

FRAUNHOFER-INSTITUT FÜR SOFTWARE- UND SYSTEMTECHNIK, ISST
FRAUNHOFER-INSTITUT FÜR ARBEITSWIRTSCHAFT UND ORGANISATION, IAO

IEDS PROJEKT SACHBERICHT TEIL I KURZBERICHT

Fraunhofer Gesellschaft

Konsortialführer und Zuwendungsempfänger: Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.	Förderkennzeichen: IEDS001
Vorhabenbezeichnung: IEDS – Incentives and Economics of Data Sharing	
Laufzeit des Vorhabens: 01.05.2021 – 31.04.2024	
Berichtszeitraum: 01.05.2021 – 31.04.2024	
Fälligkeitsdatum: 31. Oktober 2024	
Erstellungsdatum: 17.10.2024	

Ansprechperson Fraunhofer ISST:

Heinrich Pettenpohl

Fraunhofer-Institut für Software- und Systemtechnik

Speicherstraße 6, 44147 Dortmund

Heinrich.Pettenpohl@isst.fraunhofer.de

Tel.: +49 231 97677-321

Ansprechperson Fraunhofer IAO:

Dr. Holger Kett

Fraunhofer-Institut für Arbeitswirtschaft und Organisation

Nobelstraße 12, 70569 Stuttgart

Holger.Kett@iao.fraunhofer.de

Tel.: +49 711 970 2415

Kurzbericht

I. Ursprüngliche Aufgabenstellung sowie den wissenschaftlichen und technischen Stand, an den angeknüpft wurde

In der digitalen Wirtschaft sind Daten ein bedeutender immaterieller Vermögenswert mit hohem wirtschaftlichem Wert, ähnlich wie traditionelle Ressourcen. Unternehmen wie Google, Amazon und Facebook zeigen, dass Daten eine Schlüsselressource für Wertschöpfung, Innovation und Wettbewerbsvorteile sind. Sie ermöglichen Prozessoptimierung, neue Geschäftsmodelle und Wachstum der globalen Wirtschaft. Ziel des Projektes IEDS war es, im Verbund interdisziplinäre Forschung zur Ausgestaltung von Data Sharing im unternehmerischen Kontext voranzutreiben und Anreizsysteme für das Teilen von Daten abzuleiten sowie die Weiterentwicklung der Datenökonomie zu unterstützen. Dabei werden vor allem technische und ökonomische Fragen des Data Sharing analysiert, aber auch rechtliche und ausgewählte technische Aspekte einbezogen. In den Forschungsarbeiten wurde die Perspektive der Unternehmen – vor allem aus Industrie und industrienahen Dienstleistungen – in den Fokus gerückt, um die Zurückhaltung deutscher Unternehmen an der Datenökonomie zu erklären und daraus Ableitungen für die weitere Herangehensweise für die Förderung des Data Sharing treffen zu können.

II. Ablauf des Vorhabens

Die interdisziplinären Themen der Datenwirtschaft und des Data Sharing verdeutlichen die Notwendigkeit, technologische, ökonomische und rechtliche Aspekte gleichermaßen abzudecken. Dafür ist Expertise aus verschiedenen Forschungsinstitutionen erforderlich. Das IEDS-Projekt war ein Verbundprojekt von fünf Forschungseinrichtungen Leibniz-Zentrum für europäische Wirtschaftsforschung, Institut der deutschen Wirtschaft, Lehrstuhl für Industrielles Informationsmanagement der TU Dortmund, das Fraunhofer-Institut für Arbeitswirtschaft und Organisation sowie dem Konsortialführer Fraunhofer-Institut für Software- und Systemtechnik, und lief von Mai 2021 bis April 2024. Es war in zwei Module unterteilt: Modul A (acht Monate) legte die Basis für die Erforschung der Kerndimensionen Datenstrategie, Datenrecht, datengetriebene Geschäftsmodelle und Datenbewertung. Modul B (28 Monate) baute darauf auf und entwickelte ein ganzheitliches Integrationskonzept, das Unternehmen befähigt, durch Datenaustausch wertschöpfende Aktivitäten zu betreiben. Der Status quo der Datenbewirtschaftung wurde dabei jährlich evaluiert. Das Ergebnis sind Blaupausen mit Methoden und Leitfäden, die KMU und Großunternehmen bei der Teilnahme an der Datenökonomie unterstützen. Beide Module entwickeln zusammen Anreizmechanismen, die Unternehmen einen großen Mehrwert bieten.

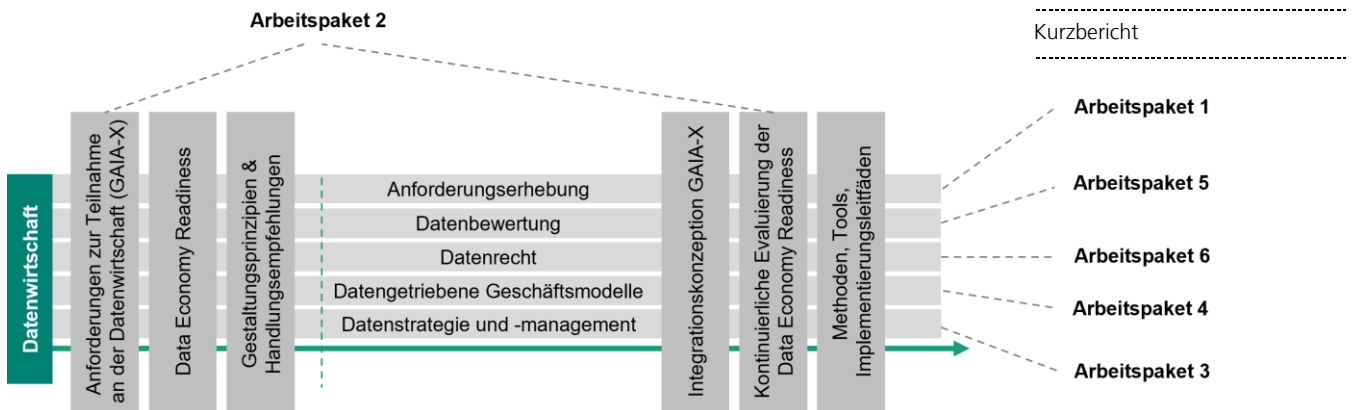


Abbildung 1 Arbeitspakete des IEDS-Projekts

III. Wesentlichen Ergebnisse sowie Zusammenarbeit mit anderen Forschungseinrichtungen

Durch die interdisziplinäre Forschung des IEDS-Projekts wurden Anreizsysteme für Data Sharing entwickelt, die Unternehmen dabei unterstützen, an der datengetriebenen Wertschöpfung teilzunehmen und die Datenökonomie in Deutschland voranzubringen. Jährliche Studien haben den Status Quo der Datenwirtschaft erfasst, Barrieren und Herausforderungen für Unternehmen aufgezeigt und Lösungsansätze entwickelt. Zudem wurden Geschäftsmodelle für datengetriebene Wertschöpfung und Unternehmensrollen in Datenökosystemen erarbeitet, sowie wissenschaftliche Werkzeuge zur Entwicklung innovativer Geschäftsmodelle auf Basis von Data Sharing.

Es wurden Prozessmodelle für die effiziente Datenbereitstellung und -suche entwickelt, um Unternehmen die Teilnahme am Data Sharing zu erleichtern. Diese Modelle helfen, Hürden in komplexen Governancestrukturen abzubauen und bessere Datenstrategien zu entwickeln. Erfolgreiche Use Cases bieten Unternehmen Orientierung bei der Strategieentwicklung.

Das Projekt untersuchte auch, wie Unternehmen Daten bewerten und welche Hemmnisse beim Teilen und Handeln bestehen, insbesondere im Hinblick auf Unsicherheiten und den kombinatorischen Wert von Daten. Für rechtliche Fragen wurden Werkzeuge wie ein Data Sharing Vertragsgenerator entwickelt.

Alle Ergebnisse des IEDS-Projekts wurden der Gesellschaft und Wirtschaft öffentlich zugänglich gemacht, unter anderem durch Vorträge und Veröffentlichungen in den jährlich erscheinenden IEDS-Whitepapers.



Abbildung 2 Jährliches IEDS-Whitepaper mit allen Projektergebnissen und Informationen

FRAUNHOFER-INSTITUT FÜR SOFTWARE- UND SYSTEMTECHNIK, ISST
FRAUNHOFER-INSTITUT FÜR ARBEITSWIRTSCHAFT UND ORGANISATION, IAO

IEDS PROJEKT SACHBERICHT TEIL II AUSFÜHRLICHER BERICHT

Fraunhofer Gesellschaft

Konsortialführer und Zuwendungsempfänger: Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.	Förderkennzeichen: IEDS001
Vorhabenbezeichnung: IEDS – Incentives and Economics of Data Sharing	
Laufzeit des Vorhabens: 01.05.2021 – 31.04.2024	
Berichtszeitraum: 01.05.2021 – 31.04.2024	
Fälligkeitsdatum: 31. Oktober 2024	
Erstellungsdatum: 17.10.2024	

Ansprechperson Fraunhofer ISST:

Heinrich Pettenpohl

Fraunhofer-Institut für Software- und Systemtechnik

Speicherstraße 6, 44147 Dortmund

Heinrich.Pettenpohl@isst.fraunhofer.de

Tel.: +49 231 97677-321

Ansprechperson Fraunhofer IAO:

Dr. Holger Kett

Fraunhofer-Institut für Arbeitswirtschaft und Organisation

Nobelstraße 12, 70569 Stuttgart

Holger.Kett@iao.fraunhofer.de

Tel.: +49 711 970 2415

Ausführlicher Bericht

I. Wichtigste Positionen des zahlenmäßigen Nachweises

Die wichtigste und größte Position des zahlenmäßigen Nachweises fällt auf Personalkosten, die zur Realisierung des Projekts verwendet wurden und notwendig waren. Mit Hilfe der Personalkosten konnten Wissenschaftler und Wissenschaftlerinnen der Fraunhofer Gesellschaft im Verbundprojekt mit anderen Wissenschaftlern die Forschungsarbeiten über den Projektzeitraum durchführen und somit die Projektergebnisse sichern. Die verwendeten Reisekosten für das Projekt wurden für Reisen zu wissenschaftlichen Konferenzen und Industrieveranstaltungen zur Verbreitung der Forschungsergebnisse verwendet. Darüber hinaus entstanden Kosten für Konsortialtreffen, die zur Koordination der Projektergebnisse und -aktivitäten durchgeführt wurden. Alle Konsortialtreffen, Projektreisen und Projektaktivitäten sind dem Projektträger in Form von Zwischenberichten mitgeteilt worden. Die erzielten Projektergebnisse werden im Verlauf dieses Sachberichts erläutert.

II. Notwendigkeit und Angemessenheit der geleisteten Projektarbeiten

Die Erforschung und Verbreitung von Anreizen für das interorganisationale Datenteilen leistet einen wichtigen Beitrag für die Datenökonomie und die Ziele der Datenstrategie der Bundesrepublik Deutschland als auch der Europäischen Kommission. Ziel des Projektes IEDS war es daher, im Verbund interdisziplinäre Forschung zur Ausgestaltung von Data Sharing im unternehmerischen Kontext voranzutreiben und Anreizsysteme für das Teilen von Daten abzuleiten sowie die Weiterentwicklung der Datenökonomie zu unterstützen. Dabei wurden technische und ökonomische Fragen des Data Sharing analysiert, aber auch rechtliche Aspekte einbezogen. Gleichzeitig sind im Projekt auch Barrieren und Hürden analysiert worden, die Unternehmen in Deutschland daran hindern an der Datenökonomie teilzunehmen oder Daten nicht mit anderen Unternehmen auszutauschen. Zu diesem Zweck sind zur Erhebung empirischer Daten deutschlandweite Umfragen durchgeführt worden, die einen Einblick in den Status quo der deutschen Datenökonomie ermöglichten, um Verbesserungspotentiale zu kristallisieren und Anreize zum Abbau von Barrieren zu entwickeln. Die jährlichen Erhebungen von Daten zum aktuellen Status der Datenökonomie in Deutschland brachten wichtige Erkenntnisse und zeigten, dass die erforschten Ergebnisse und entwickelten Werkzeuge im Projekt wichtige Anreize für vorhandene Hürden im Data Sharing für Unternehmen liefern. Auf Basis der empirischen Daten wurden technische, organisatorische, rechtliche und wirtschaftliche Hürden, die Unternehmen an der Teilnahme an der Datenökonomie oder am Data Sharing hindern, analysiert und Anreize zum Abbau dieser Hürden entwickelt. So konnten im Projekt verschiedene Ansätze für datengetriebene Geschäftsmodelle und ökonomische Wertströme analysiert und präsentiert werden, die wirtschaftliche Anreize für das interorganisationale Datenteilen aufzeigten. Ferner wurden organisatorische Prozessmodelle für die Datenbereitstellung sowie Datenverwaltung präsentiert, die es Unternehmen ermöglichen effizient und erfolgreich am Datenaustausch teilzunehmen und ihre Daten wertschöpfend zu nutzen. Darüber hinaus konnten Datenbewertungsmethoden aufgezeigt werden, anhand derer Unternehmen den Wert ihrer Daten bestimmen und somit souverän an der Datenökonomie teilnehmen können. Zur Darstellung der erfolgreichen Nutzung der Ergebnisse und Anreize des Data Sharing sind erfolgreiche Use Cases präsentiert worden, an denen Unternehmen sich orientieren können. Die Fördermittel des Projekts ermöglichten die interdisziplinäre Forschung im Verbund mit anderen Forschungs-institutionen mit verschiedenen Schwerpunkten. So

konnten die komplexen Anforderungen für den souveränen und sicheren Datenaustausch in ihrer Gesamtheit analysiert werden und Anreize für die unterschiedlichen Bereiche des wertschöpfenden Datenaustauschs bereitgestellt werden. Ohne das Verbundprojekt wäre eine solche Breite in der Analyse des Status quo der Datenökonomie und deren Anforderungen nicht möglich gewesen und die Ergebnisse wären eindimensional in ihrer Tiefe gewesen. Durch das Verbundprojekt konnten Ergebnisse erzeugt und verbreitet werden, die der Datenökonomie in Deutschland Voranschub leisten und Hürden und Barrieren für Unternehmen in Deutschland zur Teilnahme am Data Sharing abbauen.

III. Verwertbarkeit der Ergebnisse im Sinne des fortgeschriebenen Verwertungsplans

Im Projekt ist die wissenschaftliche und wirtschaftliche Verwertbarkeit sichergestellt worden. Auf wissenschaftlicher Seite konnten diverse Veröffentlichungen in hochrangigen wissenschaftlichen Konferenzen und wissenschaftlichen Publikationsreihen getätigt werden. Die Publikationen bieten dabei unterschiedliche Forschungsartefakte zu diversen Facetten des Data Sharing. So konnten eine Taxonomie zu Anreizen des industriellen Datenaustausches, Prozessmodelle für die effiziente Datenbereitstellung, Methoden zur Verbesserung der Datensuche, Methoden zur ökonomischen Bewertung von Daten und Wertschöpfungsmodelle des Data Sharing in wissenschaftlichen Outlets präsentiert und veröffentlicht werden. Die Veröffentlichungen in wissenschaftlichen Outlets stellten die qualitative Hochwertigkeit der Forschungsarbeiten sicher und ermöglichten es die Ergebnisse des Verbundprojekts einer breiten Masse an Forschern und Industrieteilnehmern zur Verfügung zu stellen. Die wissenschaftlichen Projektergebnisse erweitern den Literaturfundus des Forschungsfeldes Data Sharing signifikant, da bisher nur eine geringe Anzahl an Forschungen zu dem Thema vorhanden war. Die wissenschaftlichen Ergebnisse sind bereits während des Verbundprojekts von anderen Wissenschaftlern zitiert und genutzt worden, sodass eine wissenschaftliche Verwertbarkeit der Projektergebnisse gegeben ist. Zur Sicherstellung der wirtschaftlichen Verwertbarkeit sind die Ergebnisse und Forschungsmethoden im Verbundprojekt auf Basis von Unternehmen und ökonomischen Wertschöpfungsszenarien durchgeführt worden. Die Analyse des Status quo der deutschen Datenökonomie basierte auf empirischen Daten der Industrie und die Analyse der Ergebnisse dienen Unternehmen als Benchmark zur Einordnung der eigenen Fähigkeiten und Barrieren im interorganisationalem Datenaustausch. Die weiteren im Projekt erarbeiteten Ergebnisse stellen ökonomische Anreize zur Teilnahme an Data Sharing dar und zeigen auf, wie Unternehmen effizient und erfolgreich an der Datenökonomie teilnehmen können und ihre Daten für die Wertschöpfung bemessen können. Des Weiteren zeigen die Ergebnisse Anwendungsbeispiele und Erfolgsgeschichten des Data Sharing aus der Industrie auf, an denen sich Unternehmen richten können und neue datengetriebene Geschäftsmodelle und Wertschöpfungsmöglichkeiten entwickeln können. Die Ergebnisse wurden in verschiedenen Foren wie beispielsweise der Plattform Industrie 4.0 vorgetragen und über diverse Kanäle online der Öffentlichkeit zugänglich gemacht.

IV. Bekanntgewordener Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen

Es sind keine F&E-Ergebnisse Dritter bekannt geworden, die für die Durchführung des Projekts und deren Vorhaben beeinflussten.

V. Erfolgte und geplante Veröffentlichungen der Ergebnisse nach Nr.5 NKBF

Die Forschungsarbeiten der Fraunhofer Gesellschaft im Projekt fanden auf Basis der Arbeitspakete Data Economy Readiness, Datenstrategie und -management, datengetriebene Geschäftsmodelle, Datenbewertung sowie Projektmanagement statt.

Anreizsysteme des Data Sharing in industriellen Datenökosystemen

Die Arbeiten im Arbeitspaket **Data Economy Readiness** fokussierten sich auf die Entwicklung von Anreizen des Data Sharing in industriellen Ökosystemen, aus denen ein wissenschaftliches Paper entstand. Das Paper "**Motives and Incentives for Data Sharing in Industrial Data Ecosystems: An Explorative Single Case Study**" von Gelhaar et al. untersucht die Beweggründe und Anreize für das Teilen von Daten in industriellen Datenökosystemen am Beispiel von Catena-X, einem Datenökosystem der Automobilindustrie. Die Autoren stellen fest, dass viele Unternehmen zögern, Daten zu teilen, da der Nutzen für sie oft nicht klar ist. Um diese Forschungslücke zu schließen, analysieren sie die Motive und Anreize, die Datenanbieter dazu bewegen könnten, ihre Daten zu teilen.

Im Rahmen einer Fallstudie identifizieren die Autoren sieben verschiedene Motive und acht Anreize für das Teilen von Daten (siehe Abbildung 1). Die Motive lassen sich in wirtschaftliche, rechtliche und soziale Kategorien unterteilen. Zu den wirtschaftlichen Motiven gehören Kosteneinsparungen, Transparenz, Differenzierung, strategische Partnerschaften und direkte Vergütung. Rechtliche Motive umfassen die Einhaltung gesetzlicher Vorgaben, während soziale Motive sich auf die steigenden Anforderungen der Kunden an Nachhaltigkeit beziehen. Ein zentrales Ergebnis ist, dass der Mehrwert für die Datenanbieter hauptsächlich aus spezifischen Anwendungsfällen innerhalb des Ökosystems resultiert, anstatt aus der bloßen Teilnahme am Ökosystem selbst. Zudem zeigt sich, dass in Catena-X eine Kultur des gegenseitigen Datenaustauschs herrscht, bei der von allen Teilnehmern erwartet wird, Daten zu teilen. Die Autoren betonen, dass Anreize wie der Zugang zu datenbasierten Diensten oder die Möglichkeit zur Differenzierung durch neue Geschäftsmöglichkeiten entscheidend sind, um Unternehmen zur Teilnahme zu motivieren.

Abschließend tragen die Ergebnisse zur besseren theoretischen und praktischen Verständigung von Datenökosystemen bei und bieten Anregungen für die Gestaltung von Anreizmechanismen in solchen Systemen. Gleichzeitig erkennen die Autoren die Einschränkungen ihrer Studie an, darunter die begrenzte Übertragbarkeit der Ergebnisse auf andere Kontexte und die noch in Entwicklung befindlichen Strukturen von Catena-X.

Category	Motive	Incentive
Economic	Cost savings	Data-based services that lead to the improvement of internal processes
	Transparency	Data on the origin and authenticity of components and materials
	Differentiation	Data for new data-based business models
		Competitive advantage through information exchange and learning effects
	Strategic partnerships	Measures to enable and foster cross-company data sharing
Direct compensation	Money payment for pre-processed data	
Legal	Compliance	Data-based services that help to comply with legal requirements
Social	Customer demands for sustainability	Data-based service which can calculate the carbon dioxide footprint

Abbildung 1 Anreize des Data Sharing in industriellen Datenökosystemen aus Gelhaar et al. 2021 „Towards a Taxonomy of Incentive Mechanisms for Data Sharing in Data Ecosystems“

Forschungsarbeiten im zweiten Arbeitspaket **Datenstrategie und -management** fokussierten sich auf die strategischen Aspekte des Data Sharing sowie auf die organisatorischen Aspekte der Datenbereitstellung und Datenverwaltung.

Agile Datenbereitstellung durch DataOps Prozessinnovationen

Zum einen ist in diesem Arbeitspaket ein Report veröffentlicht worden, der Unternehmen die effiziente Datenbereitstellung aufzeigt. Der Report mit dem Titel „DataOps for Data Sharing: Herausforderungen und Anforderungen für organisationsübergreifendes Teilen von Daten“ behandelt die Bedeutung von Daten als strategische Ressource in der modernen Wirtschaft und stellt dar, wie Unternehmen diese nutzen können, um Innovationen und Wettbewerbsvorteile zu erzielen. Hauptpunkt des Berichts ist die Einführung von DataOps, einer Methode des Datenmanagements, die auf agilen und schlanken Prinzipien basiert. DataOps zielt darauf ab, Datenpipelines effizienter zu gestalten, indem sie automatisiert und standardisiert werden (Abbildung 2). Dies ermöglicht eine höhere Datenqualität und eine schnellere Bereitstellung von Daten für verschiedene Stakeholder in einer Organisation. Ein zentrales Thema ist die Herausforderung vieler Unternehmen, Daten effizient zu teilen und zu verwalten. Durch veraltete Technologien und organisatorische Barrieren, wie Datensilos, wird der Zugang zu qualitativ hochwertigen Daten oft erschwert. Hier setzt DataOps an, indem es die Zusammenarbeit zwischen verschiedenen Abteilungen fördert und die Nutzung von Daten über Abteilungs- und Unternehmensgrenzen hinweg verbessert.

Der Bericht bietet einen Überblick über die Definition von DataOps, die zugrundeliegenden Prinzipien und Frameworks sowie die zentralen Herausforderungen und Vorteile dieser Methode. Zu den Vorteilen gehören eine verbesserte Agilität, schnellere Markteinführung, bessere Entscheidungen und reduzierte Kosten. Als Herausforderungen nennt der Bericht unter anderem die Implementierung neuer Prozesse und den Widerstand gegenüber Veränderungen in der Unternehmenskultur. Ferner wird betont, dass DataOps ein wichtiger Baustein ist, um Unternehmen auf die Anforderungen der digitalen Transformation vorzubereiten und sie in die Lage zu versetzen, an den entstehenden Datenökosystemen teilzunehmen, insbesondere im europäischen Kontext mit Projekten wie GAIA-X.

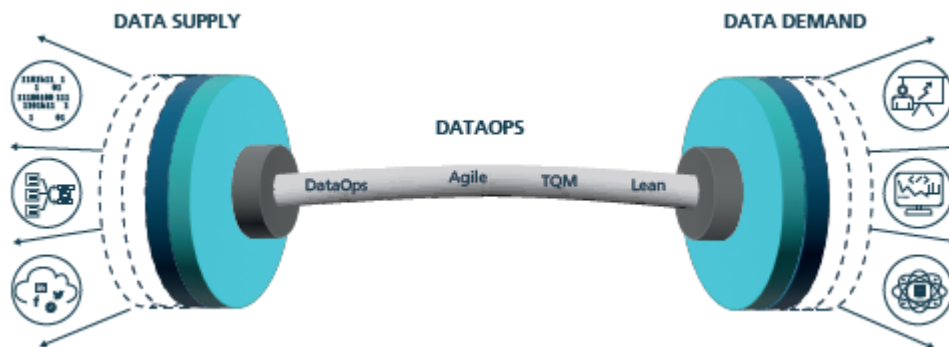


Abbildung 2 DataOps Kernelemente aus Gür 2021 „DataOps for Data Sharing: Herausforderungen und Anforderungen für organisationsübergreifendes Teilen von Daten“

Der Bericht hebt hervor, dass organisatorische Aspekte eine entscheidende Rolle bei der erfolgreichen Implementierung von DataOps und der Datenbereitstellung spielen. Dabei steht insbesondere die Komplexität der Unternehmensstrukturen sowie die Zusammenarbeit zwischen verschiedenen Abteilungen im Fokus.

Ein zentrales Problem in vielen Unternehmen sind Datensilos, also isolierte Datenbestände in einzelnen Abteilungen, die den freien Fluss von Daten innerhalb der Organisation behindern. Dies erschweren eine effiziente Bereitstellung und Nutzung von Daten, was Innovationen und datenbasierte Entscheidungen einschränken können. DataOps zielt darauf ab, diese Barrieren durch eine verbesserte Zusammenarbeit zwischen verschiedenen Teams wie Dateningenieuren, Data Scientists, IT-Sicherheits-

und Governance-Abteilungen sowie Anwendungsentwicklern abzubauen. Diese Zusammenarbeit wird als Schlüssel zur Optimierung von Datenpipelines und zur Sicherstellung der Datenqualität gesehen.

Ein weiterer organisatorischer Aspekt betrifft die Einführung von formalen Prozessen und Governance-Strukturen, um die Datenbereitstellung zu steuern und abzusichern (Abbildung 3). Die Implementierung von DataOps erfordert klar definierte Rollen und Verantwortlichkeiten, um sicherzustellen, dass alle relevanten Akteure effektiv zusammenarbeiten. Dies beinhaltet auch die Festlegung von Zugriffsrechten und die Einhaltung von Datenschutzrichtlinien, insbesondere in komplexen, datengetriebenen Organisationen. Die Entwicklung einer kollaborativen Unternehmenskultur ist ebenfalls entscheidend. DataOps fördert agile und funktionsübergreifende Teams, die effizienter arbeiten und schneller auf Datenanforderungen reagieren können. In diesem Kontext wird die Zusammenarbeit zwischen Geschäftsanwendern, Datenexperten und Technikteams als wesentlich für den Erfolg angesehen.

Zusammengefasst betont der Bericht, dass organisatorische Veränderungen, wie die Schaffung von formalen Prozessen, klaren Verantwortlichkeiten und einer kooperativen Unternehmenskultur, entscheidend sind, um die Datenbereitstellung zu verbessern und DataOps erfolgreich zu implementieren.

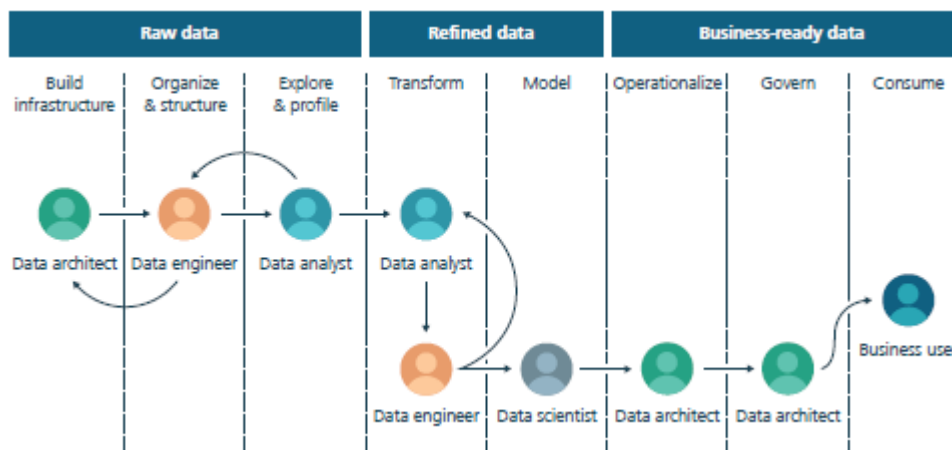


Abbildung 3 Organisationsstruktur für die agile Datenbereitstellung

Anforderungen für Unternehmen zur Verbesserung der dynamischen Fähigkeiten

Aus den konzeptionellen Arbeiten zur agilen Datenbereitstellung ist darüber hinaus ein wissenschaftliches Paper erstellt worden, welches auf der Conference on Business Informatics veröffentlicht wurde. Das Paper "Requirements for DataOps to foster Dynamic Capabilities in Organizations – A mixed methods approach" behandelt die Herausforderungen und Chancen, die Unternehmen durch die zunehmende Verfügbarkeit digitaler Daten erfahren. Die Autoren diskutieren, wie DataOps die Entwicklung dynamischer Fähigkeiten fördern kann, um Unternehmen in einem sich ständig verändernden digitalen Umfeld agiler zu machen. Ziel der Studie ist es, konzeptionelle Anforderungen für die Implementierung von DataOps zu identifizieren, um diese dynamischen Fähigkeiten zu stärken (Tabelle 1). Dazu stützt sich die Arbeit auf das Leavitt's Diamond Model und verwendet einen Mixed-Methods-Ansatz, der sowohl eine quantitative Multi-Vocal Literature Review als auch qualitative Experteninterviews kombiniert. Durch die qualitative und quantitative Analyse leiten die Autoren eine Reihe von Anforderungen an die Implementierung von DataOps ab, die auf Leavitt's Diamond Model basieren, das vier Schlüsselvariablen (Menschen, Aufgaben, Struktur, Technologie) untersucht. Die erfolgreiche Einführung von DataOps hängt stark von der organisatorischen Kultur, der Kompetenz der Mitarbeiter im Umgang mit Daten und

einer geeigneten technologischen Infrastruktur ab. Es wird betont, dass DataOps ein prozessorientierter und menschenzentrierter Ansatz ist, der auf cross-funktionaler Zusammenarbeit beruht und nicht rein technologisch orientiert ist.

Ferner stellt das Paper fest, dass die Implementierung von DataOps Unternehmen dabei unterstützen kann, schneller auf Marktveränderungen zu reagieren, fundiertere Entscheidungen zu treffen und neue datengetriebene Produkte und Dienstleistungen zu entwickeln.

Tabelle 1 Anforderungen für Unternehmen DataOps zu implementieren

<i>Dimension</i>	<i>Requirements</i>	<i>Interview #</i>	<i>Literature</i>
People	Roles to fulfill data expertise like data engineers, architects, scientists etc.	[1][2][3][4][5]	[7,8,10–13,52,55,57,58]
	Elaborated change management	[1][4][5]	[7–11,13,54]
	Training and stakeholder communication	[1][4]	[9,10,12,13,52,54]
	Organizational data culture, continuous improvement and commitment of leadership	[2][4][5]	[7,9,11,54,55,57]
Structure	Self-organizing teams	[3][4]	[8–13,52,54,57]
	Cross-functional and consumer collaboration	[1][5]	[7–13,52–55]
	Agile development sprints	[4]	[7–10,13,53,54,58]
Technology	Workflow orchestration tools	[2]	[9–13,52,53,55,57,58]
	Testing and automation tools	[1][2][5]	[8–10,12,13,52,54,55,57,58]
	Deployment and code repository tools	[2]	[9,11,12,52,53,55]
	Data capture and repository tools	[2][5]	[13,52,53]
	Collaboration and communication tools	[1][2]	[13,52,55]
	Governance tools	[2][5]	[55]
	DataOps platform	[1][2][5]	[9,11–13,52–55,57]
Task	Business requirements elevation	[1]	[10,13,55]
	ETL and data transformation	[1][2][5]	[9,13,52,55]
	Workflow orchestration	[2]	[9–13,52–55,57,58]
	Automated testing and monitoring	[1][2][5]	[7–10,12,13,52,54,55]
	Continuous deployment/integration	[2]	[7,10,12,13,52–55,58]
	Data visualization and consumption	[1][2][4][5]	[7,10,13,55]

Reifegradmodell für die strategischen Aspekte des Data Sharing

Darüber hinaus ist im Arbeitspaket Datenmanagement und -strategie ein Bericht über ein **Reifegradmodell für die strategischen Aspekte des Data Sharing** verfasst worden (Frings, Sandra / Gür, Inan, 2024, Data Sharing Reifegradmodell - Strategische Aspekte des Data Sharing). Der Schwerpunkt des Berichts liegt auf der strategischen Bedeutung des Datenaustauschs in der heutigen zunehmend digitalen Wirtschaft. Da Unternehmen immer häufiger branchenübergreifend zusammenarbeiten, um innovative Geschäftsmodelle und digitale Produkt-Service-Systeme zu entwickeln, wird der Bedarf an strategischen Ansätzen für den Datenaustausch immer wichtiger.

Um den Herausforderungen des Data Sharing zu begegnen, stellt der Bericht ein Data Sharing Reifegradmodell vor, das im Rahmen des IEDS-Projekts entwickelt wurde. Dieses Modell dient als strukturiertes Instrument, das Organisationen dabei unterstützt, ihre aktuellen Praktiken des Datenaustauschs zu bewerten und notwendige Verbesserungen zu ermitteln.

Das Reifegradmodell umfasst sechs Schlüsseldimensionen – Organisation, Prozesse, Daten, Technologie, Geschäftsmodelle und Governance –, die jeweils in spezifische Kategorien unterteilt sind, die analysiert werden, um die Bereitschaft und Fähigkeit einer Organisation zum Datenaustausch zu bewerten (siehe auch Abbildung 4):

- 1) Organisation (Strategie und Planung, Geschäftsausrichtung, Unternehmenskultur, Kompetenzentwicklung),
- 2) Prozesse (Prozessdefinition, Erfolgsmonitoring),
- 3) Wert (Chancenerkennung, Wertgewinnung),
- 4) Daten (Datenlebenszyklus, Datenmanagement und Daten-Stewardship, Datenqualität),
- 5) Technologie (Infrastruktur, Interoperabilität, Datensouveränität),
- 6) Governance (Richtlinien, Rechenschaftspflicht, Governance-Struktur)

Pro Kategorie sind Indikatoren definiert, die für die folgenden fünf Stufen erfüllt sein müssen:

- Auf der ersten Stufe (Data Sharing-Außenstehender) sind keine bzw. nur initiale Ansätze und Strukturen im Unternehmen für die inhaltlichen, strategischen Fähigkeiten vorhanden; Prozesse werden chaotisch, unbewusst, unsystematisch und unvorhersehbar durchgeführt.
- Ein reaktiver Ansatz, definierte, gesteuerte, wiederholbare und somit grundlegende Prozesse und Kompetenzen sind auf der zweiten Stufe (Data Sharing-Unerfahrener) erfüllt.
- Auf der dritten Stufe (Data Sharing-Praktiker) sind Ansätze und Prozesse stabil, formal, etabliert proaktiv geführt und Kompetenzen in Bezug auf Data Sharing fortgeschritten.
- Regelmäßige Überprüfung und Messung von Ansätzen und Prozessen sind der Fokus der vierten Stufe (Data Sharing-Spezialist).
- Auf der fünften Stufe (Data Sharing-Experte) werden Ansätze und Prozesse rund um die strategischen Fähigkeiten optimierend und nachhaltig geführt.

Das beschriebene Reifegradmodell kann auf verschiedene Weise zur Verbesserung der strategischen Ausrichtung eines Unternehmens hinsichtlich Data Sharing beitragen: Das Modell kann in Form eines Fragebogens als Werkzeug zur Erhebung der Ist-Situation und des Soll-Zustands genutzt werden. Hierzu werden die einzelnen Indikatoren zu einem Fragebogen zusammengestellt und mit einem Punktesystem hinterlegt. Es wird für jeden Indikator eine Stufe für Ist- und Soll-Zustand abgefragt. Hierfür sind 32 Antworten notwendig. Alternativ kann bei der Erhebung des Ist-Zustands für jeden der 16 Indikatoren die Übereinstimmung mit den fünf Stufen jeweils über eine 5er-Skala angegeben werden, um ein differenzierteres Bild zu erhalten. Der Soll-Zustand kann wie oben abgefragt werden. Hierzu sind 96 Antworten notwendig.

Weitere mögliche Erweiterungen sind die Ableitung von Handlungsempfehlungen aufgrund der inhaltlichen Differenzen zwischen Ist- und Soll-Zustand. Durch wiederholte Durchführung der beschriebenen Befragung können Trends erkannt und Fortschritte gemessen werden. Liegen viele durchgeführte Bewertungen vor, ist es dem Anbieter des Fragebogens möglich, zusätzlich ein Benchmarking anzubieten.

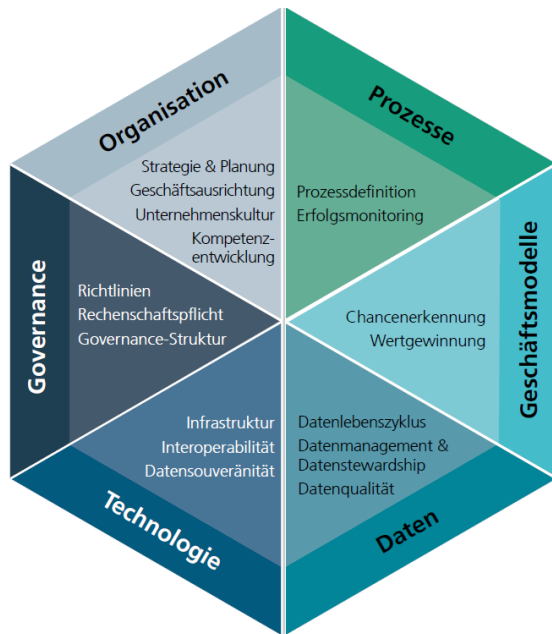


Abbildung 4 Kerndimensionen des Data Sharing Reifegradmodells

Anforderungsanalyse und Konzeption eines Matching-Verfahrens für die Datensuche

Ein weiteres Paper entstand zum Thema **Datensuche**. Eine wichtige Voraussetzung des Data Sharing ist die Zusammenführung von Datenanbieter und ihren Angeboten mit den Datennutzern und deren Bedarfen. Nur wenn ein Datennutzer von der Existenz eines potenziell nutzenbringenden Datenangebots weiß, kann er Schritte einleiten, um die Daten zu erwerben oder gegen eigene Angebote oder Leistungen zu tauschen. Auf der anderen Seite muss ein Datenanbieter sein Angebot bekannt machen, um Interessenten für seine Daten zu finden. Datenmarktplätze und Datenräume bieten einen technischen Rahmen, um den Austausch von Daten in die Wege zu leiten und zu organisieren. Das Fraunhofer IAO hatte sich im Rahmen des Projekts das Ziel gesetzt, Datenanbieter und Datenkonsumenten durch den Einsatz von agenten-basierten Systemen und KI-gestützten Matching-Verfahren dabei zu unterstützen, ihre Angebote und Bedarfe besser zusammenzuführen.

Um den Stand der Technik zu erheben und Ansatzpunkte für Verbesserungen zu identifizieren, wurden 17 Datenmarktplätze und Datenplattformen (Telekom Data Intelligence Hub, Advaneo Data Marketplace, MDM-Mobilitätsdaten-Marktplatz, Convex The Mobility Data Exchange, Caruso Dataplace, Dawex Space Data Marketplace, MyDataEconomy, Agrimetrics, otonomo data service Platform, GDELT Global Database of Events Language and Tone, Mobilty Data Space, Here, Ocean Protocol, IOTA Data Marketplace, Streamr Data Marketplace, Terbine Marketplace und Kaggle) hinsichtlich der verwendeten Such- und Navigationskonzepte und von Art und Umfang der Beschreibung der Datenangebote untersucht.

Eine Beobachtung ist, dass die beschreibenden Metadaten der angebotenen Datensätze häufig nicht sehr umfangreich sind. Die Angebote sind in der Regel durch eine Volltext-Beschreibung, den Urheber, das Datenformat, die Lizenz zur Verwendung der Daten, die Zeitpunkte der Bereitstellung und Aktualisierung, den Preis und häufig auch durch Beispieldaten beschrieben. Die Inhalte können in der Regel mit Hilfe einer Volltextsuche durchsucht oder mit Hilfe einer Navigation nach Kategorien oder Schlüsselwörtern navigiert werden. Die Suchergebnisse können in der Regel nach den Kriterien Urheber, Datenformat, Lizenz und ähnlichen Angaben gefiltert werden.

Des Weiteren wurden eine Literaturrecherche und zwei Experteninterviews mit der CAS Software AG und der Advaneo GmbH durchgeführt, um die Ergebnisse zu vertiefen. Die Anforderungserhebung lieferte die folgenden Erkenntnisse:

- Die Suche nach Daten ist ein wichtiger Bestandteil der technischen Unterstützung des Data Sharing und kommt in Datenmarktplätzen, Datenkatalogen und in Datenräumen zum Einsatz. Die einfache Auffindbarkeit von Daten stellt einen wichtigen Anreizmechanismus für das Data Sharing dar.
- Im Gegensatz zu anderen Produkten, die über Internetplattformen vertrieben werden, sind bei Datenprodukten Informationen über die Entstehungs- und Verarbeitungsgeschichte der Daten sowie deren Zeit- und Ortsbezug für die Auffindbarkeit und Bewertung von besonderer Wichtigkeit.
- Mit dem Einzug der Künstlichen Intelligenz (KI) und deren Bedarf nach großen und domänenspezifischen Datensätzen ist in den letzten Jahren ein Trend zu immer kleinteiligeren Datenangeboten für das Trainieren und Evaluieren von KI-Anwendungen zu beobachten (z. B. Sensordaten von Maschinen, Wetter, u.v.m.)

Auf Basis der Analyse der Metadaten wurde zwei relevante Anwendungsfälle für das Projekt definiert: Suche nach Sensordaten und Suche nach Wissenschaftsdaten. Sensordaten sind ein Kernbestandteil des Internet-of-Things und von Industrie 4.0. Zur Interpretation und Bewertung von Sensordaten sind Metadaten zur Erhebungsmethodik, der Maßeinheit der Datenreihe und den bereits erfolgten Vorverarbeitungsschritten von besonderer Bedeutung. Ein auf Sensordaten basierendes Anwendungsszenario aus dem Agrarbereich wurde im zweiten IEDS-Whitepaper¹ beschrieben. Wissenschaftsdaten zeichnen sich durch ihre Verknüpfung zu wissenschaftlichen Arbeiten aus.

Je spezifischer Datenangebote sind, desto stärker reduzieren sich potenzielle Nutzergruppen. Je spezifischer eine Datennachfrage ist, desto schwieriger sind die dazu passenden Datenangebote zu finden. Dabei werden aktuell Datenprodukte erst sukzessive aufgebaut, was noch viel Aufwand und vor allem Zeit benötigt. Klassische Such- bzw. Matching-Methoden von Angebot und Nachfrage sind zu statisch und träge, um die Datenprodukte und deren Beschreibungen auf die sich stets verändernden und neuen Anfragen anzupassen. Hierfür wurde der Lösungsansatz intelligenter Software-Agenten zur Repräsentation von Datenprodukten geschaffen, die ähnlich Verkäufern auf einem Markt Kundenanfragen auswerten und ihre Kundenansprache auf die Nachfrage anpassen. Dabei wird jedes Datenprodukt durch einen eigenständigen Agenten repräsentiert. Der Agent beobachtet die Suchanfragen der Kunden und wird vom Marktplatz benachrichtigt, wenn der Kunde das Datenprodukt des Agenten kauft oder in die engere Auswahl nimmt. Die Agenten nutzen die so gewonnenen Informationen über die Präferenzen der Kunden, um ihr Produkt auf dem Marktplatz besser darzustellen und passen sich so dynamisch an die sich ändernden Bedürfnisse der Datensuchenden an (Abbildung 6). Die Agenten stehen hierbei untereinander im Wettbewerb. Es sind jedoch auch Kooperationsszenarien denkbar, in denen Agenten Informationen untereinander teilen oder sich zu Produktkombinationen zusammenschließen.

¹ Anreizsysteme und Ökonomie des Data Sharing - Status Quo der deutschen Datenwirtschaft und Anwendung von unternehmensübergreifendem Datenaustausch, 2023, https://ieds-projekt.de/wp-content/uploads/2023/04/IEDS-Whitepaper_2023_finale-Version.pdf

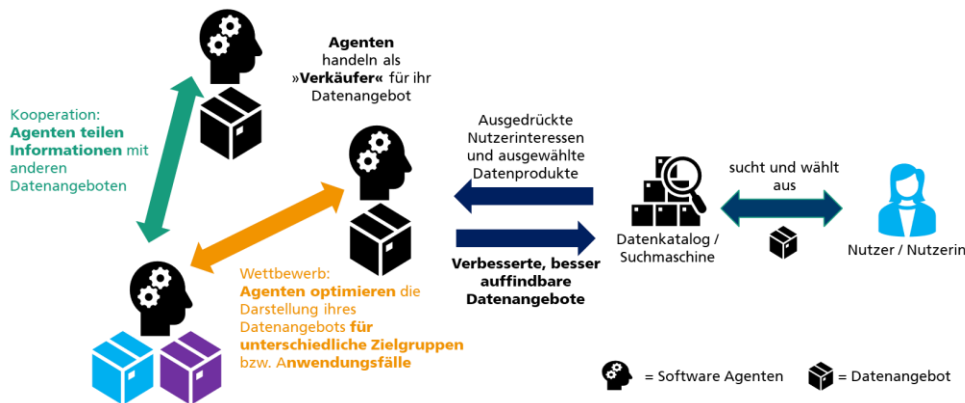


Abbildung 5 Forschungsansatz: Jedes Datenangebot wird von einem Softwareagenten repräsentiert, der als „Verkäufer“ sein Produkt bestmöglich anbietet

Der innovative Kern dieser Idee ist, dass die Optimierung der Suche auf der Angebotsseite stattfindet, während diese Optimierungen klassischerweise innerhalb der Suchmaschine umgesetzt werden (z. B. durch Anreicherung der Suchanfrage um weitere Suchbegriffe oder Synonyme). Beide Herangehensweisen ergänzen sich und können zusammen Anwendung finden.

Wie kann nun ein Agent erfahren, ob eine von ihm ausgewählte Repräsentation seines Datenprodukts in der Suchmaschine erfolgreich ist und wie kann er Rückschlüsse auf mögliche Verbesserungen ziehen? Im verfolgten Ansatz schickt jeder Agent verschiedene Repräsentationen seines Produkts ins Rennen um die Gunst der Suchenden, indem er verschiedene Varianten der Beschreibung seines Produkts in den Suchindex der Suchmaschine publiziert. Um zu verhindern, dass dasselbe Produkt mehrfach in der Trefferliste erscheint, die den Suchenden als Ergebnis einer Suchanfrage angezeigt wird, wird nur diejenige Variante des Produkts angezeigt, die von der Suchmaschine als am passendsten eingestuft wird.

Auf diese Weise erhält der Agent Informationen über den Erfolg der einzelnen Varianten und kann Rückschlüsse ziehen, welche Varianten am besten funktionieren. Auf der Basis dieser Erkenntnisse erzeugt der Agent in periodischen Abständen neue (potenziell bessere) Varianten und verwirft Varianten, die sich nicht bewährt haben. Dieses Verfahren ermöglicht es den Agenten, sich bzw. die Repräsentation seines Datenprodukts in der Suchmaschine an das sich mit der Zeit ändernde Suchverhalten der Nutzer anzupassen.

Der Erfolg einer Produktrepräsentation bedeutet in diesem Kontext, dass das Produkt 1) in einer Suchanfrage gefunden wird, 2) dass es vom Nutzer im Rahmen dieser Suchanfrage als relevant bewertet wird und 3) dass es in diesem Fall möglichst weit oben in der Liste der Suchergebnisse erscheint. Um eine objektive und quantitative Bewertung der Qualität des beschriebenen Ansatzes zu erhalten, wurden folgende Schritte durchgeführt:

- 1) Eine größere Anzahl von vorgegebenen Suchanfragen wurde ausgeführt.
- 2) Für jede Anfrage wurde die vom Suchverfahren zurückgelieferte Trefferliste, mit der für diese Suchanfrage erwarteten, „richtigen“ Trefferliste verglichen. Je größer die Übereinstimmung der tatsächlichen Trefferliste mit der erwarteten, desto höher ist die Qualität des Suchverfahrens für diese eine Suchanfrage.
- 3) Aus der Übereinstimmung der Soll- und Ist-Trefferlisten wurden Metriken zur quantitativen Bestimmung der Qualität des Suchverfahrens berechnet. Hierzu wurde in erster Linie die Metrik „Normalized Discounted Cumulative Gain“ (NDCG) benutzt, da sie in der Lage ist, das Verhalten von Suchenden realistisch abzubilden.

Zur Durchführung wurde eine Testumgebung in der Programmiersprache Python erstellt, die es erlaubt, verschiedene Suchverfahren auszuführen und objektiv zu bewerten. Um diese Bewertung für den agenten-basierte Ansatz durchzuführen, wird eine große Anzahl von Suchanfragen und den zugehörigen Relevanzeinschätzungen der jeweiligen Suchergebnisse benötigt. Da die Bewertung automatisiert und wiederholbar sein soll, wurde die Benutzerinteraktion durch Verwendung der Testdatensammlung NfCorpus durchgeführt. Diese Sammlung aus dem Bereich medizinischen Forschungsdaten enthält einen Satz von 3633 Dokumenten, 3236 Suchanfragen und 134294 Relevanzeinschätzungen, die zur Simulation der Benutzerinteraktionen mit der Suchmaschine herangezogen wurden.

Die Testumgebung führt folgende Schritte durch:

- 1) Stellen einer Suchabfrage aus der Testdatensammlung an die Suchmaschine
- 2) Vergleich der Suchergebnisse mit den erwarteten Suchergebnissen aus der Testdatensammlung. Das Ergebnis des Vergleichs wird zum einen zur Berechnung der Bewertungsmetriken benutzt. Zum anderen wird jedem in der Trefferliste enthaltenen Agenten mitgeteilt, ob er für die Suchanfrage relevant war
- 3) Die Agenten sammeln diese Rückmeldungen und optimieren ihre Repräsentation in der Suchmaschine.
- 4) Die Schritte 1) bis 3) werden für alle Suchabfrage aus der Testdatensammlung durchgeführt. Am Ende enthält der Suchindex die zu diesem Zeitpunkt erfolgreichsten Repräsentationen der zu suchende Dokumente.

Ansatz 1: Optimierung der Datenbeschreibung im Rahmen der textbasierten Suche

Ein häufiges Problem bei der Suche im Allgemeinen aber besonders bei der Suche nach Daten ist der Unterschied zwischen der Art der Datenbeschreibung und der Sprache der Suchenden. Daten werden häufig mit einer technischen von der Domäne abhängenden Sprache beschrieben während Suchanfragen häufig umgangssprachlich formuliert sind und einen anderen Wortschatz verwenden. Dies ist auch in der verwendeten Testdatensammlung NfCorpus der Fall, da es sich bei den zu suchenden Dokumenten um Abstracts medizinischer Veröffentlichungen handelt und die Suchanfragen aus einem Gesundheitsportal entnommen wurden.

Der erste im Projekt verfolgte agenten-basiert Ansatz hatte das Ziel, diese Kluft zwischen den Sprachniveaus zu verringern. Dabei sind die Dokumente in der Suchmaschine in Textform repräsentiert. Die Agenten werden wie oben beschrieben über Suchanfragen informiert, für die sie als relevant eingestuft wurden. Sie versuchen nun mit verschiedenen Strategien, die relevantesten Begriffe aus den Suchanfragen zu identifizieren und ihrer Beschreibung hinzuzufügen. Als erfolgreichste Strategie erwies sich, dass die Agenten über die Zeit alle Begriffe der relevanten Suchanfragen sammeln und in bestimmten Zeitabständen ein Topic-Modelling auf diesen Begriffen durchführen. Beim Topic-Modelling werden die Begriffe in inhaltlich zusammengehörige Cluster eingeteilt. Die Begriffe jedes Clusters wurden benutzt, um jeweils eine neue Repräsentationsvariante des Dokuments zu erzeugen.

Gegenüber dem gängigen Standardverfahren BM25, das einen NDCG@10-Wert von 0,24 lieferte, konnte das beschriebene agentenbasierte Verfahren einen deutlich

besseren NDCG@10-Wert 0.29 erreichen. Die Ergebnisse sind im Paper „Agent-based Document Expansion for Information Retrieval based on Topic Modeling of local Information“ ausführlich dargestellt, das auf der Konferenz ISCM I 2022 vorgestellt und veröffentlicht wurde. Die Präsentation wurde als beste der Session ausgezeichnet.

Ansatz 2: Optimierung der Repräsentation der Datenbeschreibung im Rahmen der semantischen Suche

Im zweiten agentenbasierten Ansatz werden die Dokumente nicht als Text in der Datenbank repräsentiert, sondern als Vektoren in einem hochdimensionalen Vektorraum. Diese sogenannten semantischen Embeddings werden von speziell trainierten Neuronalen Netzen erzeugt und haben die Eigenschaft, dass benachbarte Punkte im Vektorraum semantisch ähnliche Bedeutung haben. Eine weitere Eigenschaft dieser Embeddings ist, dass man mit ihnen wie mit Vektoren rechnen kann. Bildet man z.B. den Mittelwert zwischen mehreren Embedding-Vektoren, erhält man einen Vektor, der eine Mischung aus den semantischen Bedeutungen der einzelnen Vektoren darstellt. Eine Suche läuft so ab, dass die Suchanfrage ebenfalls in einen Vektor umgewandelt und mit den Vektorrepräsentationen der Dokumente verglichen wird. Je näher ein Dokument an der Suchanfrage liegt, desto relevanter stuft die Suchmaschine das Dokument ein.

Die Agenten nutzen den Umstand aus, dass verschiedene Embedding-Vektoren sich zu neuen semantischen Vektoren kombinieren lassen. Der Titel des Dokuments, die Sätze des Dokumenteninhalts und die Suchanfragen, für die der Agent als relevant eingestuft wurde, werden jeweils als Embedding-Vektoren kodiert. Die Agenten können acht verschiedene Kombinationsstrategien (R1-R8) nutzen, um diese Vektoren zu kombinieren:

- R1: Nur der Titel wird verwendet
- R2: Der Mittelwert aus den Inhaltsvektoren wird verwendet
- R3: Für jeden Inhaltsvektor wird eine eigene Variante des Dokuments erzeugt
- R4: Der Titel und der Mittelwert der Inhaltsvektoren wird in einer Linearkombination zusammengeführt
- R5: Aus der Linearkombination aus Titel jedem einzelnen Inhaltsvektor werden jeweils eine Variante des Dokuments erzeugt
- R6: Eine neue Variante wird aus der Linearkombination aus Titel, Mittelwert der Inhaltsvektoren und Mittelwert der relevanten Suchanfragen erzeugt
- R7: Jeweils eine neue Variante wird aus der Linearkombination aus Titel, Mittelwert der Inhaltsvektoren und jeder einzelnen relevanten Suchanfrage erzeugt
- R8: Die relevanten Suchanfragen werden geclustert und es wird jeweils eine neue Variante aus der Linearkombination aus Titel, Mittelwert der Inhaltsvektoren und jedes einzelnen gemittelten Clusters erzeugt.

Basierend auf diesen acht Strategien wurden vier agenten-basierte Experimente durchgeführt, bei denen den Agenten unterschiedliche Strategien zur Verfügung standen: A1 (benutzt R1-R5), A2 (benutzt R1-R4 und R6), A3 (benutzt R1-R4 und R7) und A4 (benutzt R1-R4 und R8). Die Ergebnisse sind in Abbildung 6 dargestellt.

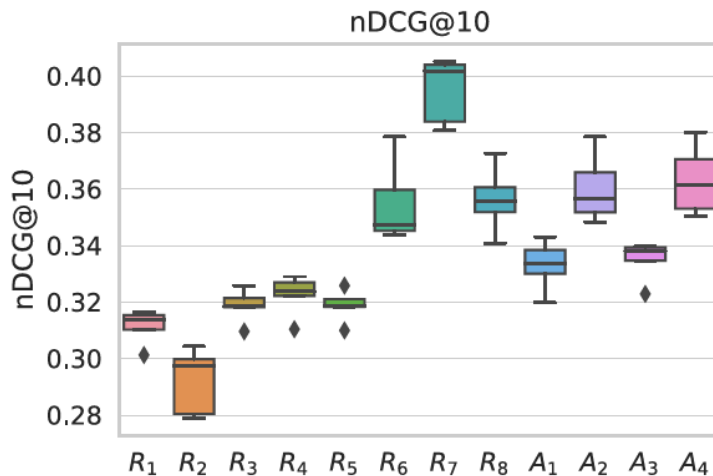


Abbildung 6 Ergebnisse der NDCG-Metrik für die einzelnen Strategien R1-R8 und die agenten-basierten Versuche A1-A4.

Der Agentenansatz A4, der auf dem Clustering der relevanten Suchanfragen aufbaut, erwies sich als der erfolgreichste. Er erreichte mit 0,36 einen deutlich höheren NDCG-Wert als die in Ansatz 1 untersuchten textbasierten Verfahren. Im Paper „Documents as Intelligent Agents: An Approach to Optimize Document Representations in Semantic Search“, das auf der WEBIST 2023-Konferenz publiziert wurde, werden die Ergebnisse eingehend dargestellt. Das Paper wurde mit dem **Best Paper Award der WEBIST 2023** ausgezeichnet.

Zusammenfassend wurde im Projekt IEDS im Bereich der Datensuche mit dem beschriebenen agenten-basierten Ansatz ein neues Verfahren entwickelt, um die Repräsentation von Dokumenten oder Datenprodukten in einem Suchverfahren zu optimieren und die Datenprodukte somit besser auffindbar zu machen. Die Agenten können dabei potenziell verteilt arbeiten und mit zusammen mit dem Datenprodukt beim Datenanbieter verbleiben. Der Ansatz ist daher gut für Data Spaces und Architekturen wie Gaia X geeignet, die diesen Ansatz aus Gründen der Datensouveränität verfolgen. Die Datensuche an sich ist ein wichtiges Bindeglied zwischen Datenanbietern und Datennutzern in einer Datenökonomie. Sie ist sowohl in der Interaktion zwischen verschiedenen Unternehmen über Datenmarktplätze und Datenkataloge wichtig als auch im Rahmen des Unternehmensinternen Datenmanagements. Insofern leisten die beschriebenen Arbeiten zur Verbesserung der Datensuche einen wichtigen Beitrag zur Förderung des Data Sharing.

Das dritte Arbeitspaket fokussierte sich auf **datengetriebene Geschäftsmodelle**.

Modellierung von Wertschöpfungskonzepten in Datenökosystemen

Dort sind Arbeiten zu verschiedenen wissenschaftlichen Papern getätigt worden. Das Paper "Hunting the Treasure: Modeling Data Ecosystem Value Co-Creation" von Azkan et al. 2022 untersucht das aufkommende Konzept von Datenökosystemen, insbesondere in Bezug auf ihre Rolle bei der interorganisationalen Wertschöpfung durch Datenaustausch. Die Forschung befasst sich mit der Lücke in empirischen Studien darüber, welchen Wert verschiedene Akteure aus der Teilnahme an Datenökosystemen ziehen. Zu diesem Zweck analysieren die Autoren 64 Anwendungsfälle aus der Gaia-X-Initiative, die darauf abzielt, sichere Datenökosysteme in Europa zu entwickeln. Das Paper identifiziert acht Schlüsselrollen innerhalb dieser Ökosysteme, darunter Datenanbieter, Dienstleister, Cloud-Plattform-Anbieter, Dateninfrastruktur-Anbieter, App-Store-Anbieter, Ökosystem-Orchestrator, Datentreuhänder und Datenmarktplatz-Betreiber.

Die Studie stellt außerdem eine Methode zur Visualisierung der Wertschöpfung in Datenökosystemen vor, indem die e3-value-Modellierungssprache verwendet wird. Diese Methode wird anhand des Agri-Gaia-Anwendungsfalls veranschaulicht, der sich auf die Verbesserung landwirtschaftlicher Prozesse durch künstliche Intelligenz konzentriert. Die Autoren validieren ihr Modell, indem sie Projektbeteiligte von Agri-Gaia konsultieren. Das Paper leistet einen Beitrag zum Verständnis von Datenökosystemen, indem es Rollen kartiert und zeigt, wie diese Rollen interagieren, um gemeinsam Wert zu schaffen. Es betont auch die Bedeutung der Perspektive von Dienstleistungsökosystemen, die davon ausgeht, dass Wert durch die Integration von Ressourcen durch verschiedene Akteure gemeinsam geschaffen wird. Die Studie schlägt vor, dass Modellierungswerkzeuge wie e3-value helfen können, die Komplexität von Datenökosystemen zu bewältigen, erkennt jedoch auch Einschränkungen an, wie beispielsweise den Fokus auf Wert im Austausch (value-in-exchange) anstelle von Wert in der Nutzung (value-in-use).

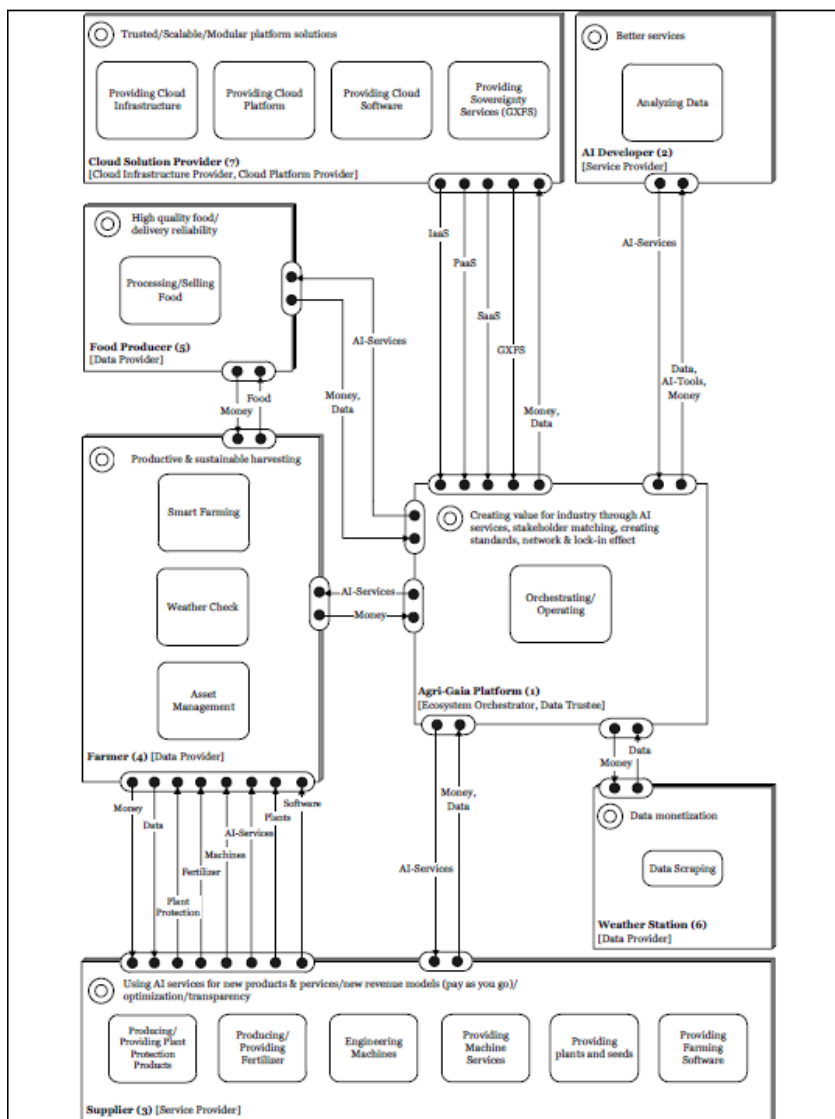


Abbildung 7 e3-Value Modell des Agri-Gaia Ökosystems aus Azkan et al. 2022 „Hunting the Treasure: Modeling Data Ecosystem Value Co-Creation“

Die Autoren schließen mit der Identifizierung potenzieller Bereiche für zukünftige Forschung, darunter die Erweiterung der Studie auf andere Industrien und Regionen, die Untersuchung der Dynamik von Rollen innerhalb von Datenökosystemen und die Entwicklung zugänglicherer Modellierungsmethoden.

Data Sharing Use-Case-Report

Im zweiten Arbeitspaket ist ein **Use Case Report** (Gür, Inan / Gieß, Anna / Jussen, Ilka / Schweihoff, Julia, 2023) erstellt worden. Der Report stellt mehrere Fallbeispiele vor, die zeigen, wie erfolgreiches Data Sharing in verschiedenen Branchen und Technologien funktioniert und welche Vorteile es bietet. Diese Use Cases veranschaulichen, wie datengetriebene Geschäftsmodelle ermöglicht und Innovationspotenziale durch die gemeinsame Nutzung von Daten erschlossen werden können. Der Bericht hebt auch die Herausforderungen hervor, denen Unternehmen beim Data Sharing begegnen, und zeigt auf, wie diese durch erfolgreiche Anwendungsfälle überwunden werden können.

Ein Beispiel im Report ist Dawex im Bereich Sport, wo Daten zwischen Einzel- und Großhändlern, Vereinen, Sponsoren und Veranstaltungsplanern geteilt werden. Dies ermöglicht die Optimierung der Fan- und Verbrauchererfahrungen, verbessert Vertragsabschlüsse und erleichtert die Planung von Großveranstaltungen.

Ein weiteres Beispiel ist Agri-Gaia im landwirtschaftlichen Sektor. Hier tauschen Landwirte, Landmaschinenhersteller und KI-Entwickler Daten aus, um präzise KI-Dienste zu entwickeln, die eine nachhaltigere Bewirtschaftung, gezielte Düngung und vorausschauende Handlungsempfehlungen ermöglichen.

Im Mobilitätsbereich wird der Mobility Data Space vorgestellt, der Fahrzeughersteller, Betreiber öffentlicher Verkehrsmittel und Forschungseinrichtungen zusammenbringt. Hier werden Sensordaten genutzt, um Verkehrssicherheit zu erhöhen, Echtzeitinformationen bereitzustellen und neue digitale Services im Mobilitätssektor zu entwickeln.

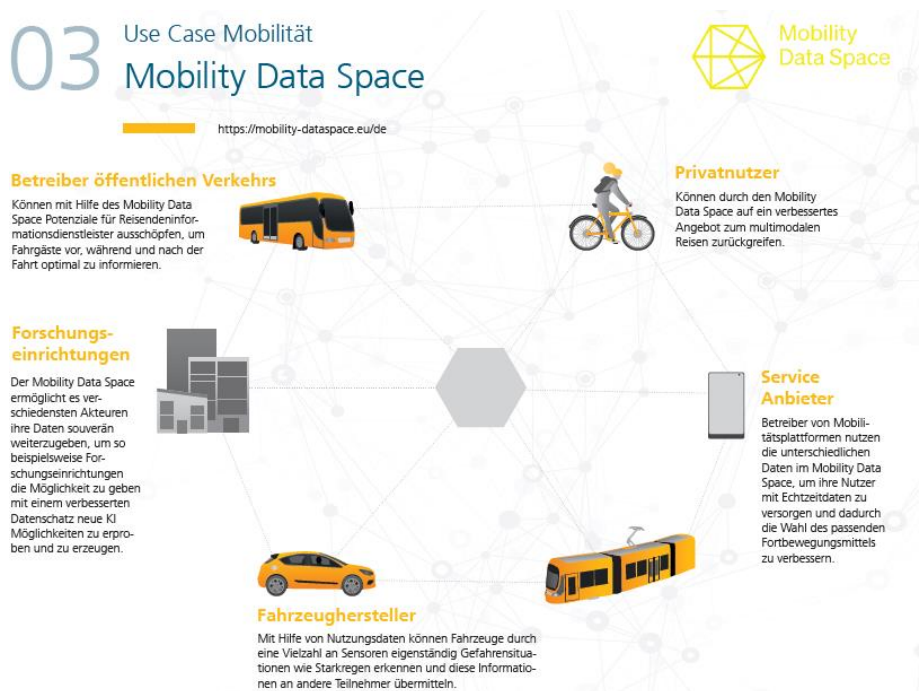


Abbildung 8 Exemplarischer Use Case aus dem IEDS Use Case Report

Designoptionen von Konnektoren für den Datenaustausch in Datenräumen

Die Veröffentlichung „What Does it Take to Connect? Unveiling Characteristics of Data Space Connectors“ von Anna Gieß et al. 2024 untersucht die essenziellen Eigenschaften von Data Space Connectors, die eine zentrale Rolle beim sicheren und effizienten Datenaustausch in sogenannten Data Spaces spielen. Diese Data Spaces ermöglichen es,

große, heterogene Datensätze aus verschiedenen Quellen ohne physische Integration zu speichern und auszutauschen. Die Verbindung zwischen diesen Datenquellen und den Data Spaces wird durch spezielle Softwarekomponenten, die sogenannten Connectors, hergestellt.

Da es keine einheitliche Definition oder umfassende Charakterisierung dieser Connectors gab, haben die Autoren eine Taxonomie entwickelt, die die wichtigsten Merkmale von Data Space Connectors systematisch beschreibt. Um dies zu erreichen, analysierten sie 23 reale Anwendungsfälle, relevante Fachliteratur sowie wissenschaftliche Arbeiten und führten eine Expertengruppe zusammen, um ein breites Spektrum an Einsichten zu gewinnen.

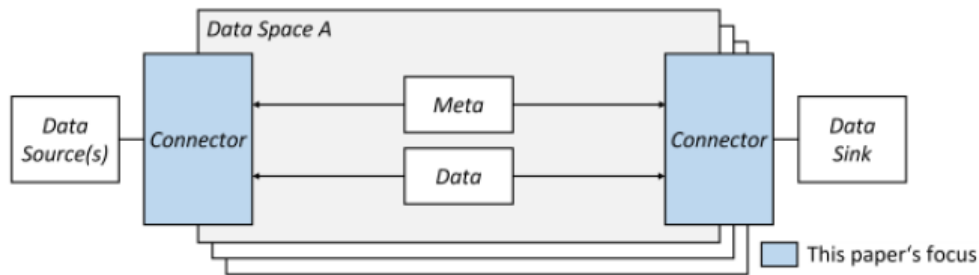


Abbildung 9 Datenaustausch in Datenräumen aus Gieß et al. 2024 „What Does it Take to Connect? Unveiling Characteristics of Data Space Connectors“

Die resultierende Taxonomie umfasst neun Dimensionen und 25 Merkmale, die unterschiedliche Aspekte der Data Space Connectors abbilden. Zu den Dimensionen gehören unter anderem die Art der Lizenz (Open-Source oder Closed-Source), Bereitstellungsoptionen wie Cloud-, Edge- oder On-Premises-Lösungen, das Service-Level, die Portabilität und die Art der Zugriffssteuerung, etwa über OAuth. Ein weiterer wichtiger Punkt ist die Interoperabilität, die durch verschiedene Protokolle und Datenverwaltungsspezifikationen, wie Identitätsmanagement und Katalogfunktionen, gewährleistet wird.

Dimension		Characteristics				E/N
License		Open source	Open source (copyleft)	Closed source (extendable)	Closed source	E
Deployment options		Edge		On-premises	Cloud	N
Service level		Connector as a service		Platform as a service	Self-service	E
Portability		Agnostic		Specific		E
Access control		OAuth		Basic auth	API key	E
Interoperability	Communication protocol	IDS protocol (Multipart / IDSCP)		Data space protocol (HTTPS)		N
	Transfer protocol	Determined		Undetermined		E
Graphical user interface		Available		Not available		E
Supported data space specifications		Identity management	Catalog	Vocabulary	...	N

Abbildung 10 Designoptionen für Datenraumkonnektoren aus Gieß et al. 2024 „What Does it Take to Connect? Unveiling Characteristics of Data Space Connectors“

Anhand von zwei Anwendungsfällen, die in der Praxis untersucht wurden, verdeutlichen die Autoren die Anwendung ihrer Taxonomie. Eines der Beispiele stammt von soivity, einem Anbieter von Data Space-Lösungen. Hier wird gezeigt, wie sich eine kostenlose Open-Source-Version eines Connectors von einer kostenpflichtigen Cloud-basierten Lösung unterscheidet, die zusätzliche Funktionen und Support bietet.

Die Veröffentlichung „Design Options for Data Spaces“ von Anna Gieß und Kollegen, präsentiert auf der 31. Europäischen Konferenz für Informationssysteme (ECIS 2023), untersucht die Gestaltungsmöglichkeiten für Data Spaces, die als zentrale Komponenten der Datenwirtschaft in verschiedenen europäischen Forschungsinitiativen dienen. Data Spaces bieten eine Infrastruktur für den sicheren, organisationsübergreifenden Datenaustausch und ermöglichen die Sammlung und Integration heterogener Datenquellen. Trotz des großen Interesses an der Entwicklung neuer Data Spaces bleibt unklar, was einen Data Space definiert und wie man ihn am besten gestaltet.

Um diese Fragen zu klären, führten die Autoren eine umfassende Analyse von 53 wissenschaftlichen Arbeiten sowie 47 realen Data Spaces durch. Auf Basis dieser Untersuchung entwickelten sie eine Taxonomie, die die wesentlichen Designoptionen für Data Spaces strukturiert und zusammenfasst. Diese Taxonomie dient sowohl als theoretisches Modell für die Forschung als auch als praktisches Werkzeug für die Implementierung von Data Spaces. Ziel ist es, sowohl Forschern als auch Praktikern ein besseres Verständnis dieser neuen Datenmanagementstrukturen zu ermöglichen.

Die Arbeit betont die Relevanz von Data Spaces als technologischem Rahmenwerk, das speziell für den interoperablen und sicheren Datenaustausch zwischen verschiedenen Organisationen und innerhalb von Datenökosystemen entwickelt wurde. Im Gegensatz zu traditionellen zentralisierten Datenplattformen basieren Data Spaces auf einer föderierten Architektur, die es ermöglicht, neue Wertschöpfungspotenziale durch die gemeinsame Nutzung von Daten zu realisieren. Die Autoren betonen die Vielfalt der möglichen Anwendungsbereiche, wie z. B. den Mobilitätsdatenraum und den Energiemarkt, und zeigen auf, wie Data Spaces in verschiedenen Domänen unterschiedlich konzipiert und gestaltet werden können.

Die entwickelte Taxonomie umfasst drei zentrale Meta-Dimensionen: wirtschaftliche, technische und governance-bezogene Aspekte. Sie untergliedert sich weiter in 17 Dimensionen und 50 Merkmale, die die wesentlichen Designoptionen für Data Spaces abdecken. Diese reichen von der Architektur des Data Spaces über die verwendeten Datenstrukturen und -typen bis hin zu den Mechanismen zur Gewährleistung von Vertrauen und Datensouveränität.

Zusammenfassend liefert die Arbeit einen wichtigen Beitrag zur Systematisierung und Analyse von Data Spaces. Die vorgeschlagene Taxonomie bietet sowohl eine theoretische Grundlage für zukünftige Forschung als auch praxisnahe Anleitungen für die Entwicklung und Implementierung von Data Spaces in verschiedenen Domänen. Die Autoren betonen, dass die Taxonomie als „Big Picture“ für die aufkommende Forschung in diesem Bereich dient und als Ausgangspunkt für weiterführende Studien und praxisorientierte Untersuchungen genutzt werden kann.

MD	Dimension (D _n)	Characteristics (C _{1,m})					E/N	
Economic	Domain	Domain-specific			Cross-domain		E	
	Funding	Public	Private		Private-public partnership		E	
	Data space access	Free			Fee		E	
	Reward	Money	Data	Service	Reputation	None	N	
	Value added services	Yes			No		E	
	Data structure	Structured		Semi-structured	Unstructured		N	
Technical	Data type	Raw data		Processed data	Metadata		N	
	Data processing	Stream			Batch		N	
	Architecture	Centralized		Decentralized	Hybrid		E	
	Data sharing logic	P2P data sharing		Data platform	Data sharing via intermediaries		N	
	Data harmonization	Data models			Data catalog		N	
	Access technology	Standardized connector			Portal		N	
	Trusted exchange	Trust by identity management			Trust by certification		N	
	Data privacy	Anonymous	Pseudonymous	Non-anonymous	Various		E	
	Data classification scheme	Domain	Origin	Topicality	Size	Data format	...	N
	Data sharing policies	Set by data space			Set by data provider		N	
Governance	Traceability and control	Space dimension	Time dimension	Use dimension	None		N	

Note: E = Exclusive, N = Non-exclusive.

Abbildung 11 Taxonomie zu Designoptionen für Datenräume aus Gieß et al. 2023 „Design Options for Data Spaces“

Im vierten Arbeitspaket wurden Forschungen zum Thema **Datenbewertung** durchgeführt.

Status Quo, Anreize und Methoden der Datenbewertung

Der Bericht „Status quo und Anreize der Datenbewertung“ untersucht die Rolle von Daten als immaterielle Vermögenswerte in Unternehmen, die zunehmend die Grundlage für den Unternehmenswert bilden. Während physische Vermögenswerte traditionell den größten Teil des Unternehmenswerts ausmachten, haben immaterielle Vermögenswerte, einschließlich Daten, heute eine entscheidende Bedeutung erlangt. Dabei bleibt jedoch die Bilanzierung von Daten herausfordernd, da bilanzierungsfähige immaterielle Vermögenswerte gemäß aktuellen Bilanzierungsstandards (z.B. IAS, HGB) nur zu ihren Anschaffungs- oder Herstellungskosten bewertet werden dürfen. Dies führt dazu, dass der tatsächliche wirtschaftliche Wert von Daten, der oft deutlich höher ist als die Kosten, in der Bilanz nicht vollständig abgebildet wird.

Der Bericht beschreibt unterschiedliche Ansätze zur Bewertung von Daten: Das marktpreisorientierte Verfahren basiert auf dem Marktwert von Daten, das kostenorientierte Verfahren auf den Anschaffungs- und Herstellungskosten, und das nutzenorientierte Verfahren schätzt den zukünftigen Nutzen, den Daten für ein Unternehmen bringen können. Besonders das nutzenorientierte Verfahren ist für die Bewertung von Daten attraktiv, da es die Gewinnerwartung über den gesamten Nutzungszeitraum berücksichtigt, jedoch auch am komplexesten ist.



Abbildung 12 Informationsmerkmale für Datenbewertung

Die Autoren betonen, dass die Bilanzierung von Daten zwar die Außendarstellung von Unternehmen verbessern könnte, insbesondere für junge und innovative Unternehmen, allerdings nicht ausreichend ist, um den Unterschied zwischen Markt- und Buchwert vollständig zu eliminieren. Unternehmen sollten sich daher bewusst über den Wert ihrer Daten sein und diese sowohl in der Bilanz berücksichtigen als auch in ihre strategischen Entscheidungen einbeziehen. Datenbewertung kann als strategisches Werkzeug genutzt werden, um Wettbewerbsvorteile zu erlangen, Prozesse zu optimieren und das eigene Geschäftsmodell datenbasiert weiterzuentwickeln. Darüber hinaus wird Data Sharing als potenzieller Ansatz für die Wertsteigerung von Daten hervorgehoben, da die gemeinsame Nutzung von Daten zwischen verschiedenen Akteuren zusätzliche Wertschöpfung ermöglichen kann.

Insgesamt wird im Bericht klar, dass die Bewertung von Daten in Zukunft an Bedeutung gewinnen wird, insbesondere angesichts der fortschreitenden Digitalisierung und der zunehmenden Bedeutung datenbasierter Geschäftsmodelle. Unternehmen, die ihre Daten als wertvolle Ressourcen erkennen und effektiv managen, können langfristig Wettbewerbsvorteile erzielen und ihre Marktposition stärken.

Datenbewertung					
finanzielle Bewertungsverfahren			nicht-finanzielle Bewertungsverfahren		
marktpreisorientiert	kostenorientiert	nutzenorientiert	qualitätsorientiert	prozessorientiert	performanceorientiert
• Preis basiert auf verfügbaren Marktpreisen	• Preis basiert auf Herstellungs- bzw. Anschaffungskosten	• Preis basiert auf der Gewinnerwartung, abzüglich der entstandenen Kosten	• Qualitätsmerkmale wie Zugänglichkeit, Aktualität oder Verwendbarkeit	• Wert basiert auf der Relation zu anderen Prozessen	• Bewertung über die Auswirkung auf KPIs

Abbildung 13 Ansätze zur Datenbewertung

Das Fraunhofer ISST hat als Konsortialführer das **Projektmanagement** verantwortet und hat alle Partner bei der Verbreitung der Ergebnisse unterstützt. Zusammen oblagen dem Fraunhofer ISST die Managementaufgaben des Projekts. Dazu zählten die stetige Überprüfung der Zeitplanung und des Projektfortschritts. Es wurde ein Risiko- und Qualitätsmanagement durchgeführt, um mögliche entstehende Risiken und Gegenmaßnahmen in den Arbeitspaketen zu identifizieren und im Ablauf des Projekts zu berücksichtigen, damit die gesetzten Ziele und Forschungsaktivitäten erfolgreich durchgeführt werden konnten. Als Instrumente der Projektüberwachung sind Meilensteinplanungen und regelmäßige Statusmeetings durchgeführt und geleitet worden. Dies ermöglichte neben der Projektkontrolle auch die Koordination der

Interaktion der projektbeteiligten Institute. Es wurden regelmäßige Meetings sowohl auf Arbeitspaketebene als auch auf Gesamtprojektebene durchgeführt, sodass alle Aktivitäten koordiniert werden konnten und eine Zusammenarbeit aller Institute mit größtmöglichem Erfolg stattfand. Darüber hinaus wurden im halbjährlichen Turnus Konsortialtreffen zum Austausch der Ergebnisse und Abstimmung des weiteren Vorgehens veranstaltet. Das Fraunhofer ISST koordinierte auch die Verbreitung der Ergebnisse, um sie einer breiten Masse zugänglich zu machen. Zu diesem Zweck wurde die IEDS-Plattform auf ieds-projekt.de aufgebaut, auf der neben allen Forschungsergebnissen auch das Gaia-X Dashboard sowie der interaktive Vertragsgenerator zu finden sind. Darüber hinaus sind die Forschungsergebnisse auf diversen Social-Media-Kanälen verbreitet worden und es gab eine Vielzahl an Fachvorträgen, Podiumsdiskussionen und Konferenzbeiträgen der IEDS-Wissenschaftler, die der Verbreitung der Ergebnisse dienen. Ferner wurde das Berichtswesen des Projekts koordiniert und dem Projektträger wurde im Rahmen von verschiedenen Berichten aller Konsortialpartner über die vorläufigen und Endgültigen Projektergebnisse sowie allen Aktivitäten informiert

Aufzählungen der getätigten und geplanten Veröffentlichungen:

Gelhaar, J. / Müller, P. / Bergmann, N. / Dogan, R., 2023, Motives and Incentives for Data Sharing in Industrial Data Ecosystems: An Explorative Single Case Study. In: Proceedings of the 56th Hawaii International Conference on System Sciences, p. 3705–3714, <https://ieds-projekt.de/wp-content/uploads/2023/03/ISST-HICCS-Gelhaar.pdf>

Frings, S. / Kett, H. / Falkner, J., 2022, Essential Areas of Action for the Smartification of Traditional Products and Services Considering the Capabilities of SMEs, In: Christine Leitner, Walter Ganz, Clara Bassano and Debra Satterfield (eds) The Human Side of Service Engineering. AHFE (2022) International Conference. AHFE Open Access, vol 62. AHFE International, USA. <http://doi.org/10.54941/ahfe1002571>

Frings, S. / Kett, H. / Falkner, J., 2023, Enhancing a Capability Maturity Model for the Smartification of Products by a Generic Approach and a Third Dimension. In: Tareq Ahram, Waldemar Karwowski, Pepetto Di Bucchianico, Redha Taiar, Luca Casarotto and Pietro Costa (eds) Intelligent Human Systems Integration (IHSI 2023): Integrating People and Intelligent Systems. AHFE (2023) International Conference. AHFE Open Access, vol 69. AHFE International, USA. <http://doi.org/10.54941/ahfe1002907>

Frings, S. / Kett, H., 2023, Method to Identify Data Related Characteristics for Detailing a Capability Maturity Model for the Smartification of Products. In: Christine Leitner, Jens Neuhüttler, Clara Bassano and Debra Satterfield (eds) The Human Side of Service Engineering. AHFE (2023) International Conference. AHFE Open Access, vol 108. AHFE International, USA. <http://doi.org/10.54941/ahfe1003101>

Gür, Inan / Spiekermann, Markus, 2022, Challenges and Requirements for interorganizational Data Sharing, ISST-Bericht, Dortmund, https://ieds-projekt.de/wp-content/uploads/2024/04/DataOps_Fraunhofer-ISST-Report.pdf

Hupperz, Marius / Groß, Tobias / Spiekermann, Marius, 2022, Datenbewertung – Status quo und Anreize der Datenbewertung, ISST Bericht, Dortmund, <https://ieds-projekt.de/wp-content/uploads/2024/04/Fraunhofer-ISST-Report-Datenbewertung-1.pdf>

Frings, Sandra / Gür, Inan, 2024, Data Sharing Reifegradmodell - Strategische Aspekte des Data Sharing, Editiert von Wilhelm Bauer and Boris Otto. Fraunhofer IAO. <https://doi.org/10.24406/publica-3117>

Gieß, Anna / Hupperz, Marius / Schoormann, Thorsten / Möller, Frederik, 2024, What Does it Take to Connect? Unveiling Characteristics of Data Space Connectors, Proceedings of the 57th Hawaii International Conference on System Sciences, Hawaii, <https://ieds-projekt.de/wp-content/uploads/2024/01/0418.pdf>

Gieß, Anna / Möller, Frederik / Schoormann, Thorsten / Otto, Boris, 2023, Design Options for Data Spaces. In: Thirty-first European Conference on Information Systems, Kristiansand, Norway, <https://ieds-projekt.de/wp-content/uploads/2023/09/ECIS2023-DESIGNOPTIONSFORDATASPACE.pdf>

Gür, Inan / Möller, Frederik / Hupperz, Marius / Uzun, Dilara / Otto, Boris, 2022, Requirements for DataOps to foster Dynamic Capabilities in Organizations – A mixed methods approach. In: 24th IEEE International Conference on Business Informatics, Amsterdam, Netherlands

Azkan, Can / Möller, Frederik / Ebel, Martin / Iqbal, Taskeen / Otto, Boris / Poppelbuss, Jens, 2022, Hunting the Treasure: Modeling Data Ecosystem Participant Value, Forty-Third International Conference on Information Systems, Copenhagen

Gür, Inan / Gieß, Anna / Jussen, Ilka / Schweihoff, Julia, 2023, Use Case Report, Dortmund, <https://ieds-projekt.de/wp-content/uploads/2023/08/Use-Case-Report.pdf>

Azkan, Can / Gür, Inan / Gelhaar, Joshua / Gieß, Anna / Groß, Tobias / Hupperz, Marius / Kett, Holger / Kutzias, Damian / Strauß, Oliver / Frings, Sandra / Büchel, Jan / Demary, Vera / Engels, Barbara / Goecke, Henry / Mertens, Arne / Röhl, Klaus-Heiner / Rusche, Christian / Scheufen, Marc/ Schröder, Bjarne / Möller, Frederik / Schweihoff, Julia / Jussen, Ilka / Dahms, Valentin / Andres, Raphaela / Erdsiek, Daniel / Niebel, Thomas / Rehse, Dominik / Rost, Vincent / Walter, Johannes / Valet, Sebastian, 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>

Azkan, Can / Gür, Inan / Gelhaar, Joshua / Gieß, Anna / Groß, Tobias / Hupperz, Marius / Kett, Holger / Kutzias, Damian / Strauß, Oliver / Frings, Sandra / Büchel, Jan / Demary, Vera / Engels, Barbara / Goecke, Henry / Mertens, Arne / Röhl, Klaus-Heiner / Rusche, Christian / Scheufen, Marc/ Schröder, Bjarne / Möller, Frederik / Schweihoff, Julia / Jussen, Ilka / Andres, Raphaela / Erdsiek, Daniel / Niebel, Thomas / Rehse, Dominik / Rost, Vincent / Walter, Johannes / Valet, Sebastian, 2023, Anreizsysteme und Ökonomie des Data Sharing - Status Quo der deutschen Datenwirtschaft und Anwendung von unternehmensübergreifendem Datenaustausch, https://ieds-projekt.de/wp-content/uploads/2023/04/IEDS-Whitepaper_2023_finale-Version.pdf

Azkan, Can / Gür, Inan / Gieß, Anna / Hupperz, Marius / Kett, Holger / Strauß, Oliver / Frings, Sandra / Bakalis, Dennis / Bolwin, Lennart / Büchel, Jan / Demary, Vera / Engels, Barbara / Goecke, Henry / Mertens, Arne / Röhl, Klaus-Heiner / Rusche, Christian / Gruben, Fabian / Kestermann, Christian / Scheufen, Marc/ Schröder, Bjarne / Möller, Frederik / Schweihoff, Julia / Jussen, Ilka / Andres, Raphaela / Erdsiek, Daniel / Niebel, Thomas / Rehse, Dominik / Rost, Vincent / Walter, Johannes / Valet, Sebastian, 2024, Anreizsysteme und Ökonomie des Data Sharing - Status quo der deutschen Datenwirtschaft und Grundlagen des unternehmensübergreifenden Datenaustausches, https://ieds-projekt.de/wp-content/uploads/2024/04/IEDS-Whitepaper_2024_Final.pdf

Jussen, Ilka / Möller, Frederik / Schweihoff, Julia / Gieß, Anna / Giussani, Giulia / Otto, Boris, 2024, Issues in inter-organizational data sharing: Findings from practice and research challenges, In: Data & Knowledge Engineering, Volume 150, <https://doi.org/10.1016/j.datak.2024.102280>

Gieß, Anna / Schoormann, Thorsten / Möller, Frederik / Gür, Inan, 2024, Discovering Data Spaces: A Classification of Design Options, In: Computers in Industry (eingereicht, noch nicht veröffentlicht)

Strauß, Oliver / Kutzias, Damian / Kett, Holger, 2022, Agent-based Document Expansion for Information Retrieval based on Topic Modeling of local Information, Proceedings of the 2022 9th Intl. Conference on Soft Computing & Machine Intelligence (ISCMI 2022), November 26-27, 2022, Toronto, Canada, <http://doi.org/10.1109/ISCMI56532.2022.10068457>, https://ieds-projekt.de/wp-content/uploads/2023/09/ISCMI_2022_Document_expansion_with_IEEE_copyright.pdf

Strauß, O. / Kett, H., 2023, Documents as Intelligent Agents: An Approach to Optimize Document Representations in Semantic Search. In Proceedings of the 19th International Conference on Web Information Systems and Technologies - WEBIST; ISBN 978-989-758-672-9; ISSN 2184-3252, SciTePress, pages 164-175. <http://doi.org/10.5220/0012239200003584>

Strauß, O. / Kett, H., 2023, Agent-based Optimization of Document Representations in Semantic Search (eingereicht bei Springer Nature Computer Science).

Wege in die Datenökonomie. Forward-Magazin 02/23 des Fraunhofer IAO. <https://public-rest.fraunhofer.de/server/api/core/bitstreams/a38721bb-3173-43b6-ab54-05c2fb89207c/content>

FRAUNHOFER-INSTITUT FÜR SOFTWARE- UND SYSTEMTECHNIK, ISST
FRAUNHOFER-INSTITUT FÜR ARBEITSWIRTSCHAFT UND ORGANISATION, IAO

IEDS PROJEKT SACHBERICHT TEIL III ERFOLGSKONTROLLBERICHT

Fraunhofer Gesellschaft

Konsortialführer und Zuwendungsempfänger: Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.	Förderkennzeichen: IEDS001
Vorhabenbezeichnung: IEDS – Incentives and Economics of Data Sharing	
Laufzeit des Vorhabens: 01.05.2021 – 31.04.2024	
Berichtszeitraum: 01.05.2021 – 31.04.2024	
Fälligkeitsdatum: 31. Oktober 2024	
Erstellungsdatum: 17.10.2024	

Ansprechperson Fraunhofer ISST:

Heinrich Pettenpohl

Fraunhofer-Institut für Software- und Systemtechnik

Speicherstraße 6, 44147 Dortmund

Heinrich.Pettenpohl@isst.fraunhofer.de

Tel.: +49 231 97677-321

Ansprechperson Fraunhofer IAO:

Dr. Holger Kett

Fraunhofer-Institut für Arbeitswirtschaft und Organisation

Nobelstraße 12, 70569 Stuttgart

Holger.Kett@iao.fraunhofer.de

Tel.: +49 711 970 2415

Erfolgskontrollbericht

I. Wissenschaftlich-technisches Ergebnis des Vorhabens, die erreichten Nebenergebnisse und die gesammelten wesentlichen Erfahrungen

Die Forschungsarbeiten im Rahmen des Projekts, geleitet von der Fraunhofer-Gesellschaft, erzielten bedeutende wissenschaftliche und praktische Ergebnisse im Bereich des Data Sharing und der datengetriebenen Wertschöpfung, insbesondere in industriellen Datenökosystemen. Diese Ergebnisse tragen wesentlich zur Weiterentwicklung der Datenwirtschaft und zur Schaffung neuer Geschäftsmodelle auf Basis von Daten bei.

Ein zentrales Ergebnis des Projekts war die detaillierte Untersuchung von Anreizen und Motiven, die Unternehmen dazu bewegen, Daten in industriellen Datenökosystemen zu teilen. Im Arbeitspaket "**Data Economy Readiness**" wurde eine Fallstudie zum Catena-X-Ökosystem, einem wichtigen Datenökosystem der Automobilindustrie, durchgeführt. Die Ergebnisse, veröffentlicht im wissenschaftlichen Paper "Motives and Incentives for Data Sharing in Industrial Data Ecosystems: An Explorative Single Case Study", zeigen, dass viele Unternehmen zögern, Daten zu teilen, weil der direkte Nutzen oft nicht klar erkennbar ist. Die Studie identifizierte jedoch sieben Motive und acht Anreize, die Unternehmen dazu bewegen könnten, Daten zu teilen. Wirtschaftliche Motive wie Kosteneinsparungen, Transparenz und strategische Partnerschaften spielen eine zentrale Rolle, aber auch rechtliche Anforderungen und soziale Aspekte, wie der Druck durch Nachhaltigkeitsforderungen der Kunden, sind von Bedeutung. Diese Erkenntnisse bieten einen wichtigen theoretischen und praktischen Beitrag zur Gestaltung von Anreizsystemen für das Data Sharing in unterschiedlichen Sektoren. Die Studie betont zudem, dass der Mehrwert für die Datenanbieter vor allem aus spezifischen Anwendungsfällen resultiert und nicht allein aus der Teilnahme an einem Ökosystem.

Im Arbeitspaket zu "**Datenstrategie und -management**" lag der Fokus auf der Optimierung der Datenbereitstellung, der Datenstrategien und der Datensuche. Der veröffentlichte Report „DataOps for Data Sharing: Herausforderungen und Anforderungen für organisationsübergreifendes Teilen von Daten“ hebt hervor, dass DataOps es Unternehmen ermöglicht, Daten effizienter und automatisierter bereitzustellen und so die Qualität und Geschwindigkeit der Datenprozesse zu verbessern. DataOps reduziert Datensilos und fördert die Zusammenarbeit zwischen verschiedenen Abteilungen innerhalb eines Unternehmens sowie organisationsübergreifend. Dies ist von besonderer Bedeutung, da viele Unternehmen Schwierigkeiten haben, qualitativ hochwertige Daten rechtzeitig zur Verfügung zu stellen. Die Implementierung von DataOps kann diese Herausforderungen durch standardisierte und automatisierte Datenpipelines überwinden.

Darüber hinaus wurden neue Ansätze für die Datensuche und das Matching von Datenangeboten und -nachfragen entwickelt. Ein besonderer Fokus lag auf der Anwendung von agentenbasierten Systemen und KI-gestützten Matching-Verfahren, die Datenmarktplätze und Datenräume effizienter gestalten. Die Forschung zeigte, dass durch den Einsatz intelligenter Software-Agenten die Darstellung von Datenprodukten auf Marktplätzen optimiert und somit die Auffindbarkeit verbessert werden kann. Dieser Ansatz ist besonders zukunftsweisend für den Einsatz in Data Spaces wie Gaia-X, da er die Datensouveränität der Anbieter wahrt und gleichzeitig eine dynamische Anpassung der Datenangebote an die Bedürfnisse der Nutzer ermöglicht. Die entwickelte Lösung trägt entscheidend dazu bei, die Herausforderungen bei der Suche nach spezifischen Daten zu überwinden und schafft eine Grundlage für eine effizientere und zielgerichtete Nutzung von Daten in verschiedenen Sektoren, von der Industrie bis hin zur Wissenschaft.

Ein weiteres bedeutendes Ergebnis war die Entwicklung eines Reifegradmodells für die strategischen Aspekte des Data Sharing, das Unternehmen dabei unterstützt, ihre gegenwärtigen Praktiken des Datenaustauschs zu bewerten und Verbesserungspotenziale zu identifizieren. Der Bericht "Data Sharing Reifegradmodell - Strategische Aspekte des Data Sharing" analysierte die zunehmende Bedeutung des Datenaustauschs für die Entwicklung innovativer Geschäftsmodelle und digitaler Produkt-Service-Systeme. Das Reifegradmodell umfasst sechs Schlüsseldimensionen – Organisation, Prozesse, Wert, Daten, Technologie und Governance – und bietet Unternehmen ein strukturiertes Werkzeug, um ihre Data Sharing-Fähigkeiten zu bewerten und zu verbessern. Diese Dimensionen decken alle relevanten Bereiche ab, die für eine erfolgreiche Integration von Daten in die Geschäftsstrategien notwendig sind. Das Modell kann flexibel auf verschiedene Branchen und Unternehmensgrößen angewendet werden und bietet eine wertvolle Orientierungshilfe für Unternehmen, die ihre strategischen Fähigkeiten im Umgang mit Daten stärken möchten. Durch die Anwendung dieses Modells können Unternehmen langfristig Wettbewerbsvorteile erlangen, indem sie ihre Datenressourcen gezielt nutzen und in neue Geschäftsmodelle integrieren.

Im Rahmen des Arbeitspakets "**datengetriebene Geschäftsmodelle**" wurden neue Ansätze zur Modellierung von Wertschöpfung in Datenökosystemen entwickelt. Insbesondere das wissenschaftliche Paper "Hunting the Treasure: Modeling Data Ecosystem Value Co-Creation" befasste sich mit der Rolle von Datenökosystemen bei der interorganisationalen Wertschöpfung. Die Studie untersuchte 64 Anwendungsfälle der Gaia-X-Initiative und analysierte, wie unterschiedliche Akteure in Datenökosystemen zusammenarbeiten, um gemeinsam Wert zu schaffen. Die Studie identifizierte acht Schlüsselrollen innerhalb solcher Ökosysteme, darunter Datenanbieter, Cloud-Plattform-Anbieter, Ökosystem-Orchestratoren und Datenmarktplatz-Betreiber. Durch die Verwendung der e3-value-Modellierungssprache wurde veranschaulicht, wie diese Akteure interagieren, um durch den Austausch von Daten neue Werte zu schaffen. Diese Erkenntnisse sind besonders wertvoll für die Praxis, da sie Unternehmen helfen, ihre Rolle in Datenökosystemen besser zu verstehen und zu optimieren. Die Studie unterstreicht zudem die Bedeutung von Dienstleistungsökosystemen, bei denen der Wert nicht nur im Austausch (value-in-exchange), sondern auch in der Nutzung (value-in-use) entsteht, was eine neue Perspektive auf die gemeinsame Wertschöpfung durch Daten bietet. Darüber hinaus untersuchten Forschungsarbeiten die essenziellen Eigenschaften von Data Space Connectors und deren zentrale Rolle beim sicheren und effizienten Datenaustausch in sogenannten Data Spaces. Die Autoren entwickelten eine Taxonomie, die 23 reale Anwendungsfälle und wissenschaftliche Arbeiten analysiert, um die wichtigsten Merkmale von Data Space Connectors systematisch zu beschreiben. Diese Taxonomie umfasst neun Dimensionen und 25 Merkmale, wie Lizenzarten, Bereitstellungsoptionen und Interoperabilität. Zudem beleuchtet sie praktische Unterschiede zwischen Open-Source- und Cloud-basierten Connector-Lösungen. Eine weitere Untersuchung analysierte die Designoptionen für Data Spaces, die als technologische Rahmenwerke für den organisationsübergreifenden Datenaustausch fungieren. Auch hier wurde eine Taxonomie entwickelt, die 17 Dimensionen und 50 Merkmale umfasst, um Data Spaces strukturell und funktional zu klassifizieren. Diese Taxonomie bietet sowohl ein theoretisches Modell für die Forschung als auch praktische Anleitungen zur Implementierung von Data Spaces in verschiedenen Domänen, etwa im Mobilitäts- oder Energiemarkt. Die Arbeit stellt einen wichtigen Beitrag zur Systematisierung und Weiterentwicklung von Data Spaces dar, indem sie die komplexen Aspekte dieser Strukturen umfassend analysiert.

Im Zusammenhang mit der **Datenbewertung** wurden Forschungsarbeiten veröffentlicht, die die Rolle von Daten als immaterielle Vermögenswerte in Unternehmen untersucht. Die Autoren betonten, dass die Bewertung von Daten eine zentrale Rolle für den Unternehmenswert spielen kann, insbesondere in Zeiten der fortschreitenden

Digitalisierung. Verschiedene Bewertungsansätze wurden analysiert, darunter marktpreisorientierte, kostenbasierte und nutzenorientierte Verfahren. Der Bericht zeigt auf, dass das nutzenorientierte Verfahren, das den zukünftigen wirtschaftlichen Nutzen von Daten berücksichtigt, das größte Potenzial zur Wertsteigerung bietet, jedoch auch am komplexesten ist. Die Forschung macht deutlich, dass Unternehmen ihre Daten als wertvolle Ressourcen betrachten und in ihre strategischen Entscheidungen einbeziehen sollten, um langfristig Wettbewerbsvorteile zu sichern.

Als Konsortialführer war das Fraunhofer ISST zudem für das **Projektmanagement** und die Verbreitung der Ergebnisse verantwortlich. Durch regelmäßige Statusmeetings und Meilensteinplanungen wurde der Fortschritt des Projekts kontinuierlich überwacht und gesteuert. Die Verbreitung der Ergebnisse erfolgte über verschiedene Kanäle, darunter die IEDS-Plattform, Social-Media-Präsenzen und eine Vielzahl von Fachvorträgen und Konferenzen. Zudem wurden die Forschungsergebnisse in wissenschaftlichen Publikationen veröffentlicht, um eine breite Fachöffentlichkeit zu erreichen. Diese Maßnahmen trugen dazu bei, die Erkenntnisse des Projekts einer breiten Masse von Unternehmen und Forschern zugänglich zu machen und somit die praktische Anwendung der entwickelten Ansätze zu fördern.

Insgesamt leistete das Projekt einen wichtigen Beitrag zur Förderung des Data Sharing und zur Weiterentwicklung der Datenwirtschaft. Die entwickelten Anreizsysteme, Reifegradmodelle, Innovationsansätze im Bereich der Datensuche und die praktischen Anwendungsfälle bieten wertvolle Impulse für Unternehmen, die ihre datengetriebenen Geschäftsmodelle optimieren und ihre Position in der globalen Datenökonomie stärken möchten. Die Ergebnisse des Projekts zeigen deutlich, dass die effektive Nutzung von Daten in der modernen Wirtschaft nicht nur einen Wettbewerbsvorteil darstellt, sondern zunehmend zur Schlüsselfähigkeit für den langfristigen Erfolg von Unternehmen wird.

II. Fortschreibung des Verwertungsplans. Diese soll, soweit im Einzelfall zutreffend, Angaben zu folgenden Punkten enthalten:

a. Erfindungen/Schutzrechtsanmeldungen und erteilte Schutzrechte

Es gibt seitens der Fraunhofer-Institute ISST und IAO keine eingereichten Erfindungsmeldungen, die in direktem Zusammenhang mit den Ergebnissen des Projekts stehen. Weiterhin liegen keine Rechte Dritter vor, die einer Verwertung der Projektergebnisse entgegenstehen könnten. Fraunhofer ISST und Fraunhofer IAO halten keine eigenen Schutzrechte an den im Rahmen von IEDS erzielten Ergebnissen.

b. Wirtschaftliche Erfolgsaussichten nach Projektende

Die Ergebnisse des Forschungsprojekts leisten wirtschaftlichen Akteuren und allen voran der Datenökonomie in Deutschland Vorschub leisten. Das IEDS-Projekt hat nicht nur aufgezeigt wie Data Sharing effizient und erfolgreich durchgeführt werden kann, sondern auch welche Vorteile eine datengetriebene Wertschöpfung Unternehmen bieten. Durch das Teilen von Daten können Unternehmen ineffiziente Prozesse optimieren, Ressourcen besser planen und Kosten senken. Zum Beispiel können Unternehmen durch den Datenaustausch innerhalb einer Lieferkette Transparenz über Bestände und Nachfragen gewinnen, was den sogenannten Bullwhip-Effekt reduziert, bei dem Nachfrageschwankungen zu Über- oder Unterproduktion führen. Data Sharing ermöglicht es, Daten für Forschungs- und Entwicklungsprojekte zu nutzen, was neue Innovationen in verschiedenen Industrien antreibt. Unternehmen können datenbasierte Erkenntnisse verwenden, um neue Produkte und Dienstleistungen zu entwickeln, die auf Trends und Kundenbedürfnissen basieren. Je mehr Daten in einem Ökosystem geteilt werden, desto wertvoller können diese für alle beteiligten Akteure sein. Daten, die von

mehreren Unternehmen oder Organisationen genutzt werden, haben eine höhere Wirkung und können breiter angewendet werden. Auch die öffentliche Hand kann von Data Sharing profitieren, beispielsweise im Gesundheitswesen, in der Verkehrsplanung oder bei der Bekämpfung des Klimawandels. Der Zugang zu öffentlichen und privaten Daten kann helfen, soziale Herausforderungen zu bewältigen, Innovationen in der öffentlichen Verwaltung voranzutreiben und die Lebensqualität zu verbessern. Data Sharing treibt wirtschaftliches Wachstum, Innovation und Effizienz voran und schafft die Basis für neue Geschäftsmodelle und Kooperationen. Das IEDS-Projekt hat auf Basis empirischer Daten aufgezeigt, welche Hürden und Herausforderungen in Deutschland für Unternehmen noch bestehen und hat Lösungsansätze geboten, um die datengetriebene Wertschöpfung in Deutschland voranzubringen.

c. Wissenschaftliche und/oder technische Erfolgsaussichten nach Projektende

Im Verlauf des IEDS-Projekts sind diverse wissenschaftliche Veröffentlichungen auf hochrangigen und weitverbreiteten Konferenzen getätigt worden. Die wissenschaftlichen Arbeiten und Veröffentlichungen fanden Anklang und wurden bereits in weiterführender Forschung genutzt und zitiert. Des Weiteren fließen die Inhalte in universitären Lehrveranstaltungen ein und erweitern die Kompetenzen im Bereich Data Sharing des Forschungsstandorts Deutschland und verbessern deren Standing, da die Ergebnisse die Forschung zur Datenökonomie in Deutschland und Europa fördern und einen wichtigen Beitrag für die anwendungsorientierte Forschung zur Etablierung neuer digitaler Technologien und Wertschöpfungsmethoden in der Industrie leisten. So konnten die Ergebnisse Beiträge zur Forschungsarbeiten zu dezentralen Dateninfrastrukturen leisten, die in diversen öffentlichen Forschungsprojekten im Rahmen des Gaia-X und International Data Space Association Initiativen genutzt werden.

d. Wissenschaftliche und wirtschaftliche Anschlussfähigkeit für eine mögliche notwendige nächste Phase bzw. die nächsten innovatorischen Schritte zur erfolgreichen Umsetzung der Ergebnisse

Die wissenschaftliche und wirtschaftliche Anschlussfähigkeit der Projektergebnisse ist durch die Verwendung der Forschungsarbeiten in deutschland- und EU-weiten Datenraumprojekten gegeben. Die verschiedenen branchenweiten Datenrauminiciativen wie Factory-X, PLATOON; AgriGaia, Green Deal Dataspace, Mobility Data Space, Catena X oder Silicon Economy können die verschiedenen Anreizsysteme für das Data Sharing sowie die Erkenntnisse zu den Hürden und Herausforderungen nutzen, um die Skalierung und Verbreitung von dezentralen Datenintegrationsarchitekturen voranzubringen und somit der deutschen und europäischen Datenstrategie Vorschub zu leisten. Darüber hinaus können Unternehmen die Benchmarks auf Basis der empirischen Daten zur eigenen Verortung nutzen sowie die präsentierten Use Cases als Blaupause für eigene neuartige datengetriebene Wertschöpfungsmethoden verwenden.

III. Angaben zu Arbeiten, die zu keiner Lösung geführt haben

Die inhaltlichen Ziele und Projekthinhalte für die Arbeitspakete konnten im besagten Projektzeitraum erfüllt werden.

IV. Angaben über die Einhaltung der Ausgaben- und der Zeitplanung

Die Ausgaben- und Zeitplanung der Antragstellung wurde eingehalten.

