

Einleitung

„Man könnte Design als einen Plan zur Anordnung von Elementen beschreiben, um einen bestimmten Zweck zu erreichen.“^a

Charles Eames

^aIm Original: One could describe Design as a plan for arranging elements to accomplish a particular purpose.

Diese Dissertation gestaltet ein Referenzmodell für föderierte Datendienste in industriellen Ökosystemen. Gemäß des Architekten und Designers Charles Eames, ist Design ein „Plan zur Anordnung von Elementen zur Erfüllung eines bestimmten Zwecks“ (Buur 2020). Im Folgenden werden daher die Motivation zu der Gestaltungsaufgabe und die zugrunde liegenden Probleme erläutert, die anschließend zu konkreten Zielen der Arbeit formuliert werden. Darauf folgt die Erläuterung des Forschungskontextes und des Forschungsdesigns, mit dem die Gestaltung erfolgt. Ebenfalls werden die Zielgruppen aufgezeigt, welche die Ergebnisse der Dissertation unterschiedlich nutzen können. Den Abschluss dieses ersten Kapitels bildet ein Überblick über die Struktur der Arbeit.

1.1 Motivation und Problembeschreibung

Die Welt befindet sich im Wandel. Schlagworte wie Industrie 4.0 und Gesellschaft 4.0 bezeichnen die Digitalisierung aller Lebensbereiche und deren Nutzung zur Generierung von Mehrwerten (Kollmann und Schmidt 2016, S. 3–4; Lasi et al. 2014, S. 261–262). Durch den Einsatz digitaler Technologien erzeugen Wirtschaft und Gesellschaft mehr Daten denn je. Ihre Nutzung bestimmt die Wohlfahrt aller Nationen (Drexl 2021, S. 11).

Digitale Technologien und Daten für die Wohlfahrt

„Digitale Technologien verändern die Wirtschaft und die Gesellschaft. Sie sind die Grundlage für neue Produkte und Dienstleistungen. Digitale Technologien werden durch die Nutzung von Daten entwickelt, z. B. für das Training von Systemen der künstlichen Intelligenz, und ihre Anwendung erzeugt unzählige neue Daten. Die bestmögliche Nut-

zung dieser Technologien und der von ihnen erzeugten Daten in voller Übereinstimmung mit den Grundwerten unserer vielfältigen und demokratischen Gesellschaft ist der Schlüssel zum wirtschaftlichen, sozialen und öffentlichen Wohlergehen.“^a

(Drexl 2021, S. 11)

^aIm Original: „Digital technologies transform the economy and society. They are the basis of new products and services. Digital technologies are developed by using data, such as for the training of artificial intelligence systems, and their application generates myriads of new data. Making best use of these technologies and the data they generate, in full compliance with the fundamental values of our diverse and democratic society, is key for economic, social and public welfare.“

Aus Sicht der Industrie ergeben sich dadurch Möglichkeiten zur Entwicklung innovativer datengetriebener Dienstleistungen und Produkte, wodurch ganz neue Geschäftsmodelle entstehen (Azkan et al. 2020, S. 1–2; Lusch und Nambisan 2015, S. 167). Außerdem können interne und betriebsübergreifende Prozesse optimiert werden, wodurch Durchlaufzeiten und Kosten gesenkt werden können (Vargo und Lusch 2011, S. 186; Wiebe et al. 2018, S. 17177). Viele dieser Dienstleistungen basieren auf der Nutzung künstlicher Intelligenz und dem Training von statistischen Modellen durch eine enorme Menge von Daten (Roh et al. 2021, S. 8). Ein Beispiel hierfür ist Airbus Skywise, das unter anderem vorausschauende Wartung für Flugzeuge anbietet. Hierfür werden die Flugdaten aller Maschinen eines bestimmten Typs ausgewertet um darin Muster zu erkennen mit denen Wartungsintervalle prognostizierbar gemacht werden können. Damit kann die Einsatzzeit der Maschinen erhöht und damit deren Betriebskosten gesenkt werden (Airbus o. J.).

Mit der zunehmenden Relevanz digitaler Geschäftsmodelle, datengesteuerter Optimierung und der Notwendigkeit von Informationsvorteilen müssen Unternehmen somit auf eine vielfältige und große Menge von Daten zugreifen (Legner, Hüner et al. 2020; Otto, Mohr et al. 2020). So titelt *The Economist* (2017) „die wertvollsten Ressourcen der Welt sind nicht mehr Öl, sondern Daten“.¹ Nur wenige Organisationen können auf diese Mengen zurückgreifen, was zu einem Ungleichgewicht im Wettbewerb zwischen großen Plattformfirmen und kleinen und mittleren Unternehmen sowie großen Unternehmen wie Automobil-OEMs führt. Während große Plattformunternehmen² durch die Bereitstellung der Infrastruktur für weite Teile der Industrie und des Internets Zugang zu der notwendigen Datenmenge besitzen, haben kleinere Unternehmen das Nachsehen (Otto, Mohr et al. 2020; Nuccio und Guerzoni 2019, S. 323). Die Marktmacht der Plattformunternehmen steigt dadurch weiter, wenngleich in der Europäischen Union

¹Im Original: „The world’s most valuable resource is no longer oil, but data“.

²Beispielsweise Amazon, Microsoft oder Alphabet (ehem. Google).

regulatorisch gegengesteuert wird.³ Dies führt dazu, dass der Wettbewerbsdruck auf Industrieunternehmen stetig ansteigt, da ein Teil der Wertschöpfung nicht mehr innerhalb der Industrie, sondern durch die Technologieunternehmen, als neue Marktteilnehmer, erzielt wird (Biglaiser et al. 2019, S. 3–5; Steur 2018, S. 45).

Aus der steigenden Nachfrage nach Daten ergeben sich für Unternehmen weitere Hindernisse. Die Datennutzung lässt sich schwer skalieren. So werden beispielsweise mehr als 80 % des (Zeit-)Budgets von Datenanalyseprojekten für die Aufbereitung der Daten und nicht für Ergebnisse eingesetzt (Gabernet 2017). Mehr Daten würden das bestehende Problem nur noch verschärfen. Unternehmen haben bereits mit der Datenproduktivität zu kämpfen und ertrinken in ihren Data Lakes. Zum anderen scheinen trotz der Datenschwemme immer wieder die richtigen Daten zu fehlen, zu wenige Signale gehen in immer mehr Rauschen unter (Fan et al. 2014, S. 298). Zudem explodiert die Datenmenge seit der Einführung digitalisierter Prozesse, wie dem Internet der Dinge oder der allgemeinen Prozessautomatisierung (IDC 2020). Dies führt letztlich zu einer enormen Komplexität bei der Verarbeitung und Speicherung von Daten (Hsinchun Chen et al. 2012, S. 1166). Unternehmen nutzen ihren Datenschatz nicht effektiv, und so werden derzeit nur 0,5 % der gesammelten Daten analysiert (Burn-Murdoch 2012). Darüber hinaus prognostiziert die *International Data Corporation*, dass die Menge der gesammelten Daten in den nächsten Jahrzehnten dramatisch ansteigen wird, was diese Probleme noch verschärft (IDC 2020).

Zur Lösung dieses Problems wird die gemeinsame Nutzung von Daten als erfolgversprechende Maßnahme gesehen (European Commission 2021; Legner, Hüner et al. 2020; Otto, Mohr et al. 2020). Hiervon können viele Unternehmen profitieren, die aktuell nicht in der Lage sind derartige Modelle zu entwickeln, da ihnen der Zugang zur erforderlichen Menge an Daten für deren Training fehlt. Dieses Ziel hat auch die Europäische Kommission erklärt und in diesem Kontext den *Data Governance Act* eingebracht, der darauf abzielt, die gemeinsame Nutzung der erzeugten Daten innerhalb der Europäischen Union zu fördern. Die gemeinsame Nutzung von Daten wird somit allgemein als Schlüssel für eine Revolution hin zu einer datengesteuerten Wirtschaft und damit als entscheidend für künftiges Wirtschaftswachstum und Wohlstand angesehen. Der für den Binnenmarkt zuständige EU-Kommissar Thierry Breton kommentierte dies wie folgt:

Data Governance Act

„Mit dem heute verabschiedeten Data Governance Act legen wir einen gemeinsamen Ansatz für die gemeinsame Nutzung von Daten fest - den

³Unter anderem soll der *Digital Markets Act* der EU den digitalen Sektor fairer gestalten und dazu dienen, die großen Plattformunternehmen zu regulieren.

europäischen Weg. Wir erleichtern den Fluss der zunehmenden Industriedaten über Sektoren und Mitgliedstaaten hinweg, um Europa dabei zu helfen, der Datenkontinent Nummer eins in der Welt zu werden.“^a

(European Commission 2021)

^aIm Original: „With the Data Governance Act agreed today, we are defining a common approach to data sharing – the European way. We are facilitating the flow of growing industrial data across sectors and Member States to help Europe become the world’s number one data continent.“

Die gemeinsame Nutzung von Daten scheint also eine Antwort auf die unzureichende Nutzung von Datenbeständen zu sein und damit deren Wert zu steigern (Lavie 2006, S. 644–645). Im Gegensatz zum klassischen Datenaustausch, der im Fall von SWIFT⁴ bereits seit 1977 weite Verbreitung findet und die damalige Telex-Systeme ersetzte⁵, wird bei der gemeinsamen Datennutzung der direkte Zugriff auf Daten gewährt (Jussen et al. 2023, S. 4; Vesselkov et al. 2019, S. 303). Eine der zentralen Herausforderungen bei der gemeinsamen Nutzung von Daten besteht jedoch darin, die Interessen des Einzelnen (Datenanbieter) und der Gemeinschaft (Ökosystem, Gesellschaft usw.) auszugleichen, um die Netzwerkeffekte von Daten zu nutzen (Gregory et al. 2021; Jarke et al. 2019). Somit bleibt die Frage der Datensouveränität eines der Schlüsselthemen der Wirtschaft der gemeinsamen Datennutzung (Jarke et al. 2019). Wie Cloud- und Webdienste zeigen, können gemeinsam genutzte Ressourcen die Effizienz und Flexibilität komplexer Systeme über Organisationsgrenzen hinweg erhöhen (Baskerville, Cavallari et al. 2010, S. 48). Während die gemeinsame Nutzung von Infrastrukturen, z. B. mit Cloud Computing und Anwendungen wie Web Services, zum Standard gehört, ist die gemeinsame Nutzung von Daten heute weitgehend unerforscht und wird nicht ausreichend genutzt (Legner, Hüner et al. 2020; Opriel et al. 2021). Grundsätzlich streben Organisationen danach, Wettbewerbsvorteile durch den Erwerb von Ressourcen und Fähigkeiten zu erlangen (Barney 1991, S. 99; Wade und Hulland 2004, S. 108). Aufgrund der zunehmenden Vernetzung von Unternehmen ist die gemeinsame Nutzung von Ressourcen über Unternehmensgrenzen hinweg zu einer wertvollen Quelle geworden, da sie die Auslastung der Ressourcen erhöht (Gulati und Gargiulo 1999, S. 413–414; Lavie 2006, S. 644–645; Phelps 2010, S. 905–906). Aus Sicht der Industrie ist dies zu einem der wichtigsten Zukunftsthemen geworden, wie ein Architekt eines der größten Telekommunikationsausrüster der Welt bestätigt.

⁴Die „Society for Worldwide Interbank Financial Telecommunication (SWIFT)“ wurde 1973 von 239 Banken gegründet, um die Kommunikation bei grenzübergreifenden Zahlungen zu standardisieren (SWIFT o. D.).

⁵Das Telex-System wurde bereits 1933 von der deutschen Reichspost eingeführt (Union 1984).

Data Sharing ist die Zukunft

„Ich glaube, das ist das, was kommen wird. Das Potenzial ist enorm! Das ist ja letztendlich die Digitalisierung, wovon wir reden und Industrie 4.0 und Digital Twin. Das ist unausweichlich, die werden kommen. Und vor allem das produzierende Gewerbe, die Industrie, muss sich fit machen. Nur dann werden sie die Möglichkeit haben, an den Kunden tatsächlich die Dienste in fünf Jahren anzubieten, ansonsten haben sie in fünf Jahren selbst keine Möglichkeit mehr an den Kunden heran zu kommen.“

Chief Architect Telekommunikation

Während die Relevanz gemeinsamer Datennutzung für Politik und Wirtschaft deutlich ist, steht die Erforschung der Grundlagen gemeinsamer Datennutzung noch am Anfang. Das Phänomen datenzentrierter Plattformen ist bislang nur unzureichend beschrieben (Otto und Jarke 2019, S. 576) und es fehlt laut Jussen et al. (2023, S. 8) an „Instrumenten oder Vorgaben, die die Komplexität des Themas für den Endnutzer verringern und somit die Anfänge der gemeinsamen Nutzung von Daten leichter zugänglich machen würden“⁶. Darüber hinaus fehlt ein Verständnis darüber wie der organisatorische Rahmen der gemeinsamen Datennutzung aussehen kann. Aufgrund der hohen Dynamik und Anforderungen an die Selbstorganisation solcher Meta-Organisationen, wird in dieser Arbeit das Ökosystem als Organisationsform vorgeschlagen (Guggenberger, Möller, Haarhaus et al. 2020, S. 12; Otto 2022a, S. 45). Zuletzt gibt es keine Einigkeit darüber, welche technischen Fähigkeiten für die gemeinsame Datennutzung notwendig sind und wie diese innerhalb eines föderierten Systems verteilt werden können (Bechtsis et al. 2022, S. 4412).

Mehrere Forschungsprojekte, politische Initiativen und Praxiskonsortien arbeiten an der Bereitstellung von Organisationsformen, Standards und Technologie für die gemeinsame Nutzung von Daten, um diese Probleme anzugehen. Im Zentrum der Projekte stehen dabei föderierte Datendienste in industriellen Ökosystemen. Ökosysteme, die auf föderierten Strukturen basieren, sind weniger anfällig für Kartell- und Monopolbildung, da sie dezentral und ohne eine zentrale Macht organisiert sind (Guggenberger, Möller, Haarhaus et al. 2020, S. 9–10). Sie werden von einem Konsortium verwaltet, das die Souveränitätsprozesse koordiniert und kontrolliert (Otto 2022b, S. 11–13). Das erste Projekt, das dies adäquat adressierte, war International Data Spaces (IDS). Im Jahr 2019 wurde die aktuelle Version des Referenzarchitekturmodells in der Version 3.0 (RAM 3.0) veröffentlicht. Auf Basis des RAM 3.0 wurden mehrere funktionsfähige Implementierungen umgesetzt, so ein Prototyp von T-Systems der Deutschen Tele-

⁶Im Original: „However, there is currently a lack of tools or guidelines that would reduce the complexity of the topic for the end-user and thus make the beginnings of data sharing more accessible“.

kom und der Niederländischen Organisation für Angewandte Naturwissenschaftliche Forschung (TNO). Auf dieser Architektur aufbauend wurden Gaia-X und weitere Datenraumprojekte wie Catena-X, das den Großteil der deutschen Automobilindustrie umfasst, und der Mobility Data Space, der sich auf den Mobilitätssektor bezieht, gestartet.

Das Ökosystem, das im Kern aller dieser Projekte steht, stellt eine Meta-Organisation dar (Kretschmer, Leiponen et al. 2020, S. 6). Ökosysteme sind also „Organisationen, die mehrere rechtlich autonome Einheiten umfassen“ (Gulati, Puranam et al. 2012, S. 571) und das gemeinsame Ziel haben, Daten zwischen Organisationen zu teilen. Daher sind Ökosysteme von zunehmender Bedeutung für Zwecke, die kollektive Aktionen zwischen lose gekoppelten Organisationen erfordern (Berkowitz und Dumez 2016, S. 150). Letztlich trägt dies zu den oben genannten Vorteilen bei und fördert die Möglichkeiten des dynamischen Zugriffs auf Daten aus verschiedenen Quellen. Föderierte Datendienste für industrielle Ökosysteme, die Autonomie und Flexibilität für alle Parteien ermöglichen, sind geeignet, diese Anforderungen zu erfüllen. Im Gegensatz zu zentralisierten Plattformkonzepten, die regelmäßig von einzelnen Akteuren dominiert werden (Jacobides et al. 2018, S. 2257), wird das Konsortium zum zentralen Akteur. Diese Meta-Organisation entscheidet kollaborativ über wesentliche Modalitäten der Plattform und damit über das gesamte Ökosystem. Projektübergreifend gibt es jedoch kaum einen Konsens über die Fähigkeiten und Dienste, die erforderlich sind, um gemeinsame Datennutzung souverän über föderierte Datendienste zu ermöglichen.

Data Ecosystem at BMW

„Für die BMW Group ist ein konsistentes Daten-Ökosystem entscheidend, um Portabilität, Interoperabilität sowie Interkonnektivität und Datenhoheit für verschiedenste Anwendungsbereiche zu erhöhen. Der DataSpace Connector ist eine innovative, zentrale Komponente, die den souveränen Austausch von Daten nach definierten Richtlinien und Prinzipien im Sinne von Gaia-X ermöglichen kann.“ ^a

Marco Görgmaier, Head of Data Transformation & Artificial Intelligence at BMW Group

^aIm Original: „For the BMW Group a consistent data ecosystem is crucial to increase portability, interoperability as well as interconnectivity and data sovereignty for a wide variety of fields of application. The DataSpace Connector is an innovative, central component, which can enable the sovereign exchange of data according to defined guidelines and principals in line with Gaia-X.“

1.2 Ziele der Arbeit

Die vorliegende Dissertation zielt darauf ab, die zuvor bezeichneten Problemstellungen zu adressieren und verfolgt dabei das übergeordnete Ziel, ein Modell für Datendienste zu entwickeln, welches die gemeinsame Datennutzung zwischen Organisationen unterstützt. Das Modell soll in Form eines Referenzmodells für die Gestaltung föderierter Datendienste in industriellen Ökosystemen erarbeitet werden. Durch die Entwicklung eines Referenzmodells soll ein Konsens über die Projekte, die sich mit föderierten Datendiensten beziehungsweise Datenräumen beschäftigen, hergestellt werden. Durch das wiederverwendbare und konfigurierbare Modell soll letztlich die Entwicklung von föderierten Datendiensten in industriellen Ökosystemen standardisiert werden und damit die Konstruktionsleistung für die Anwender reduziert werden, indem sie sich diese auf das Modell beziehen können (vom Brocke und Fettke 2019).

Daraus ergibt sich die Hauptforschungsfrage, welche leitend für diese Thesis ist und nach der Gestaltung von föderierten Datendiensten in industriellen Ökosystemen fragt. Ausgehend von der Leitfrage werden vier Teilfragen abgeleitet, deren Beantwortung dem konventionellen Vorgehensmodell einer wissenschaftlichen Arbeit folgt und sich von einem sehr generellen und breiten Blick auf die Problematik zu einer spezifischen und detaillierten Untersuchung der leitenden Fragestellung entwickelt. Tabelle 1.1 zeigt eine Übersicht aller Forschungsfragen und deren Beiträge. Dabei werden zwei Gruppen von Fragen unterschieden: Während die ersten zwei Forschungsfragen deskriptiver Natur sind, sind die Folgenden präskriptiv.

Tabelle 1.1: Zusammenfassung der Forschungsfragen und jeweiligen Beiträge.

#	Forschungsfrage	Beitrag
FF	Wie können föderierte Datendienste in industriellen Ökosystemen gestaltet werden?	Gestaltungsprinzipien für föderierte Datendienste in industriellen Ökosystemen
FF1	Welche Ökosystemtypen lassen sich identifizieren?	Typologie der Ökosystemkonzepte in der Literatur
FF2	Welche Eigenschaften haben diese Typen?	Analyse der Eigenschaften der Typen aus FF1
FF3	Wie muss eine Organisation gestaltet sein, die gemeinsame Datennutzung ermöglicht?	Anforderungen an föderierte Datendienste für industrielle Ökosysteme
FF4	Welche Dienste werden für föderierte Datendienste in industriellen Ökosystemen benötigt?	Referenzmodell für Datendienste in industriellen Ökosystemen

In dieser Arbeit soll ein Referenzmodell für föderierte Datendienste in industriellen Ökosystemen entwickelt werden, welches eine interorganisationale Organisation zur gemeinsamen Nutzung von Daten ermöglicht. Damit muss ein interorganisationales Informationssystem (IOS) gestaltet werden, welches die notwendigen Anforderungen erfüllt und damit zweckgerichtet ist. IOS haben durch mannigfaltige technologische Entwicklungen in den letzten Jahrzehnten eine starke Wandlung erlebt, die unter anderem zur Entstehung sogenannter Plattformökosysteme geführt haben. Neben dem Konzept der Plattformökosysteme lässt sich in der Literatur eine Vielzahl unterschiedlicher Konzeptionen dieses Begriffs identifizieren, was eine einheitliche Charakterisierung des Begriffs erschwert. Ausgehend von dem Ursprungskonzept des Geschäftsökosystems (Moore 1993) werden heute über 23 unterschiedliche Konzepte verwendet (S. Hyrynsalmi und S. M. Hyrynsalmi 2019), welche zum Teil überlappende Charakteristika und Spezifizierungen aufweisen und damit zu einer Unschärfe der Begriffe führen (S. Hyrynsalmi und S. M. Hyrynsalmi 2019, S. 1). Aus diesem Grund zielt die erste Forschungsfrage darauf ab, die Charakteristika von den jeweiligen Konzepten zu entkoppeln und zu allgemeinen Ökosystemtypen neu zu arrangieren. Dies ist relevant für die Entwicklung des Referenzmodells, da erst hierüber das industrielle Ökosystem eindeutig spezifiziert werden kann. Daraus ergibt sich nun die erste Forschungsfrage:

Forschungsfrage 1

Welche allgemeinen Ökosystemtypen lassen sich identifizieren?

Die zweite Forschungsfrage baut unmittelbar auf der ersten Frage auf und fragt nach den jeweiligen Eigenschaften der identifizierten Typen. Hierbei werden die Eigenschaften und Charakteristika der ursprünglichen Ökosystemkonzepte verallgemeinert und auf die identifizierten Typen zurück übertragen. Somit können die Ökosystemtypen, unabhängig von ihrer ursprünglichen Ausrichtung, diskutiert werden und bilden somit idealtypische Muster. Diese Forschungsfrage zielt vor allem darauf ab, dass die Charakteristika eines geeigneten Ökosystemtyps für föderierte Datendienste identifiziert werden können. Für die Entwicklung des Referenzmodells ist dies relevant, da diese Eigenschaften einen Einfluss auf die Gestaltung der föderierten Datendienste in industriellen Ökosystemen haben.

Forschungsfrage 2

Welche Eigenschaften haben diese Typen?

Aufbauend auf den vorherigen Forschungsfragen wird in der Folge nun ein Typ genutzt, welcher aufgrund seiner spezifischen Eigenschaften die Zielerreichung bestmöglich unterstützt. Dieser Idealtyp eines industriellen Ökosystems wird im Folgenden weiter spezifiziert und aufgrund seines föderalen Charakters als Meta-Organisation verstanden (Gulati, Puranam et al. 2012; Kretschmer und Khashabi 2020). Innerhalb dieser

Organisation soll die gemeinsame Nutzung von Daten durch föderierte Datendienste ermöglicht werden, was im Gegensatz zu klassischen Beziehungsstrukturen bspw. Supply Chain Management Plattformen, eine neuartige Austauschbeziehung darstellt. Aus diesem Grund können auch physische Prozesse nicht einfach digital abgebildet werden, sondern sie müssen vielmehr eine Gestalt annehmen, die diese Austauschbeziehung unterstützt. Daraus ergibt sich nun Forschungsfrage 3:

Forschungsfrage 3

Wie muss eine Organisation gestaltet sein, die gemeinsame Datennutzung ermöglicht?

Die zuvor definierten grundlegenden Anforderungen an die Organisation müssen im nächsten Schritt mit einer Funktionsarchitektur hinterlegt werden. Damit wird also die Frage nach den konkreten Diensten zur Durchführung dieser Austauschbeziehungen und zur Etablierung der Organisation adressiert. Die Menge an Diensten und ihre spezifischen Interaktionen stellen letztlich das Referenzmodell für föderierte Datendienste in industriellen Ökosystemen dar.

Forschungsfrage 4

Welche Dienste werden für föderierte Datendienste in industriellen Ökosystemen benötigt?

1.3 Forschungskontext und -design

Das Hauptziel dieser Dissertation ist die Entwicklung eines Referenzmodells für die Gestaltung föderierter Datendienste in industriellen Ökosystemen, mit der eine zwischenbetriebliche Organisation zur gemeinsamen Nutzung von Daten realisiert werden kann. Grundsätzlich ist dieser Forschungsbeitrag damit in der Wirtschaftsinformatik (WI)⁷ zu verorten, die sich als eigenständige wissenschaftliche Disziplin etabliert hat. Mit der Verbreitung von Informations- und Kommunikationstechnologie (IuK-Technologie) ab den 1950er Jahren hat sich die WI vor allem aus der Betriebswirtschaftslehre heraus entwickelt und ihren Gegenstandsbereich über die Zeit deutlich erweitert und vertieft (L. J. Heinrich, Heinzl et al. 2011, S. 13–14). Der maßgebliche Forschungsgegenstand wird heute als das Informationssystem und dessen einzelne Elemente betrachtet. WI kann gleichermaßen als Realwissenschaft als auch Ingenieurwissenschaft betrachtet werden (Nunamaker, M. Chen et al. 1990, S. 97–98; L. J. Heinrich, Heinzl et al. 2011, S. 48–49). Als Realwissenschaft hat die WI das Ziel, Informationssysteme zu beschreiben und zu verstehen und aus dem verallgemeinerten Wissen über diese Systeme Artefakte für

⁷Beziehungsweise deren Angloamerikanischen Pendant, dem Information Systems Research (ISR).

deren zielgerichteten Einsatz zu entwickeln (A. S. Lee und Baskerville 2003, S. 241; Hevner et al. 2004, S. 76). Letzteres begründet die WI als Ingenieurwissenschaft. Das übergreifende Ziel der WI ist das Verständnis dafür, wie aus Wert aus Informationen generiert wird und diese Wertschöpfung optimiert werden kann (Nunamaker, Briggs et al. 2015, S. 11). Im Folgenden wird ausgehend von dieser Einordnung der Kontext diskutiert.

Diese Dissertation folgt dem *Design Science Research (DSR)* Paradigma. Im Gegensatz zur Verhaltensforschung, die sich mit dem Verständnis von Zusammenhängen innerhalb und zwischen Organisationen und Individuen beschäftigt und daraus Theorien zur Erklärung und Vorhersage von Phänomenen ableitet (Ursache-Wirkung-Fokus), wird im DSR Gestaltungswissen über Artefakte erzeugt (Stahl 2009, S. 118; Hevner et al. 2004, S. 76; S. T. March und G. F. Smith 1995, S. 253–254). „Gestaltungswissen bezieht sich auf Mittel-Zweck-Beziehungen zwischen Problem- und Lösungsräumen“ (vom Brocke, Winter et al. 2020, S. 520; Venable et al. 2016, S. 17). Typische Ergebnisse der Verhaltensforschung sind mathematische Modelle, die den Einfluss unabhängiger Variablen auf abhängige Variablen angeben. Ergebnisse im DSR sind Gestaltungstheorien und Artefakte (Gregor und Hevner 2013, S. 339–341). Gestaltungstheorien gehen über diese normativen und deskriptiven Theorien hinaus und bieten präskriptive Handlungsempfehlungen für Gestaltungsprozesse (Walls et al. 1992, S. 49). Artefakte hingegen sind künstlich erzeugte Dinge, die einen Schöpfer, Zweck und Funktionalitäten aufweisen. Sie verbinden durch einen zielgerichteten Gestaltungsprozess den Problem- mit dem Lösungsraum. Typische Artefakte in der WI sind Modelle, Methoden, Konstrukte und Instantiierungen (S. T. March und G. F. Smith 1995, S. 253). Im Folgenden werden die in dieser Dissertation gestalteten Artefakte, in der Reihenfolge ihrer Entwicklung, erläutert:

- *Typologie*: In dieser Dissertation wird eine Typologie von Ökosystemen gestaltet, die eine Klassifikationsmethode darstellt (Bailey 1994, S. 1; Paré et al. 2015, S. 184). Nach Bailey (1994, S. 4) ist eine Klassifikation der „generelle Prozess der Gruppierung von Entitäten durch Gemeinsamkeiten“⁸. Es lassen sich zwei Ansätze zur Klassifikation von Merkmalsträgern unterscheiden: Bei einer Typologie werden *Typen* konzeptionell hergeleitet, bei einer Taxonomie erfolgt eine empirische Ableitung von *Taxonen* (Lambert 2015, S. 53; Baden-Fuller und M. S. Morgan 2010, S. 161). Im Gegensatz zu einer Taxonomie, bei der die Merkmale empirisch identifiziert werden, werden hier Merkmale aus der Theorie abgeleitet und damit Idealtypen oder „Gedankenbilder“ (Weber 1949, S. 90) erzeugt. Damit wird eine Möglichkeit geschaffen, Abweichungen realer Objekte von die-

⁸Im Original: „... general process of grouping entities by similarity.“

sen Idealtypen festzustellen und eine Gruppe von idealisierten Merkmalsträgern geschaffen, anhand derer weitere Untersuchungen durchgeführt werden können (Doty und Glick 1994, S. 243; Blalock 1969, S. 30; McKinney 1966, S. 65–66).

- *Gestaltungsprinzipien*: Gestaltungsprinzipien sind Meta-Artefakte, die eine allgemeine Lösung für eine Klasse von Artefakten darstellen (Baskerville et al. 2018; Iivari 2015) und damit eine Gestaltungstheorie darstellt (Gregor und Hevner 2013, S. 341). Sie bieten kodifiziertes Gestaltungswissen als präskriptive Richtlinien (Hevner et al. 2004, S. 95; Möller, Guggenberger et al. 2020, S. 2), die, innerhalb spezifischer Randbedingungen (Gregor, Kruse et al. 2020, S. 1637; Chandra Kruse, Seidel und Gregor 2015, S. 4044), Handlungen anleiten und gleichermaßen einschränken (Baskerville und Pries-Heje 2010, S. 272). Damit ermöglichen Gestaltungsprinzipien eine Hilfestellung für Entwickler, um den Gestaltungsprozess effizienter zu gestalten (Chandra Kruse, Seidel und Puraio 2016, S. 39; McAdams 2003, S. 357). Darüber hinaus sind sie ein Medium, um Erkenntnisse an Manager und Technologieexperten zu vermitteln (Hevner et al. 2004, S. 95). In dieser Thesis werden Gestaltungsprinzipien entwickelt, die den präskriptiven Aspekt des Referenzmodells abdecken und damit einen Beitrag zur Wissensakkumulation im Bereich der föderierten Datendienste in industriellen Ökosystemen liefern.
- *Referenzmodell*: Referenzmodelle sind Modelltypen⁹, welche „ein gewisses Maß der Allgemeingültigkeit aufweisen“ (Leimeister 2015, S. 222), durch welche spezifische Modelle der gleichen Klasse gestaltet werden können (Leimeister 2015, S. 222; Winter, Gericke et al. 2009, S. 3). Referenzmodelle stellen somit Idealmodelle dar, welche für die Lösung einer Klasse von Probleme herangezogen werden können (L. J. Heinrich, Heinzl et al. 2011, S. 29). Einerseits kann damit die Entwicklung, bspw. von Organisationssystemen oder Anwendungssystemen, geplant werden, andererseits kann ein Referenzmodell für den Vergleich von Istmodellen (oder Systemen) herhalten (Schlagheck 2000, S. 54; Schütte 1998, S. 69). Der Nutzer eines Referenzmodells passt dieses an die Anforderungen des spezifischen Falls an (Thomas und Scheer 2016, S. 3). In dieser Dissertation beschreibt zunächst das Referenzmodell der Dienste deskriptiv die Klasse der Modelle, die Gestaltungsprinzipien wiederum sind ein präskriptiver Vorschlag, nach dem die Klasse von Modellen gestaltet werden kann. Referenzmodelle werden mit der Referenzmodellierung entwickelt (Becker und Knackstedt 2002; vom Brocke 2007).

⁹Modelle bilden ein natürliches oder künstliches Realweltobjekt ab, welches wiederum ein Modell sein kann. Dabei wird die Abbildung in der Regel verkürzt und nur wesentliche, für die Anwendung relevante Aspekte abgebildet. Das Original muss dem Modell dabei nicht eindeutig zugewiesen sein, eine Ersetzungsfunktion wird für bestimmte Nutzer, bestimmte Zeitabschnitte und bestimmte theoretische und praktische Operationen erfüllt (Stachowiak 2013, S. 131–133). Ein Modell ist immer das Ergebnis eines Gestaltungsprozesses (Frank 2007, S. 123).

Zur Gestaltung dieser Artefakte werden unterschiedliche *Methoden* zur Datenerhebung, -analyse und Konstruktion eingesetzt, die im Folgenden dargelegt werden. Jede Methode wird in einem separaten Abschnitt detailliert diskutiert und ihre Nutzung begründet.

- Bei der *empirischen Datenerhebung* werden einerseits Dokumente identifiziert und andererseits Experteninterviews durchgeführt. Die Datenerhebung wird angelehnt an die Fallstudienforschung nach Yin (2017) durchgeführt. Dazu werden einerseits Dokumente zu den Fällen von Catena-X, Gaia-X und dem International Data Space untersucht. Diese Daten werden dann ergänzt durch eine Interviewstudie. Experteninterviews sind eine der verbreitetsten und damit wichtigsten Datenerhebungsverfahren der qualitativ-empirischen Forschung (Baur und Blasius 2014, S. 53; Myers und Newman 2007, S. 2–3). Darüber hinaus werden zur Evaluierung der Ergebnisse ebenfalls empirische Daten in Form von Expertenbefragungen durchgeführt.
- Die *Datenanalyse* wird mittels qualitativer Methoden durchgeführt. So wird an mehreren Stellen die Literaturanalyse, als eine der wesentlichen Formen der Sekundärdatenanalyse, angewandt. Vor allem zur Bestimmung der theoretischen Grundlagen, zur Ermittlung des Standes der Technik und der Charakteristika der Ökosysteme. Ergänzend wird in Letzterem die qualitative Inhaltsanalyse eingesetzt, eine Methode zur Auswertung von Texten, die im Rahmen empirischer Sozialforschung anfallen (Mayring und Fenzl 2014, S. 543). Durch die Anwendung der Inhaltsanalyse erfolgte die Auswertung der Literatur strukturierter und formeller, wodurch ein hoher Grad der Intersubjektivität der Ergebnisse erzielt werden kann (Mayring und Fenzl 2014, S. 543). Die qualitative Inhaltsanalyse kommt ferner zum Einsatz, um die mittels einer Experteninterviewstudie generierten Primärdaten auszuwerten.
- Zudem werden *Konstruktionstechniken der Referenzmodellierung* eingesetzt, die „spezifische Regeln zur Übernahme, Adaption und Ergänzung von Modellinhalten“ für die Wiederverwendbarkeit von Referenzmodellen darstellen (vom Brocke 2018). Bei der Gestaltung des Referenzmodells wird eine Kombination aus der Konfiguration und der Analogie angewandt (Vgl. vom Brocke 2003, S. 239–284, 308–315). Die Konfiguration folgt dem Prinzip der Reduktion eines Gesamtmodells und die Analogie bezeichnet eine Technik, bei der das Ergebnismodell (Anwendung) mit dem Ursprungsmodell (Referenzmodell) hinsichtlich eines spezifischen Merkmals übereinstimmt (Becker und Knackstedt 2002, S. 26–28; vom Brocke 2018). Die Auswahl liegt darin begründet, dass die Anwendungsdomäne des Modells sehr komplex ist, weswegen die Konstruktionsleistung für die ein-

zelenen Dienste zum Anwender des Modells verlagert werden (Vgl. hierzu vom Brocke 2018). Andererseits sollen in der Gesamtbetrachtung die gesamten Modellinhalte bereitgestellt und auch die Schnittstellen zwischen den Teilmodellen (Dienste) definiert werden. Dadurch kann die Anwendungskomplexität reduziert werden, da die Anwender für einzelne Dienste einen Überblick über das Gesamtsystem haben und gleichzeitig die Ausgestaltung des Dienstes in Abhängigkeit vom grundsätzlichen Funktionsprinzip weitgehend selbst entwickeln können.

Durch die Integration formaler Methoden und Erhebung empirischer und praxisrelevanter Daten kann die Forderung des DSR nach Rigorosität und Relevanz adressiert werden (Hevner et al. 2004, S. 79–80; Gregor und Hevner 2013, S. 349–350). Ein Unternehmensarchitekt hat dies deutlich auf den Punkt gebracht:

Zwei Seiten vom Pferd fallen

„Ich meine auf der einen Seite hast du da die Theorie und die Forschung. Da ist es natürlich fein, wenn man versucht möglichst tief den Sachverhalt zu verstehen, die Architektur zu verstehen. Das andere Extrem ist dann die Anwendung. Und da spielt dann die Exzellenz und die Vollständigkeit von einem theoretischen Modell manchmal nicht die treibende Rolle, sondern da ist dann die spannende Frage, welche Probleme kannst du denn lösen und wie kommst du über eine taktische Ebene dann in diese Fachlichkeit rein. Und das ist auch eigentlich genau der Klinsch, den man dann meistens hat. Also man kann da meistens von zwei Seiten vom Pferd fallen. Entweder indem man zu pragmatisch dran geht und dann nicht genug Substanz hat um da wirklich in eine Skalierung zu kommen oder indem man zu viel von der Theorie kommt und nicht in eine praktische Anwendung kommt und dann investiert man einen Haufen Geld in Struktur und Methoden und Abläufe und man löst genau null Probleme und dann ist es auch kaputt und das ist genau, würde ich sagen, diese Balance abzumischen, das was ich versuche in Bosch, als auch in Catena-X.“

Unternehmensarchitekt Automotive

Zusammenfassend werden in dieser Dissertation Ökosysteme als Analyseeinheit untersucht, innerhalb derer die föderierten Datendienste die gemeinsame Verwendung von Daten ermöglichen sollen (Vgl. Tabelle 1.2). Der Fokus liegt dabei auf der Industrie, es werden also keine Endkunden in die Betrachtung mit eingezogen. Innerhalb des Forschungsprozesses wird sowohl deskriptives als auch präskriptives Wissen generiert. Letztlich wird durch die Gestaltungsprinzipien ein wichtiger Beitrag zur Theorieentwicklung von Datendiensten in industriellen Ökosystemen geleistet.

Tabelle 1.2: Rahmenwerk dieser Dissertation.

Begriff	Anwendung
Analyseeinheit	(industrielle) Ökosysteme
Untersuchungsgegenstand	Hauptsächlich Catena-X, nebensächlich Datenraum Mobilität, International Data Spaces und Gaia-X
Domäne	Industrie
Wissenstypen	Deskriptive Beschreibung von Sachverhalten und präskriptives Wissen in Form von Gestaltungsprinzipien und einem Referenzmodell

1.4 Zielgruppe und Nutzen des Referenzmodells

Diese Dissertation adressiert aufgrund ihres interdisziplinären Charakters eine Reihe unterschiedlicher Zielgruppen.¹⁰ Dabei ist besonders der hohe Praxisbezug des Forschungsgegenstandes hervorzuheben, aufgrund dessen einerseits strategische Initiativen als auch Anwendungs- und Architekturentwickler von den Ergebnissen profitieren können. Ferner soll die Relevanz für die Politik betont werden, deren Fokus auf den Aufbau industrieller Ökosysteme unverkennbar ist.¹¹ Zuletzt adressiert diese Arbeit selbstverständlich die Gruppe der Forschenden der Wirtschaftsinformatik.

- *Organisationsentwickler und Strategiemanager* können die Typologie der Ökosysteme nutzen, um ihre Stellung in der Vielzahl von Ökosystemen, in denen Organisationen heute partizipieren, besser zu verstehen und strategische Initiativen zielgerichtet auszurichten. Das Referenzmodell und besonders das Rollen- und Interaktionsmodell hilft ihnen dabei, detaillierte Maßnahmen abzuleiten und das Potenzial, das die gemeinsame Datennutzung bietet, vollständig zu nutzen.
- *Softwareentwickler und -architekten* können die Artefakte dieser Dissertation auf unterschiedliche Art und Weise nutzen. Für sie ist besonders das Referenzmodell und die Beschreibung der einzelnen Dienste relevant, da sie hierüber Softwareartefakte erstellen können und diese vertreiben beziehungsweise selbst betreiben können. Das Referenzmodell bietet dafür einen Ausgangspunkt für die Zusammenhänge innerhalb des zu gestaltenden Systems, die Schnittstellen und die Funktionsstruktur der Subsysteme, also der einzelnen Dienste. Entsprechend

¹⁰Insgesamt leiten sich die jeweiligen Beiträge aus der Motivation dieser Arbeit ab, sollen aber an dieser Stelle noch einmal spezifiziert und nach Zielgruppen getrennt aufgezeigt werden, um die Relevanz für die jeweiligen Gruppen darzustellen.

¹¹Hervorzuheben ist in diesem Kontext besonders der *Data Act* und der *Data Governance Act* (Vgl. Abschnitt 1.1).

kann durch die sachgemäße Anwendung des Referenzmodells die Abstimmung zwischen den einzelnen Subsystemen vereinfacht werden und funktionale Kompatibilität erzeugen.

- *Politische Entscheidungs- und Mandatsträger* können die Typologie der Ökosysteme nutzen, um spezielle Initiativen zu fördern. Außerdem kann das Referenzmodell dazu dienen, dass besonders zentrale Dienste, die eine Schlüsselrolle für die Gleichberechtigung der Teilnehmer an einem Ökosystem und die Markteffizienz einnehmen, besonders zu betrachten. Hierüber können auch gezielte Fördermaßnahmen oder staatliche Eingriffe, in Anlehnung an den Data Governance Act oder Digital Markets Act, abgeleitet werden.
- *Forschende der Wirtschaftsinformatik* profitieren in zweierlei Hinsicht von den Ergebnissen dieser Dissertation. Vor allem die Systematisierung der Ökosystemkonzepte und die Entwicklung des Referenzmodells für föderierte Datendienste stellen für diese Gruppe einen besonderen Mehrwert dar, da hier wichtige Forschungslücken geschlossen werden. Die Gestaltungsprinzipien für das Referenzmodell stellen ein wesentliches Artefakt für die Forschung dar, da sie einen Beitrag zur Theorieentwicklung leisten.

1.5 Struktur der Dissertation

Zur Beantwortung der Forschungsziele gliedert sich diese Arbeit in drei Teile: Zunächst werden die theoretischen Grundlagen diskutiert, anschließend erfolgt ein deskriptiver Ergebnisteil und abschließend die Erarbeitung des Referenzmodells für föderierte Datendienste für industrielle Ökosysteme, der als präskriptiver Ergebnisteil verstanden werden kann. Letztlich werden diese drei Teile in acht Kapitel gegliedert.

In *Kapitel 1* wurde eine Einführung in die Forschungsarbeit dieser Dissertation gegeben. Zunächst wurde dazu die Motivation dargelegt, die das Forschungsinteresse geweckt hat. Daraus wurden die einzelnen Forschungsziele abgeleitet und das Forschungsdesign entwickelt. Im Folgenden wird nun die Struktur dieser Monographie präsentiert.

Kapitel 2 diskutiert die theoretischen Grundlagen dieser Arbeit. Dabei werden die systemtheoretischen Grundlagen der Organisation dargestellt, und mit der Definition der Organisation als Informationssystem abgeschlossen. Anschließend werden die Grundlagen des Daten- und Informationsmanagements von Organisationen dargelegt, was den Ausgang für den Aufbau einer überbetrieblichen Dateninfrastruktur begründet. Die Grundlagen von Dienstleistungen und Datendiensten stellen den nächsten wesentlichen

Bereich der theoretischen Grundlagen dar, wobei zuletzt das Wesen von Datendiensten beschrieben wird. Der letzte Teil der Grundlagen behandelt unterschiedliche Formen von betriebsübergreifenden Organisationen, die für gemeinsame Datennutzung in Frage kommen könnten.

Kapitel 3 umfasst den Stand der Technik, der den konzeptionellen Bezugsrahmen für das Referenzmodell beschreibt. Dabei wird zunächst, ausgehend von dem Ziel dieser Dissertation, der Suchraum der Erhebung festgelegt. Dieser umfasst vier Konzeptgruppen. Dabei werden allgemeine Konzepte, wie Ansätze für die Informationssystemgestaltung, und spezifischere Konzepte, wie die Implementierungen von Datenräumen, beschrieben und deren Beitrag für die Entwicklung des Referenzmodells abgeleitet. Zuletzt wird daraus eine Grobstruktur entwickelt, welche die Grundlage für die eigentliche Entwicklung des Referenzmodells darstellt.

Die ersten Forschungsergebnisse werden in *Kapitel 4* vorgestellt. Hier wird die Literatur der Ökosysteme systematisch aufbereitet, deren Charakteristika analysiert und mittels einer Clusteranalyse fünf Idealtypen identifiziert. Diese Typen weisen wesentliche Unterschiede hinsichtlich ihrer Organisationsstruktur und damit der Eignung für föderierte Datendienste auf.

Kapitel 5 beinhaltet den zweiten Hauptteil dieser Arbeit. Hier wird einer der zuvor identifizierten Typen näher betrachtet und dessen spezifische Eigenschaften für das Referenzmodell genutzt. Dabei werden zunächst die Anforderungen erhoben, die für eine solche Organisation notwendig sind. Ausgehend davon werden Gestaltungsprinzipien erarbeitet und damit ein Referenzmodell für föderierte Datendienste erstellt. Dazu wird eine Prozess- und Funktionssicht bereitgestellt und die einzelnen Dienste detailliert beschrieben. Den Abschluss des Kapitels bildet die Evaluation des Referenzmodells.

Kapitel 6 bildet den Abschluss der Dissertation. Dabei werden zunächst die Ergebnisse zusammengefasst und die Beiträge zu Wissenschaft und Praxis diskutiert. Abschließend werden die Limitationen der Dissertation dargelegt und ein daraus abgeleiteter Ausblick auf mögliche Forschungsansätze aufgezeigt.