 20.3.23).

Scassa, T. (2020). Designing Data Governance for Data Sharing. Lessons from Sidewalk Toronto. Technology and Regulation 2020, 44-56

Schneider, I. (2022). Datentreuhandschaft durch Intermediäre. Chancen, Herausforderungen und Implikationen. Vortrag 2 der Reihe „Zu treuen Händen“ | Januar 2022.

Specht Riemenschneider, L.; Kerber, W. (2022). Datentreuhänder – Ein problemlösungsorientierter Ansatz. URL: https://www.kas.de/de/einzeltitel/-/content/datentreuhaender-ein-problemloesungsorientierter-ansatz (zuletzt abgerufen am 14.03.23).

Specht-Riemenschneider, L., & Kerber, W. (2022). Designing Data Trustees–A Purpose-Based Approach.

Stalla-Bourdillon, S. (2021). A Maturity Spectrum for Data Institutions. IEEE Security & Privacy 19(5) https://doi.org/10.1109/MSEC.2021.3094985.

Stevens, G.; Boden, A. (2022). Warum wir einen parteiische Datentreuhänder brauchen. Zum Modell der Datentreuhänderschaft als stellvertretende Deutung der Interessen individueller und kollektiver Identitäten. Vortrag 6 der Reihe „Zu treuen Händen“ | Februar 2022.

Yoon, A., Lee, Y.Y. (2019). Factors of trust in data reuse. Online Information Review, 43(7), 1245–1262. https://doi.org/10.1108/OIR-01-2019-0014.

Young et al. (2019). Beyond Open vs. Closed: Balancing Individual Privacy and Public Accountability in Data Sharing. https://doi.org/10.1145/3287560.3287577.

Zhang (2021). A commentary of Data trusts in MIT Technology Review 2021. https://doi.org/10.1016/j.fmre.2021.11.016.

Zrenner et al. (2019). Usage control architecture options for data sovereignty in business ecosystems. https://doi.org/10.1108/JEIM-03-2018-0058.

technopolis group

www.technopolis-group.com