SCHOOL OF BUSINESS AND SOCIAL SCIENCES
AARHUS UNIVERSITY

Enabling Dataspace Value Chain at Edge for Industry 4.0 and Beyond

Technologies Convergence and Cross-Domain Data Integration over Distributed Edge Architectures

PhD dissertation

Parwinder Singh

Department of Business Development and Technology
Aarhus BSS
Aarhus University

2025

Thesis submitted:	January 31, 2025
PhD supervisor:	Associate Professor Michail J. Beliatis Aarhus University, Denmark
Co-supervisors:	Associate Professor Mirko Presser Professor René C. Goduscheit
PhD committee:	Associate Professor, Francesco Chinello Aarhus University, Denmark
	Professor, Klaus Mößner TU Chemnitz, Germany
	Associate professor, Inigo Flores Ituarte Tampere University, Finland
PhD Series:	Department of Business Development and Technology Aarhus BSS, Aarhus University, Denmark

Dedicated to My Sweet Family

To my wife, Neha Varshney – the unwavering pillar of my strength, my partner in every sense, and my greatest source of encouragement. Your patience, love, and belief in me have guided me through late nights, long days, and countless challenges, lifting me up even when the road seemed most difficult. Your sacrifices, resilience, and steadfast support have been invaluable, allowing me to pursue my dreams confidently. I am endlessly grateful for your strength, understanding, and the warmth you bring to my life. This achievement is as much yours as it is mine, and I dedicate it to you with all my heart.
To my son, Rehaansh Veer Singh – my good luck, cheerful stress buster, and a constant source of joy.
To my mothers, Surinder Kaur & Kusum Lata, whose vision and encouragement have inspired me.
To my fathers, Bir Singh & Dinesh Chandra Gupta, and brother, Sukhwinder Singh, whose wisdom and belief in me have fueled this journey.

I would also like to extend my heartfelt gratitude to my friends Nidhi & Nakul, Michael Lystbæk, and our neighbor, Gitte Kristensen, who stood by us during challenging times with their care and support as a family in Denmark. Their kindness and generosity created countless wonderful memories that we will cherish and smile upon in the years to come.

This accomplishment would not have been possible without your love, unending support, and steadfast confidence that you have shown in me. Thank you for being my foundation and standing beside me every step in accomplishing my longstanding PhD dream.

Preface

The journey of this PhD work has been a transformative exploration into the realm of distributed edge ecosystems and the convergence of digital technologies aimed at revolutionizing cross-domain resources integrated industrial value chains. The fourth industrial revolution, or Industry 4.0, along with technological advancements, has reshaped our understanding of digitalization in industrial ecosystems, and it is in this space that this PhD research finds its significance. This PhD work started with a vision of improving distributed edge operations, bringing digital traceability, transparency, and trust within complex and multi-stakeholder environments by leveraging the combination of modern technologies known collectively as IDEAL (**I**oT, **D**istributed Ledger Technology, **E**dge, **AI**-based **L**earning systems).

The primary research question guiding this work, ***"How can distributed edge architectures and convergence/integration of (IDEAL) technologies be applied to build an Edge-driven value-chain ecosystem in Industry 4.0 and beyond contexts?"*** reflects the ambition to advance Edge-driven architectures that have the capabilities to address the evolving demands of industries. This research explores various technological convergence or resource integration capabilities, challenges, and practical applications through six research articles, culminating in theoretical contributions and developing valuable frameworks for Edge-driven cross-domain value-chain constituting a **Dataspace** oriented ecosystem.

In short, this dissertation, reflecting the PhD research work, addresses technical challenges of digitalization, cross-domain data integration, and distributed architectures, aiming to support the Industry 4.0 ecosystem and as well as lay the foundation for Industry 5.0, which envisions human-centric (i.e., multi-stakeholder), data-driven, trusted, sovereign, adaptable, and sustainable industrial ecosystems. This dissertation comprises six research papers, presented in the order of their completion. An introductory chapter precedes the papers, offering a comprehensive background, motivation, and a cohesive narrative linking each paper to the overall research context and objectives. All papers were developed collaboratively with coauthors, with my role as the primary author and principal contributor. Finally, this dissertation ends with the conclusion chapter, which discusses and summarises all outcomes, contributions, and key takeaways from this research work.

Acknowledgement

In conducting this research, I have had the privilege of collaborating with experts in industry and academia. This collaborative work has enriched the research process, providing invaluable insights and perspectives. My co-authors and friends - Nidhi, Asim, and Michael Lystbaek; mentors and colleagues — particularly my supervisors Michail J. Beliatis, Mirko Presser, and Rene C. Goduscheit — have all contributed in meaningful ways to shape the outcomes of this work. The guidance I received has not only deepened my technical expertise but has also strengthened my understanding of the broader societal and industrial implications of this research.

Special thanks go to Aarhus University's Department of Business Development and Technology, which provided both the academic environment and the support necessary to pursue this research. I am also deeply grateful to Professor Nirvana Meratnia for hosting me during my research stay abroad at Eindhoven University of Technology, Netherlands, guiding me, reviewing my research work, and giving me many useful insights along with support and encouragement.

Parwinder Singh.

Parwinder Singh
Herning, 31st April 2025

Abstract

The digital transformation enabled by Industry 4.0 (I4.0) has redefined industrial processes, driving innovation through connectivity, data-centric operations, and decentralized architectures. However, realizing the full potential of digitalization advancements, particularly within multi-stakeholder industrial ecosystems (e.g., supply chain or manufacturing-related producer and customer-dependent production operations) which are distributed in nature, need a paradigm shift from traditionally centralized to decentralized architectures backed by Edge-driven, interoperable, trusted, and sovereign data operations. Motivated by this pressing need along with the vision to enable a sustainable and collaborative cross-domain data-driven value chain at Edge, this PhD research investigates the role of distributed edge architectures and the convergence of *IDEAL (IoT, Distributed Ledger, Edge, AI-based Machine Learnings)* technologies in developing scalable, secure, and context-aware (at processing, service and data level) adaptable ecosystems for futuristic industrial applications, for instance Industry 5.0.

This dissertation addresses the challenges of distributed edge architectures and the complexities of Cross-domain Data Integration (CDDI) and sovereignty, emphasizing data traceability, heterogeneity harmonization, and semantic modeling. It presents a novel Distributed Edge Network Operations oriented Semantic (DENOS) model that extends the traditional Cloud-Edge-Device (CED) continuum to offer a dynamic and semantic processing contextual model that enables Edge-driven resource pooling of resources (i.e., infrastructure, service, or data) for Cross-domain Resource Integration (CDRI) requirements. In addition, it has also demonstrated *Edge-driven Dataspace* development, compliant with International Data Space (IDS) standards, enabling secure, regional, and cross-domain resource sharing while respecting data ownership, sovereignty, and privacy. This research identifies heterogeneity as a significant barrier at the data, system, and interface levels, highlighting the necessity of meta-standards and harmonization frameworks to bridge disparate systems, service interfaces, and data models to ensure interoperability during CDRI operations.

Another key contribution of this dissertation is introducing an *extended data life cycle model* to support sovereign CDDI operations. Furthermore, a *participatory motivation model* has also been introduced to trigger cooperation among multiple stakeholders or entities in building Edge-driven cross-domain ecosystems, i.e., Dataspace, by participating in it rather than competing, which is essential for the sustainability of future industrial ecosystems. This participation model highlights the need for breaking down data silos to enable CDRI or Dataspace ecosystem to establish "win-win" collaborative business models, which would be crucial for I4.0 and the upcoming human-centric Industrial 5.0 systems.

This dissertation contributes theoretically and practically both by demon-

strating semantic modeling covering technical and business perspectives, enabling modular, adaptable implementations through the DENOS model to reduce development and deployment efforts while delivering significant industrial impact. Additionally, this PhD work has also relaxed the definition of Edge to see it more abstractly to position the Edge as an organizational entity or organizational Edge that provides the required resources to build a CDDI/CDRI ecosystem leveraging IDEAL technologies. Doing this unlocks constrained Edge deadlock to look for new opportunities for decentralized architectures and operations that can be designed and scaled effectively across domains.

This dissertation, leveraging Design Science Research methodology over six research articles, offers theoretical advancements and practical insights into key aspects of digitalization, cross-domain data integration, distributed edge architectures, and the convergence of IDEAL technologies. Thereby laying the foundation for modern data-driven value-chain enablement and sovereign resource sharing within industrial ecosystems.

Resumé

Den digitale transformation muliggjort af Industri 4.0 (I4.0) har redefineret industrielle processer og fremmet innovation gennem konnektivitet, datacentrerede operationer og decentraliserede arkitekturer. For at realisere det fulde potentiale af digitaliseringsfremskridt, især inden for multistakeholder-industrielle økosystemer (f.eks. forsyningskæder eller produktionsrelaterede producent- og kundeafhængige produktionsoperationer), der er naturligt distribuerede, kræves et paradigmeskifte fra traditionelt centraliserede til decentraliserede arkitekturer, understøttet af Edge-drevne, interoperable, pålidelige og suveræne dataoperationer. Motiveret af dette presserende behov og visionen om at muliggøre en bæredygtig og samarbejdsorienteret tværdomæne-datadrevet værdikæde ved Edge, undersøger denne ph.d.-forskning rollen af distribuerede Edge-arkitekturer og konvergensen af *IDEAL (IoT, Distributed Ledger, Edge, AI-baserede Machine Learning (ML) metoder)* teknologier i udviklingen af skalerbare, sikre og kontekstbevidste (på process, service og dataniveau) tilpasningsdygtige økosystemer til Industri 4.0 og fremtidige industrielle applikationer som f.eks. Industri 5.0.

Denne afhandling adresserer udfordringerne ved distribuerede Edge-arkitekturer og kompleksiteterne ved *Cross Domain Data Integration* (CDDI) og suverænitet, med vægt på datasporbarhed, harmonisering af heterogenitet og semantisk modellering. Den præsenterer en ny *Distributed Edge Network Operations-orienteret Semantic* (DENOS) model, der udvider det traditionelle *Cloud-Edge-Device* (CED) kontinuum for at tilbyde en dynamisk og semantisk behandlingskontekstmodel, der muliggør Edge-drevet ressourcepuljering af ressourcer (dvs. infrastruktur, service eller data) til *Cross Domain Resource Integration* (CDRI). Derudover demonstreres udviklingen af *Edge-drevet Dataspace*, der er i overensstemmelse med *International Dataspace* (IDS)-standarder, hvilket muliggør sikker, regional og tværdomæne-ressourcedeling med respekt for dataejerskab, suverænitet og privatliv. Forskningen identificerer heterogenitet på data-, system- og grænsefladeniveau som en væsentlig barriere og fremhæver behovet for meta-standarder og harmoniseringsrammer for at bygge bro mellem forskellige systemer, servicegrænseflader og datamodeller for at sikre interoperabilitet under CDRI-operationer.

En anden vigtig bidrag af denne afhandling er introduktionen af en *udvidet datalivscyklusmodel* til støtte for suveræne CDDI-operationer. Desuden introduceres en *participatory motivation model* for at fremme samarbejde blandt flere interessenter eller enheder i opbygningen af Edge-drevne tværdomæne-økosystemer, dvs. Dataspace, ved at deltage frem for at konkurrere, hvilket er afgørende for fremtidige industrielle økosystemers bæredygtighed. Denne deltagelsesmodel fremhæver nødvendigheden af at bryde datasiloer for at muliggøre CDRI- eller Dataspace-økosystemer og etablere "win-win" samarbejdsbaserede

forretningsmodeller, som vil være afgørende for I4.0 og de kommende menneskecentrerede Industri 5.0-systemer.

Denne afhandling bidrager både teoretisk og praktisk ved at demonstrere semantisk modellering, der dækker tekniske og forretningsmæssige perspektiver, hvilket muliggør modulære, tilpasningsdygtige implementeringer gennem DENOS-modellen for at reducere udviklings- og implementeringsindsatsen, samtidig med at der leveres betydelig industriel effekt. Desuden har denne ph.d.-afhandling udvidet definitionen af Edge til at betragte det mere abstrakt som en organisatorisk enhed eller organisatorisk Edge, der leverer de nødvendige ressourcer til at opbygge et CDDI/CDRI-økosystem ved hjælp af IDEAL-teknologier. Dette åbner for nye muligheder for decentraliserede arkitekturer og operationer, der kan designes og skaleres effektivt på tværs af domæner.

Denne afhandling, der anvender Design Science Research-metoden gennem seks forskningsartikler, bidrager til teoretiske fremskridt og praktiske indsigt i nøgleaspekter af digitalisering, tværdomæne-dataintegration, distribuerede Edge-arkitekturer og konvergensen af IDEAL-teknologier. Dermed lægger den fundamentet for en moderne datadrevet værdikædeaktivering og suveræn ressourcedeling inden for industrielle økosystemer.

Table of Contents

List of Figures

List of Tables

Glossary

A-La-Carte It is an approach for artifact (i.e., services) packaging related to the Dataspace implementation and explained in Paper V under section 7.4.3.. 11, 180

Blockchain A decentralized ledger that records transactions across multiple computers.. 7

Convergence of Technologies The blending of multiple technologies such as IoT, AI, blockchain, and edge computing to create integrated and innovative solutions.. 2

Cross-Domain Pertaining to the integration or interaction of different knowledge areas, industries, or systems to achieve a common goal.. 2

Dataspace A conceptual framework or infrastructure that enables the integration, sharing, and interoperability of distributed data across various domains and organizations.. 6

Edge An entity (which could be an organization, system, gateway, or devices), aligned with edge computing principles, having computing capabilities located close to the data source, capable of processing IoT data and hosting AI and Blockchain workloads. . 1–12, 14–18, 175, 183

Industry 4.0 The fourth industrial revolution characterized by smart manufacturing, cyber-physical systems, IoT, and AI-driven processes.. 5, 173

Semantic Modelling The process of creating a structured representation of data that defines its meaning, relationships, and context within a given domain.. 16

Value Chain A series of activities or processes that add value to a product or service from conception to delivery to the end customer.. 4

Acronyms

6G Sixth Generation Mobile Network. 18

AI Artificial Intelligence. 1–3, 7, 173

AIOTI Alliance for AI, IoT and Edge Continuum Innovation. 14, 16

BPMN Business Process Modelling Notations. 9, 11

CDDI Cross Domain Data Integration. 6, 15, 18, 19

CDRI Cross Domain Resource Integration. 3, 5, 6, 15, 183

CED Cloud-Edge-Device. 3, 4, 6, 8, 11, 12, 16, 176

DENOS Distributed Edge Network Operations-oriented Semantic. 4, 8, 11, 12, 16, 19

DLT Distributed Ledger Technology. 2, 3, 7, 11, 14, 173

DSR Design Science Research. 9, 11

I4.0 Industry 4.0. 1, 3–7, 9, 14, 19

I5.0 Industry 5.0. 1, 9, 15, 17, 18

IDEAL **I**nternet of Things, **D**istributed Ledger Technology, **E**dge Computing and **A**rtificial Intelligence-based Machine **L**earning. 1, 5, 16, 173

IDS International Data Space. 8, 9, 11, 12, 18, 19

IoT Internet of Things. 1–3, 6, 7, 10, 14, 15, 17, 173

RO Research Objective. 5, 6, 9, 12, 176

RQ Research Question. 5–7, 9, 173

SLR Systematic Literature Review. 9, 11

SRQ Sub Research Question. 5, 6, 9, 173

WP Work Package. 6–9, 12

Chapter 1

Introduction

1.1 Background and Motivation

The fourth industrial revolution has brought a modern era of digitalization, where industries are undergoing a paradigm shift towards Industry 4.0 (I4.0) and beyond [1]. The demand for intelligent machines, autonomous devices, and distributed data architectures is driving this paradigm shift with the **I**nternet of Things, **D**istributed Ledger Technology, **E**dge Computing and **A**rtificial Intelligence-based Machine **L**earning (IDEAL) technologies and playing crucial roles in digital transformation [2], [3]. As industries continue to adopt I4.0, it becomes increasingly evident that traditional business models are still struggling with the required digitalization [4], [5]. However, this transformation is crucial for industrial growth and sustainability in the long run. Therefore, the role of different technologies is significantly increased for businesses to achieve sustainable development by bringing digitalization aligning with the vision of I4.0 and beyond, i.e., Industry 5.0 (I5.0) [6], [7]. For example, emerging SMEs rely on the Internet of Things (IoT) ecosystem with Artificial Intelligence (AI) capabilities for real-time data analytics [2], and mature industries in I4.0 are adopting distributed edge architectures to benefit from computing closer to the device layer [8], [9]. Therefore, the need for convergence of technologies and Edge paradigm is growing, where diverse technologies can be leveraged to create values over distributed resources across multiple stakeholders and cross-domain ecosystems [10], [11]. In this context, the concept of Edge-driven value-chain ecosystems has emerged as a promising approach in I4.0.

Edge driven value-chain ecosystems refer to networks of interconnected Edge entities (devices, gateways, or systems running at the edge of the network with proximity to data sources) or stakeholders that desire to create values by leveraging the potential of distributed edge computing resources, IoT, AI, and other emerging technologies [12], [3]. These ecosystems are characterized by decentralized architectures, where data is processed and analyzed at the edge, enabling real-time decision-making and improved efficiency [8]. Therefore, distributed edge architectures have emerged as a potential solution for resource-constrained industrial environments, supporting domain-specific decision-making by moving computation and storage closer to data sources [13]. In addition, edge computing complements cloud solutions, aligning well with I4.0 goals of low latency, scalability, increased security, expanded interoperability, reduced storage costs, and improved connectivity for resilient communication [13], [14].

However, distributed architectures focusing on edge computing present certain challenges, such as resource constraints, trust, privacy, discovery, monitoring, and fault tolerance [15], [16]. Recent technological advancements like IoT can address real-time sensing and data processing requirements; AI can bring intelligent insights into the collected data for impactful decision-making, and Distributed Ledger Technology (DLT) can address some of the relevant security, privacy, and trust issues through tamper-proof encryption, integrity, transparency, immutability, and auditable functionality [17]–[19]. However, hosting DLT and AI workloads further complicates the situation for an Edge already dealing with constrained resource limitations [20].

These limitations of Edge-driven architectures indicate a pressing need for architectural extensions that incorporate semantic resource discovery and pooling, support distributed security, harmonize heterogeneity, and establish sovereign, governed, and trusted frameworks [11]. Such architectural advancements would enable reusability by sharing resources, integrated capabilities, and interoperable functional flows across Edge-assisted and cross-domain operations. Therefore, this research has focused on investigating the theoretical backgrounds and extending the traditional distributed edge architectural paradigms, as well as exploiting the convergence capabilities of IDEAL technologies and analyzing what, where, and how this technological convergence can be applied in real-world perspective and thereby bring some values to the industry. These values include novel methods of distributed security features, digital traceability of events, seamless data sharing, and service chaining of resources, systems, and services via Dataspace enablement at the Edge [21], [22]. This will, in turn, unlock new cross-domain integrated architectural ecosystem-driven business constellations wherein actors can engage directly at the Edge to get the benefits of transformed business ecosystems [10], [9], [23].

1.2 Identified Research Gaps and Challenges

Different challenges and gaps have been identified while investigating different studies for this PhD research. This has already been listed as part of the published papers, which are listed in Table 1.2. The overview of relevant research gaps and challenges is given below:

1. **Convergence of Technologies at Edge:** The main research question revolves around how to converge or integrate IDEAL technologies to solve real-world challenges and what would be the benefit of that. The technologies meaning here should be interpreted as services developed using specific technology-driven concepts and tools. For example, IoT can generate data through environmental sensing capabilities and enable real-time data processing. Similarly, Edge can be referred to as available computing closer to the data sources. AI can be used to develop analytical tasks or services, and DLT can be used for secure, immutable, and tamper-proof data events. So, when referred to as the convergence of technologies, it means running the technological services in the same computing namespace (representing the same system or domain). The term integration means running services in different computing namespaces, and that is where Cross-Domain comes into the picture. As the focus of the research is on

the IDEAL set of technologies, the Edge appears to be the default choice as it provides computing infrastructure for other technologies-related tasks to run, i.e., IoT, DLT and AI workloads. However, the problem is that traditionally, the Edge is perceived as a device that only has constrained resources. So, naturally, a question arises here: How can Edge be used with resource constraint limitations to host the technologies that demand computing resources? This perspective creates a deadlock for the Edge to evolve further. There are two possible approaches here: 1) to extend or optimize the utilization of Edge computing capabilities or 2) to extend the horizon or definition of Edge. For this PhD research, the latter approach has been chosen. As a lot of research is already going on in the first approach, and many areas (e.g., heterogeneity, integration, or interoperability) of the first approach need different research perspectives through the second approach. In addition, the second approach provides many benefits and opportunities, like cross-domain integration and Edge-driven value chain development.

Therefore, in this PhD work, the Edge layer has been perceived more abstractly (or relaxed) for the first time in literature. For example, the articles of this PhD dissertation discussed Edge as the organizational Edge (as factory unit, system, gateway, or device) having sufficient computing capabilities to host IoT, AI, and DLT workloads. This perspective actually also fulfills the definition of Edge, offering computing closer to the data source. So when an Edge is perceived as an organizational Edge or vice versa, lots of opportunities start emerging, such as creating value chains between the organizational Edge(s) or focusing on the Edge as a domain or Edge domain that enables cross-domain data, services, or system integration pathways. Then, naturally, opportunities for Cross Domain Resource Integration (CDRI) can be foreseen. A Resource here can be defined as an object associated with infrastructure, service, or data entities. However, developing these CDRI-based value chains presents another challenge: How to control this cross-domain sharing of resources sovereignly? This brought the Dataspace into the research context to provide trust and sovereign control at the design level. The relevant gap of considering Edge as an organizational Edge or see Edge in an abstract manner, to build Edge driven cross-domain value chain, has been addressed in Paper IV, Paper V and Paper VI.

2. **Distributed Architecture** - I4.0 typically expects to follow the Cloud-Edge-Device (CED) continuum. The CED continuum historically comprises Cloud, Edge, and Device layers. However, it can be envisioned as an ecosystem of multiple domains with multiple stakeholders involved at each layer. This means all classical challenges, such as heterogeneity, interoperability, distributed security and trust, and resources management, also apply here. On top of that, considering modern-day requirements, there is a need for opportunities for cross-domain data or resource integration, but there are also requirements to deal with the classical challenges.

Therefore, to address these challenges, CED continuum needs to be extended with additional capabilities leveraging the combined strength of IDEAL technologies. These challenges have been identified and addressed based on a systematic literature review in Paper IV.

3. **Heterogeneity at Multiple Levels:** Heterogeneity has always been an issue for decades whenever two entities or points of integration are involved. In this research work, it is observed that this classical heterogeneity issue is also applicable here. There is heterogeneity in almost all aspects of distributed architecture involved. Heterogeneity is always the dominating factor when there is a desire to integrate diverse technological services associated with diverse systems, interfaces, and data models. In addition, when Edge-driven cross-domain (multi-stakeholder environment) resource integration is required, this issue appears to be the major blocker. Therefore, to solve heterogeneity, there is a need to identify the factors behind it and explore a relevant heterogeneity harmonization approach, which has been discussed in Paper III.

4. **Diverse Use Case Requirements:** From the I4.0 perspective, different use cases have been explored in this PhD work, one from the wind energy domain and the other from the manufacturing domain. The context of IDEAL technological capabilities remains the same irrespective of the application domain. However, their implementation keeps changing depending on the target use case requirements for the technological context. This demands effort, cost and time. Therefore, this raises a fundamental question: Can we create a distributed edge architectural framework that can be implemented dynamically or have a dynamic implementation context to fulfill the diverse use case requirements? This is exactly being addressed through the proposed Distributed Edge Network Operations-oriented Semantic (DENOS) model in Paper IV, which offers a dynamic implementation context through a semantic processing model.

5. **Distributed Security and Sovereignty** - Distributed edge architectures have been found to be vulnerable on the security front because of their distributed nature. In a certain domain of operations, e.g., a manufacturing unit/factory, control (meaning the devices themselves) and data (only device/machine-specific data) plane-related operations are involved. On top of that, when Edge-to-Edge cross-domain integration is perceived, the ownership of data flowing from one domain to another becomes paramount. So then the question arises of how to safeguard the interest of the data owner under cross-domain data integration flow. For this purpose, extension of the data life cycle to include integration and reuse phases has been explained in Paper III, and the use of IDS standards-based negotiation, sovereignty, and identity-based access management to share data in cross-domain has been presented in Paper VI.

6. **Value Chain and Resources in Silos**: In Paper II and Paper III, it has been discussed that the resources (data or services) across domains remain

in silos and are incapable of utilizing to their full potential due to the missing CDRI architectures. These silos can be broken by leveraging diverse technologies, including IDEAL, to establish Dataspace-driven value chains based on distributed, Edge-driven, and cross-domain resources integration. These value chains can be designed and developed over cross-domains by sharing common interests or assets of a specific domain, such as reusable data, services, or systems. The systematic literature review in Paper IV and other papers has revealed that very limited knowledge is available in this area. Consequently, this PhD work aims to provide a foundational framework for advancing the knowledge of cross-domain resource-integrated value chains.

1.3 Research Question and Objectives

Following the above motivation and to address the relevant challenges, this research work has set the following primary research question:

Research Question (RQ): How can distributed edge architectures and convergence/integration of IDEAL technologies be applied to build an edge-driven value-chain ecosystem in Industry 4.0 and beyond contexts?

Following this RQ, the Sub Research Questions (SRQs) and corresponding Research Objectives (ROs) for this PhD research have been laid down, which are given as follows:

- **SRQ1:** What are the key challenges and opportunities in integrating and converging IDEAL technologies for I4.0 applications?

  RO1: To address the SRQ1, the first research objective investigates the integration and convergence of IDEAL technologies for relevant challenges and value addition in relevant I4.0 applications or use case contexts.

- **SRQ2:** What architecture needs to be designed to secure interoperable and traceable operations for heterogeneous domains or data sources in multistakeholder I4.0 environments?

  RO2: For SRQ2, the second research objective focused on exploring distributed, interoperable, integrated, traceable, and immutable secure Edge operations for heterogeneous devices or data sources, cross-domain services, and infrastructural resources in a multi-stakeholder environment.

- **SRQ3:** What theoretical advancements are necessary in the traditional computing paradigm to support cross-domain value chains in I4.0 and beyond?

RO3: As per the requirement of the SRQ3, extensions for the theoretical groundings for the traditional CED paradigm and the cross-domain value chain development design need to be investigated under this objective.

- **SRQ4:** How can we address data sovereignty, sharing, and heterogeneity challenges for distributed operations in cross-domain?

  RO4: To address SRQ4, this research objective investigates the design for a cross-domain and decentralized data integration framework for data-sharing that follows data sovereignty, security, and heterogeneity harmonization principles and leverages IDEAL strengths to enhance real-time intelligent and interoperable operations along with trust, traceability, and data governance at the Edge.

- **SRQ5**: How can the overall research context be validated to demonstrate cross-domain value chain development among different stakeholders?

  RO5: To fulfill the SRQ5, this research objective validates the research context through practical case studies, i.e., examining proposed models, convergence or integration of technologies at the Edge, and Cross Domain Data Integration (CDDI) to facilitate relevant value chains among different stakeholders in real-world scenarios.

Scope: This PhD research investigates several focal areas: exploring literature in a systematic way for identifying challenges and opportunities related to the RQ and ROs, extending theoretical groundings for CED continuum, leveraging diverse technologies for solving different real-world problems, such as blockchain for secure traceability in the supply chain, implementing IoT and edge computing to enhance SMEs' operational efficiency, and enabling CDRI at Edge through the Dataspace concept. These focal points are structured to contribute toward a broader vision of I4.0 and 5.0, where interoperable, secure, and scalable systems drive decision-making across industries. In doing so, the research aligns with the overarching goal of making architectural and technological advancements accessible and beneficial to diverse stakeholders, including large enterprises and SMEs alike.

1.3.1 Work Packages and Vision Alignment

This PhD work is structured into four Work Packages (WPs), each addressing specific academic and/or industry challenges. By addressing these challenges, the research contributes to practical insights and theoretical groundings that will advance the ongoing evolution toward distributed edge architectures for cross-domain value-chain-enabled industrial ecosystem systems in the future.

1. **WP 1 – Theoretical foundation and literature review on the convergence of IDEAL technologies for Distributed Edge in I4.0**

This WP establishes a comprehensive foundation for the RQ through a systematic literature review and initial testbed development to understand the convergence of IDEAL technologies for digitizing industrial processes under distributed architectures, especially for I4.0. The reflections of this WP can be seen in all papers where the literature review findings in narrative contexts have been done. For example, Paper I has practically explored the industrial use case related to the wind industry supply chain, where the use of DLT, IoT, and Edge has been applied to solve the digital traceability of events in real-time and building secure and trusted ecosystem in a distributed and multi-stakeholder environment. Similarly, distinct literature reviews have been done continuously in different papers to identify relevant gaps and solve specific challenges. Paper IV presents the systematic literature review summarizing our findings on technology convergence, detailing key challenges, and proposing an IDEAL convergence capabilities framework. The literature review has focused on how each technology (IoT, Edge, AI, DLT, etc.) is individually and collectively used and identifies relevant challenges, with particular emphasis on their role in distributed edge architectures.

2. **WP 2 – Validating research context through impactful use cases of industrial contexts** The objective of the WP is to present the RQ's relevance practically by delving into real-world use cases. Therefore, three industrial use cases are chosen to validate IDEAL convergence in practical industry applications, especially within wind energy, manufacturing, and cross-domain data integration contexts. These use cases have been presented in Paper I, Paper II, Paper V, and Paper VI. Paper I presents a Blockchain application in a Wind Industry supply chain demo case where it explores the specific application of blockchain within the wind energy sector, demonstrating the IoT and Blockchain convergence at the Edge that together enables digital traceability for multi-stakeholders environment in I4.0. Paper II presents a case study on Roll-to-Roll Printing Operations in the manufacturing domain and represents the need for business modeling, innovation, and cross-domain data integration ecosystem, i.e., Industrial Dataspace, showing a possible technological impact on the case company's production efficiency, sustainability and business growth in the long run. Paper V and Paper VI demonstrate the cross-domain data integration scenarios through Dataspace applications covering general and standard perspectives based on the wind turbine use case presented in Paper I. These use cases validate the significance of distributed architectures and IDEAL technologies at the Edge and in an Edge-driven value chain through data sharing, digital traceability, and integration frameworks in real-world industrial settings. In short, they demonstrate the business and technical impacts and provide empirical evidence for our research context. Please check the summary section 1.6 or dive directly into the relevant chapter for more details.

3. **WP 3 – Investigation and development of Edge-driven distributed**

architectures that support cross-domain resource integration in diverse contexts leveraging IDEAL capabilities. This WP focuses on investigating distributed edge architectures from the theoretical and practical point of view. Therefore, *DENOS* is designed and validated to extend the traditional CED paradigm and supports cross-domain resources integration, facilitating semantic adaptation at the Edge. The main research outcome for this WP, which explains the proposed DENOS model, is presented in Paper IV. In addition, Paper V also presents an edge-enabled context-aware Datspace model that supports this WP.

4. **WP 4 – Decentralized, distributed and cross-domain resources and operations management for data sharing with security, trust and sovereignty to build Edge-driven value chains**: This WP focuses on decentralized operations management by identifying the requirement for heterogeneity harmonization in cross-domain integration of resources to support interoperable operations. This is addressed in Paper III, which introduces the extended data life cycle and participatory motivation model useful in building cross-domain data integration ecosystems (i.e. Dataspace) and demonstrates different levels of integration possibilities when we perceive Edge-to-Edge integration for developing Edge-driven Dataspace wherein the Edge acts as an organizational Edge as explained in Paper III and Paper V. In addition, this WP brings trusted, traceable, and governance methods during cross-domain data sharing to safeguard the interests of the data owner. This has been achieved through blockchain methods in Paper I, Paper IV, and International Data Space (IDS) standards-compliant data sharing with sovereign controls in Paper VI. This shows how an Edge-driven ecosystem can be developed to support secure and reliable cross-entity or cross-domain operations. This WP is useful for developing decentralized frameworks for data sharing and secure data management in multi-stakeholder environments with standard-compliant methods that address key challenges such as digital traceability, trust, security, and data sovereignty.

In summary, each work package contributes toward building a distributed, scalable, and Edge-driven cross-domain value-chain ecosystem. For instance, WP1 establishes the foundational literature and testbed for IDEAL convergence, setting the stage for further research. WP2 explores the relevant use cases in the industry and demonstrates the use of different technologies together and validates the research context in real-world use cases, assessing their business and technical viability. WP3 made a deep dive into the literature to extend the theoretical grounds for the CED paradigm to address the relevant challenges of distributed edge architectures, solving complexities of cross-domain data integration, and developing dynamic implementation contexts suiting the diverse needs of different use cases. Finally, WP4, leveraging WP3 and WP2 groundwork, tried to address the heterogeneity and interoperability challenges for cross-domain data integration to enable a data-driven value chain across an industrial ecosystem that operates in decentralized and distributed environments. In addition, WP4 also demonstrates how data can be shared securely and in a sovereign manner following IDS standards, principles and semantic models, supporting trust and traceability of events.

Through WPs, this PhD research contributes to developing a foundation for cross-domain value chains on distributed edge architectures that industries can tailor to evolving demands. This ensures both immediate relevance and long-term sustainability in I5.0, which focuses on human-centric and adaptable systems.

1.4 Methodology and Research Design

The primary research approach for this PhD research is illustrated in Figure 1.1 and is based on the Design Science Research (DSR) methodology [24]. DSR is chosen because it fulfills the design objective of this research, which includes the design process, i.e., it simultaneously generates knowledge about the method used to design an artifact and the design or the artifact itself [25], [26]. Using the DSR as an overarching approach for this PhD, solutions or artifacts have been designed and developed to address specific real-world problems or challenges, serving as the foundation for theoretical knowledge extensions. However, each of the individual research articles, as shown in Table 1.1, followed a suitable research methodology tailored to the specific requirements of the respective article. This includes Action Research focusing on architecture and prototype for Paper I, Business Process Modelling Notations (BPMN) with Architectural Design for Paper II, DSR for Paper III, Systematic Literature Review (SLR) with prototype and experimentation for Paper IV, Onion Architectural Design for Paper V, and Constructive Research for Paper VI, considering the nature of the investigation and research context.

Relevant use cases focusing on I4.0 have been identified following RQ, SRQs and ROs. This includes three use cases, namely - the wind turbine supply chain, label printing, and Edge assisted Dataspace use cases. The first one related to the local wind turbine industry is presented in Paper I and then reused in other Paper IV, Paper V and Paper VI to exploit the real world environment under different research contexts. The second use case concerns the automation and digitalization of label printing operations and processes in an SME case company and is presented in Paper II . In addition, the literature review is also continuously carried out to build theoretical groundings for each WP and relevant use case. The final use case covers the real-world Edge-driven Dataspace enablement scenarios from academic and industrial point of view, and this has been used in Paper III, Paper V and Paper VI.

Most phases of research design, shown in Figure 1.1, run in parallel. The implementation phase provides the prototype development and validation for the use case scenarios and proposed architectural model, as well as offers a controlled environment to experiment with various WPs related to RQ. There are multiple prototypes (close to the real world) developed, which include Paper I, Paper IV,

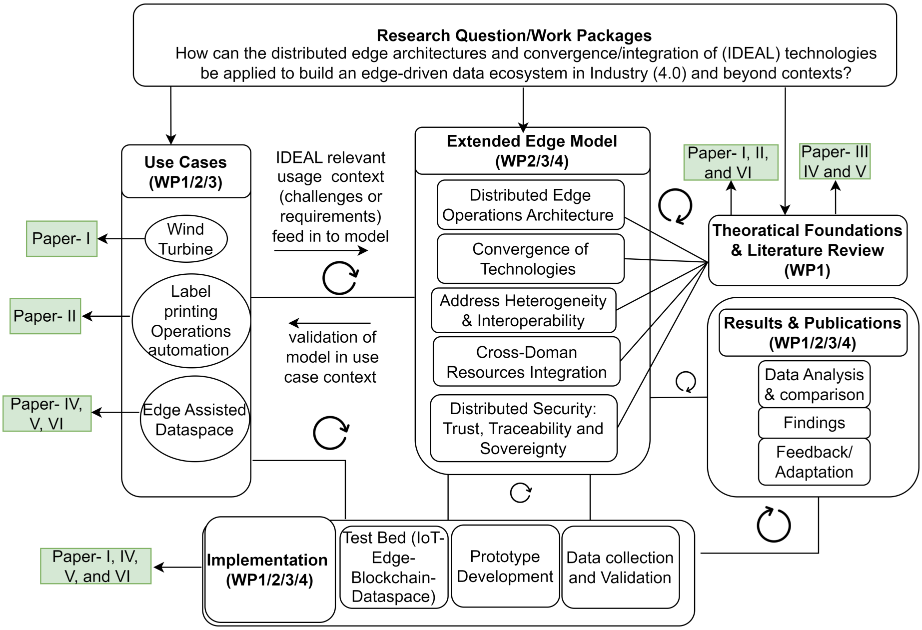


Figure 1.1: PhD Research Design

Paper V, and Paper VI. Details of the prototype implementation have already been given in the relevant paper. Consequently, relevant experimental data has also been collected for concept validation and research analysis. In addition, diverse data sources have been used for this research, as shown in Table 1.1. Data from the systematic and narrative literature review and experimental data (from experiments) are involved. In addition, empirical data and observations are collected by conducting workshops and discussions with the relevant stakeholders regarding different use cases. The analysis of empirical data is used to generate the high-level requirements for defining the relevant technical system architecture. Eventually, all the research development and analysis findings are published as research papers, which have been described as outcomes in the next section.

1.5 Research Outcome

This PhD dissertation has included six main research papers, each addressing a distinct industry/academic challenge and generating relevant outcomes, as shown in Table 1.1. The Paper I explores blockchain's role in improving the traceability and transparency of commodity components within the wind industry and also demonstrates how diverse technologies like IoT, Blockchain, and Edge can be converged to solve industry-specific problems of digital traceability and trusted operations in a multi-stakeholder ecosystem. The Paper II investigates digital solutions to enhance error management, digitalization, and Dataspace-driven cross-domain data integration in a label printing manufacturing SME. The

Table 1.1: Summary of Data, Methodology, Challenges, and Outcomes of Different Articles

Paper	Data Type-/Methodology	Challenges	Outcome
Paper I	Empirical, experimental/ Action research	Implementation of digital traceability for wind supply chain events	• Industrial use case • DLT prototype over distributed architecture • Emergence of technologies convergence
Paper II	Empirical/ BPMN and Architectural	Digitalization and data points augment in SME operating with outdated technological machinery	• Industrial use case and business process maps • Conceptualized Dataspace architecture and Edge abstraction idea
Paper III	Simulation, literature/ DSR	Cross-Domain Data Integration Motivation, and Heterogeneity	• Extended data life cycle • Participatory motivation model • Harmonized integration model • Identify needs of Meta Standards
Paper IV	Literature, empirical, experimental/ SLR	• Traditional CED limitations or challenges • Adoption of modern technologies at Edge • Dataspace enablement	• Identification of challenges for Edge, Dataspace, and technologies convergence • DENOS Model • Edge-Blockchain integration method
Paper V	Empirical, experimental, literature/ Onion architectural	Conceptualizing Dataspace at Edge for cross-domain value chain enablement	• Dataspace architecture and prototype • A-La-Carte approach for distributed microservices deployment
Paper VI	Literature, experimental/ Constructive research	Adoption of International Data Space Standards in local context	IDS enabled Dataspace Prototype for regional industrial context

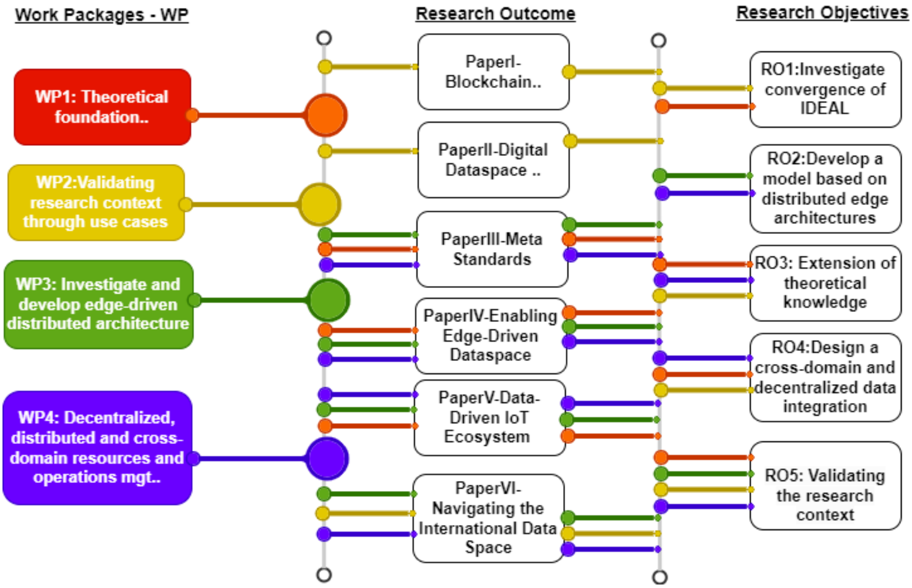


Figure 1.2: Research Mapping

Paper III explores how distributed architectures in a cross-domain ecosystem can be developed at the Edge, what challenges it needs to resolve, and how the relevant stakeholders can be motivated to build such an ecosystem. Therefore, considering the insights and emerged requirements of the previous three papers, Paper IV has developed a systematic literature review followed by a conceptual model, called DENOS, for CED continuum extension to address relevant classical challenges and facilitate modern Edge-driven cross-domain data integration (i.e., Dataspace) operations, and leveraging the convergence of IDEAL technologies. Finally, Paper V and Paper VI demonstrate the real-world scenario of cross-domain data sharing across organizational boundaries using generalized Dataspace concepts and IDS standards-based compliances covering data sovereignty, security, and negotiation aspects among different stakeholders.

Overall, this PhD work has yielded six research articles (one Journal and five conferences) and one book chapter as a major contributor or first author. In addition, two journal article publications as co-authors are also in progress. The overall research outcome during the entire PhD duration has been shown in Table 1.2. However, for this dissertation, only six research papers (highlighted ones) are included, which have been explained in the next chapter covering Paper I to Paper VI, and their summary has been presented in the next section. Figure 1.2 shows the relevant research process mapping to establish relationships or interconnections among WPs, ROs, and research papers to map the overall PhD research context.

Table 1.2: PhD Research Output

Article	Reference
Paper I	Singh, P., Holm, K., Beliatis, M. J., Ionita, A., Presser, M., Prinz, W., & Goduscheit, R. C. (2022). Blockchain for Economy of Scale in Wind Industry: A Demo Case. In Lecture Notes in Computer Science (LNCS, vol. 13533). Springer.
Paper II	Singh, P., Nidhi, J. K., Beliatis, M. J., & Presser, M. (2023). Digital Dataspace and Business Ecosystem Growth for Industrial Roll-to-Roll Label Printing Manufacturing: A Case Study. In SENSORCOMM 2023, The Seventeenth International Conference on Sensor Technologies and Applications. Porto, Portugal, September 25-29, 2023.
Paper III	Singh, P., Nidhi, A. U. H., & Beliatis, M. J. (2023). Meta Standard Requirements for Harmonizing Dataspace Integration at the Edge. In 2023 IEEE Conference on Standards for Communications and Networking (CSCN).
Paper IV	Singh, P., Beliatis, M. J., & Presser, M. (2024). Enabling Edge-Driven Dataspace Integration Through Convergence of Distributed Technologies. Internet of Things, 101087. Elsevier, January 2024.
Paper V	Singh, P., Nidhi, Beliatis, M. J., & Presser, M. (2024). Data-Driven IoT Ecosystem for Cross Business Growth: An Inspiration Future Internet Model with Dataspace at the Edge. Internet2024 Conference - IARIA, Athens, Greece, March 10 - 14, 2024.
Paper VI	Singh, P., Meratnia, N., Beliatis, M. J., & Presser, M. (2024). Navigating the International Data Space To Build Edge-Driven Cross-Domain Dataspace Ecosystem. Global IoT and Edge Computing Summit, GIECS 2024. Communications in Computer and Information Science, vol 2328. Springer, Cham. Brussels, Belgium, September 24, 2024.
Article-In-Progress	P. Singh, M. Gaspari, N. Meratnia, and M. Presser. AI-Driven Harmonized Data Ingestion: A Vision for Seamless Cross-Domain Data Integration. IEEE Transactions on Knowledge and Data Engineering, [Status: To be Submitted].
Book Chapter	Singh P, Beliatis M, Emilie Mathilde J, Presser M. Digital Business IoT Maturity Patterns from EU-IoT Ecosystem. In: Sofia RC, Soldatos J, editors. Shaping the Future of IoT with Edge Intelligence: How Edge Computing Enables the Next Generation of IoT Applications. 2024 Jan. Chapter 18. PMID: 38564562.
Article-In-Progress	Nidhi, Singh, P., Tsironis, G., Tsagarakis, K., Beliatis, M. J. (2025). Footprint of FiberGlass Enterprises for Embracing Circular Economy and Dataspace Enabled Digital Product Passport: A Study in European Baltic Countries. Journal of Circular Economy. [Status: To Be Submitted]

1.6 Summary of Papers

1.6.1 Paper I

Challenge: *Develop a Blockchain-based solution to enhance traceability, transparency, and efficiency in the wind industry's multi-stakeholder supply chain.*
Summary: This paper explores the use of blockchain technology to enhance the traceability and transparency of supply chain events within the wind industry, focusing on commodity components like bolts and fasteners. A prototype demonstrator in a lab environment is developed through qualitative feedback from industry experts. This study maps the supply chain events associated with turbine components belonging to multiple stakeholders. This study has also explored the convergence of DLT, IoT, and Edge technologies. DLT is validated to achieve immutable data sharing, ownership, confidentiality, and trust among multiple stakeholders. At the same time, IoT has been used to generate real-time data about turbine operations at Edge devices and send data to traditional ERP systems. This study provides a scalable foundation for extending blockchain's application to other commodity components in the wind industry, emphasizing the convergence of DLT, Edge, and IoT as a step towards the main research question of technology integration or convergence at the Edge.

Award Recognition: This paper research work was honored with the *Recognition Alliance for AI, IoT and Edge Continuum Innovation (AIOTI) Award 2023 in SME category* at the *AIOTI Signature, Belgium, 2023* event. The award acknowledges the significance and potential impact of interdisciplinary research that focuses on business and technological domains simultaneously to give an end-to-end holistic industrial solution.

1.6.2 Paper II

Challenge: *Investigate and develop smart error management and operational strategies to boost productivity in a label printing SME*

Summary: This study examines a Danish SME in the label printing and bundles manufacturing domain that seeks digital transformation. It addresses a crucial aspect of I4.0 in our main research question, exploring how digitalization can be introduced within SMEs to improve productivity and optimize operations, especially when they rely on older machinery. The case study highlights a significant gap in existing research and industrial applications; for instance, despite operating for over five decades, this industry lacks detailed process maps, hindering effective technology integration.

The research is structured into two phases. In the first phase, process maps of the production environment are created to establish a foundational understanding of workflow and dependencies. In the second phase, a service-oriented Dataspace architecture is designed for cross-domain data and service-integrated operations. This phased approach provided insights that led us to redefine the *Edge* concept, expanding it to view the organization itself as an *Edge entity.* This research paper identifies real-world challenges while introducing digital solutions to modernize operations over old machinery using technological combinations, i.e., IoT, Edge, Blockchain, and Cloud computing. Through process mapping and proposed service-oriented architecture, the study outlines how these technologies can enhance productivity by resolving errors in machinery, reducing waste and managing legacy machinery, providing a model for SMEs with similar constraints.

In discussions with SME stakeholders, it has been found that *Data*, when shared cross-domain, can serve as a *valuable* asset to develop new business models. For example, it is envisioned that printed label bundles upon transportation from the production unit, i.e., manufacturer end to the customer facility, can be enabled to trace events and manage operations remotely based on the shared services model and information embedded in the bundles to turn them into smart bundles. However, they emphasized the need for data sovereignty in cross-domain operations, ensuring that production owners control their data and associated operations. This realization introduced the concept of Dataspaces as a key solution, enabling cross-organizational data sharing while safeguarding data sovereignty. This led to the development of a distributed architecture model that further reinforced Dataspace as a focal point in our research and to achieve seamless, secure, and sovereign cross-domain integrated operations for SMEs in a modernized digital ecosystem.

Award/Recognition: This study is selected as a potential paper and invited to transform the conference into a journal paper.

1.6.3 Paper III

Challenge: *Develop a framework to motivate organizations to participate in an Edge-driven Dataspace ecosystem, enhancing cross-organizational data sharing and service integration.*

Summary: This paper addresses the potential of Dataspaces in I5.0, where industrial IoT and cyber-physical systems generate vast data volumes. It suggests that organizations should be viewed as *Edges.* This creates opportunities to build value chains by sharing data across boundaries, creating a cross-organizational resource pool that powers shared and context-driven services. This study proposes a *Participatory Motivation* system model for motivating entities to participate in a coordinated Edge-driven Dataspace ecosystem, fostering a value chain based on the mutual gain of sharing and reusing data and services across organizations. Therefore, it introduced a modified *Data Life Cycle* for this purpose and also emphasized the need for harmonization of heterogeneity induced due to diverse standards being used in cross-domain services and systems. This study lays a foundation for further research in Meta Standards, which are needed for CDRI or CDDI, interoperability, and sustainable, semantically integrated ecosystems.

1.6.4 Paper IV

Challenge: *Develop an architectural framework that integrates or converges IDEAL to address heterogeneity, security, scalability, and resource management for diverse use case requirements and Dataspace applications.*

Summary: The exploration for this study began with understanding how different technologies could converge to develop value chains based on common interests like data, services, and systems. The Edge emerged as the natural choice for enabling this convergence due to its proximity to data sources and computing infrastructure. Therefore, this study reviews the literature on the convergence of the IDEAL for supporting Dataspace integration efforts. The study identifies key challenges, including heterogeneity, security, scalability, and resource management. A proposed architectural framework, DENOS, is presented to address these challenges and enable secure, semantic integration of resources and technologies.

The DENOS framework extends the traditional CED continuum, which historically served as a hierarchical computing infrastructure, to address the relevant challenges and modern requirements for decentralized systems. The CED continuum's classical challenges, including heterogeneity, interoperability, distributed security, and efficient resource management, are coupled with emerging opportunities like cross-domain data sharing and resource integration. DENOS meets these needs by introducing three additional layers - Semantics, Convergence, and Dataspace Integration - into the CED continuum, providing a robust architectural framework to overcome both technical and business obstacles through dynamic realization. The semantic layer offers Semantic Modelling, which enables dynamic implementation context to diverse use cases with contextual modeling contexts, namely Processing Context (at the infrastructure level), Service Context (at the service level), and Data Context (at the data level). The convergence layer provides execution of these contexts to provide dynamic implementation with optimized resource utilization capabilities at the Edge. These dynamically executed contexts ensure flexibility to meet varied operational needs from diverse use case needs. The Dataspace Integration layer under the DENOS framework supports seamless cross-domain integration at multiple levels (at processing, service, or data level), building cross-domain value chains by facilitating resource convergence across infrastructure, service, and application layers, aligning technical operations with strategic business goals.

Award/Recognition: This research work has won the *AIOTI Award 2024 in Research Category* at the *AIOTI Days Award Event, Belgium, 2024*. The award acknowledges the novelty, significance, and potential impact of the proposed research contributions that extend the traditional distributed edge architectural continuum.

1.6.5 Paper V

Challenge: *Develop an Edge-enabled and context-aware Dataspace model to facilitate secure, collaborative cross-domain data integration and service sharing for local data-driven value chains.*

Summary: This paper, by leveraging the insights of Paper II, Paper III, and Paper IV, presents a novel *Edge-enabled, context-aware Dataspace model* aimed at facilitating cross-domain and cross-organization data integration and service sharing in local or regional contexts. It highlights the critical role of data as a driving force for business growth, innovation, and sustainability in competitive markets. It enables businesses to share and reuse data and services in a secure and governed manner through the proposed DS model, which fosters collaborative and innovative ecosystems, aligning with the broader vision of I5.0 and beyond. Rather than encouraging competition, the model promotes a cooperative and win-win approach where organizations leverage shared resources to build innovative services and business models, creating value chains that benefit all stakeholders. The proposed DS model considers the Edge layer as a critical component. Traditionally, edge computing focuses on processing data near the source to reduce latency and save bandwidth, but it has always been considered as a constrained Edge with limited computing capabilities. However, this study has also broadened the Edge role, aligning with the earlier studies. Each participating organization, system, or entity has been considered to act as an Edge, contributing computing resources to develop the Edge-driven DS ecosystem. This Edge-enabled approach allows for creating localized DS ecosystems that can operate efficiently and securely, meeting the needs of stakeholders across various domains.

The proposed model suggests including semantic modeling capabilities to enable context-aware data operations, which are essential for handling modern industries' diverse data integration and reusability needs. By employing semantic adaptation and context-aware data lakes, the model can deliver shared data and services aligned with different domains' specific requirements. For example, the sharing organization's data usage rules and governance protocols are respected by receiving organizations, ensuring data sovereignty and compliance. This approach empowers businesses to seamlessly integrate resources while maintaining control over their data, thus creating robust cross-domain collaboration opportunities.

To validate the proposed DS model, the study employed a local IoT Edge-cluster testbed and prototyped the DS in a controlled lab environment. A wind turbine use case was used to demonstrate the model's applicability, showcasing how the Edge-oriented DS infrastructure could support integration and governance requirements in real-world scenarios. By orchestrating the Edge resources using predefined service catalog artifacts leveraging the extended A-La-Carte (ALC) approach, the model effectively met the identified functional requirements. These validations not only demonstrate the technical feasibility of the DS model but also underline its practical relevance in addressing modern data integration challenges.

The paper also provides insights into how DS ecosystems can be realized on a smaller scale in local or regional contexts by leveraging Edge resources. By combining semantic modeling, smart governance, and context-aware capabilities, the DS model enhances the efficiency of cross-domain data management operations and value chain development. It simplifies the integration of diverse data sources and services, enabling businesses to unlock new opportunities for collaboration and innovation. Furthermore, the study identifies the broader implications of the proposed DS model for the evolving I5.0 landscape and the future internet ecosystem. By incorporating advanced semantic capabilities, the model aligns with the anticipated demands of emerging technologies such as Sixth Generation Mobile Network (6G), which will require efficient and scalable solutions for managing vast amounts of cross-domain data. This positions the DS model as a forward-thinking solution capable of meeting the needs of increasingly interconnected and data-driven industries.

Award/Recognition: This paper is honored with the *Best Paper Award* at the *INTERNET 2024* conference at Athens, Greece, March 2024.

1.6.6 Paper VI

Challenge: *Investigate Dataspace-oriented standards, specifications, and reference implementations to build a local or regional Dataspace for Edge-driven cross-domain data integration.*

Summary: This study achieves the same objective as reflected in Paper V but this time by fulfilling the compliances of EU-driven IDS standards. This paper explores data as a pivotal asset in the modern industrial era that drives intelligence and enables informed decision-making across the business, technical, economic, and social domains through advanced machine learning techniques and methods. Therefore, data generated in one domain often finds its utility in others, necessitating *CDDI* and sharing across Edges. For successful CDDI and sharing, maintaining *data sovereignty* is critical to govern usage contexts. Such governance aligns with the European Union's *Dataspace* vision, which aims to create a unified, data-driven European market through frameworks like the *IDS*. IDS, supported by the European Commission, provides essential standards and specifications to streamline CDDI activities across industries. However, despite the huge volume of information available on the IDS ecosystem, it is found to be fragmented and difficult to navigate, hindering its adoption and practical implementation. Therefore, this study addresses that gap by consolidating theoretical and practical insights into building IDS-based, Edge-driven Dataspaces. By focusing on a real-world use case in the *wind industry supply chain*, the study demonstrates the applicability of CDDI through a locally developed IDS platform. This practical validation not only showcases the potential of IDS-driven CDDI for Edge domains but also highlights its transformative value in enabling efficient and sovereign cross-domain data sharing within the European market and local industrial environments.

Award/Recognition: This research work has contributed to fixing issues in the main code repository for the IDS community.

1.7 Significance of this PhD Research

This PhD research has investigated the distributed edge architectures that extend traditional distributed computing paradigms to accommodate dynamic implementation contexts for diverse use case needs and modern data integration requirements through the convergence of IDEAL technological and CDDI-based Dataspace capabilities. As a result, this research has introduced a context-aware Dataspace and IDEAL convergence DENOS model to meet the diverse needs of industrial sectors like smart manufacturing or wind energy and the emerging Dataspace applications in I4.0 and beyond industrial contexts.

References

[1] Wan, J., Cai, H., and Zhou, K., "Industrie 4.0: Enabling technologies," in *Proceedings of 2015 international conference on intelligent computing and internet of things*, IEEE, 2015, pp. 135–140.

[2] Matt, D. T. and Rauch, E., "Sme 4.0: The role of small-and medium-sized enterprises in the digital transformation," in Springer International Publishing, 2020, pp. 3–36.

[3] Singh, P., Beliatis, M. J., and Presser, M., "Enabling edge-driven dataspace integration through convergence of distributed technologies," *Internet of Things*, p. 101 087, 2024.

[4] Elhusseiny, H. M. and Crispim, J., "Smes, barriers and opportunities on adopting industry 4.0: A review.," *Procedia Computer Science*, vol. 196, pp. 864–871, 2022.

[5] Singh, P., Nidhi, N., Karpavice, J., Beliatis, M. J., and Presser, M., "Digital dataspace and business ecosystem growth for industrial roll-to-roll label printing manufacturing: A case study," in *Thinkmind Digital Library*, 2023.

[6] Philbin, S., Viswanathan, R., and Telukdarie, A., "Understanding how digital transformation can enable smes to achieve sustainable development: A systematic literature review," *Small Business International Review*, vol. 6, no. 1, e473, 2022.

[7] Lin, J., Yu, W., Zhang, N., Yang, X., Zhang, H., and Zhao, W., "A survey on internet of things: Architecture, enabling technologies, security and privacy, and applications," *IEEE internet of things journal*, vol. 4, no. 5, pp. 1125–1142, 2017.

[8] Khan, W. Z., Ahmed, E., Hakak, S., Yaqoob, I., and Ahmed, A., "Edge computing: A survey," *Future Generation Computer Systems*, vol. 97, pp. 219–235, 2019.

[9] Porter, M. E., Heppelmann, J. E., *et al.*, "How smart, connected products are transforming companies," *Harvard business review*, vol. 93, no. 10, pp. 96–114, 2015.

[10] Westerlund, M., Leminen, S., and Rajahonka, M., "Designing business models for the internet of things," 2014.

[11] Singh, P., Haq, A. U., Beliatis, M., *et al.*, "Meta standard requirements for harmonizing dataspace integration at the edge," in *2023 IEEE Conference on Standards for Communications and Networking (CSCN)*, IEEE, 2023, pp. 130–135.

[12] Carlini, S., "The drivers and benefits of edge computing," *Schneider Electric–Data Center Science Center*, vol. 8, 2016.

[13] Cao, J., Zhang, Q., Shi, W., Cao, J., Zhang, Q., and Shi, W., *Challenges and opportunities in edge computing.* Springer, 2018, pp. 59–70.

[14] Silva, P., Costan, A., and Antoniu, G., "Investigating edge vs. cloud computing trade-offs for stream processing," in *2019 IEEE International Conference on Big Data (Big Data)*, IEEE, 2019, pp. 469–474.

[15] Yeow, K., Gani, A., Ahmad, R. W., Rodrigues, J. J., and Ko, K., "Decentralized consensus for edge-centric internet of things: A review, taxonomy, and research issues," *IEEE Access*, vol. 6, pp. 1513–1524, 2017.

[16] Zehir, S. and Zehir, M., "Internet of things in blockchain ecosystem from organizational and business management perspectives," in Springer, 2020, pp. 47–62.

[17] Du, Y., Wang, Z., and Leung, V. C., "Blockchain-enabled edge intelligence for iot: Background, emerging trends and open issues," *Future Internet*, vol. 13, no. 2, p. 48, 2021, Times cited: 2 publisher: MDPI.

[18] Singh, P., Beliatis, M., Jakobsen, E. M., and Presser, M., "Digital business IoT maturity patterns from eu-IoT ecosystem," in *Shaping the Future of IoT with Edge Intelligence: How Edge Computing Enables the Next Generation of IoT Applications* (River Publishers Series in Communications and Networking), Sofia, R. C. and Soldatos, J., Eds., River Publishers Series in Communications and Networking. Germany: River Publishers, Oct. 2023, ch. 18, pp. 377–403, Available. DOI: 10.13052/rp-9788770040266.

[19] Singh, P., Beliatis, M. J., and Presser, M., "Enabling edge-driven dataspace integration through convergence of distributed technologies," *Internet of Things*, p. 101 087, 2024.

[20] Pan, J., Wang, J., Hester, A., Alqerm, I., Liu, Y., and Zhao, Y., "Edgechain: An edge-iot framework and prototype based on blockchain and smart contracts," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4719–4732, 2018.

[21] Singh, P., Holm, K., Beliatis, M. J., Ionita, A., Presser, M., Wolfgang, P., and Goduscheit, R. C., "Blockchain for economy of scale in wind industry: A demo case," in *Internet of Things: 5th The Global IoT Summit, GIoTS 2022, Dublin, Ireland, June 20–23, 2022, Revised Selected Papers*, Springer, 2023, pp. 175–186.

[22] Singh, P., Beliatis, M., Presser, M., *et al.*, "Data-driven iot ecosystem for cross business growth: An inspiration future internet model with dataspace at the edge," in *INTERNET 2024: The Sixteenth International Conference on Evolving Internet, ISBN: 978-1-68558-133-6*, 2024.

[23] Lamperti, S., Cavallo, A., and Sassanelli, C., "Digital servitization and business model innovation in smes: A model to escape from market disruption," *IEEE Transactions on Engineering Management*, pp. 1–15, 2023, Times cited: 3 event-title: IEEE Transactions on Engineering Management, ISSN: 1558-0040. DOI: 10.1109/TEM.2022.3233132.

[24] Hevner, A. R., March, S. T., Park, J., and Ram, S., "Design science in information systems research," *MIS Quarterly*, vol. 28, no. 1, pp. 75–105, 2004.

[25] Gregor, S. and Hevner, A. R., "Positioning and presenting design science research for maximum impact," *MIS Quarterly*, vol. 37, no. 2, pp. 337–355, 2013.

[26] Hevner, A., "The three cycle view of design science research," *Scandinavian Journal of Information Systems*, vol. 19, no. 2, p. 87, 2007.

Chapter 2

List of Papers

Background

This chapter consolidates the core research contributions of this PhD dissertation, encompassing six peer-reviewed research publications. These publications include five conference papers and one journal paper, all previously published with my role as the primary contributor. Relevant co-author statements are also provided in Appendix B.

The following table 2.1 maps each paper to its corresponding chapter for clarity. Please note that, in this dissertation, the term "Paper" is used to refer to the chapter in which the relevant research work is presented.

Table 2.1: Mapping of Dissertation Chapters to Published Papers

Paper	**Chapter**	**Publication Details**
Paper I	Chapter 3	*Blockchain for Economy of Scale in Wind Industry: A Demo Case*, Lecture Notes in Computer Science, Springer, 2022.
Paper II	Chapter 4	*Digital Dataspace and Business Ecosystem Growth for Industrial Roll-to-Roll Label Printing Manufacturing: A Case Study*, SENSORCOMM 2023.
Paper III	Chapter 5	*Meta Standard Requirements for Harmonizing Dataspace Integration at the Edge*, IEEE CSCN, Nov. 2023.
Paper IV	Chapter 6	*Enabling Edge-Driven Dataspace Integration Through Convergence of Distributed Technologies*, Internet of Things Journal, Elsevier, 2024.
Paper V	Chapter 7	*Data-Driven IoT Ecosystem for Cross Business Growth: An Inspiration Future Internet Model with Dataspace at the Edge*, INTERNET 2024 Conference.
Paper VI	Chapter 8	*Navigating the International Data Space To Build Edge-Driven Cross-Domain Dataspace Ecosystem*, GIECS 2024, Communications in Computer and Information Science, Springer.

This dissertation includes only the pre-print versions of the published papers, which are documented as chapters, as indicated in the mapping shown above in Table 2.1. Please refer to the following list for the original published versions along with their full reference details.

Paper I

Singh, P., Holm, K., Beliatis, M.J., Ionita, A., Presser, M., Wolfgang, P., Goduscheit, R.C. "**Blockchain for Economy of Scale in Wind Industry: A Demo Case**". In: *González-Vidal, A., Mohamed Abdelgawad, A., Sabir, E., Ziegler, S., Ladid, L. (eds) Internet of Things. GIoTS 2022. Lecture Notes in Computer Science. Springer, Cham.* Vol. 13533, 2022, pp. 175–186. DOI: 10.1007/978-3-031-20936-9_14.

Paper II

Singh, P., Nidhi, Karpavice, J., Beliatis, M. J., Presser, M. "**Digital Dataspace and Business Ecosystem Growth for Industrial Roll-to-Roll Label Printing Manufacturing: A Case Study**". In: *SENSORCOMM 2023: The Seventeenth International Conference on Sensor Technologies and Applications.* September 2023, ISBN 978-1-68558-090-2, pp. 7–13. DOI: 978-1-68558-090-2

Paper III

Singh, P., Nidhi, Haq, A. U., Beliatis, M. J., Presser, M. "**Meta Standard Requirements for Harmonizing Dataspace Integration at the Edge**". In: *IEEE Conference on Standards for Communications and Networking (CSCN).* Nov 2023, pp. 130–135. DOI: 10.1109/CSCN60443.2023.10453211.

Paper IV

Singh, P., Beliatis, M. J., Presser, M. "**Enabling Edge-Driven Dataspace Integration Through Convergence of Distributed Technologies**". In: *Internet of Things, Elsevir* Vol. 25, 2024, ISSN 2542-6605. DOI: 10.1016/j.iot.2024.101087.

Paper V

Singh, P., Nidhi, Beliatis, M. J., Presser, M. "**Data-Driven IoT Ecosystem for Cross Business Growth: An Inspiration Future Internet Model with Dataspace at the Edge.**" In: *INTERNET 2024:The Sixteenth International Conference on Evolving Internet.*, March 2024, ISSN: 2308-443X, pp. 19-25. DOI:978-1-68558-133-6.

Paper VI

Singh, P., Meratnia, N., Beliatis, M. J., Presser, M. "**Navigating the International Data Space To Build Edge-Driven Cross-Domain Dataspace Ecosystem.**" In: *Global Internet of Things and Edge Computing*

Summit. GIECS 2024. Communications in Computer and Information Science, Springer, Cham, Sep 2024, volume 2328, pp. 151-168. DOI: 10.1007/978-3-031-78572-6_10.

Chapter 3

Paper I - Blockchain for Economy of Scale in Wind Industry: A Demo Case

Parwinder Singh, Kristoffer Holm, Michail J. Beliatis, Andrei Ionita, Mirko Presser, Prinz Wolfgang, René C. Goduscheit

Originally published in *Lecture Notes in Computer Science, Springer Nature Switzerland AG*, January 2023, volume 13533, pp. 175–186. DOI: 10.1007/978-3-031-20936-9_14.

Abstract

Abstract. This paper summarizes the key findings of a qualitative study based on feedbacks of experts from the wind industry followed by a demo case of blockchain technology. This study includes investigation on mapping of supply chain and commodity products related end-to-end life cycle associated data and operational events. Furthermore, identification and blueprinting of the requirements have been pursued for enabling traceability at various stages of their life cycle utilizing blockchain. In this study context, blockchain offers digital traceability of operational events and associated data sharing with complete immutability, ownership, confidentiality, trust, and transparency across distributed supply chain which comprises of multiple stakeholders. In addition, digital technology intervention like IoT has been leveraged to support quality of operations in quantitative manner through real time data driven digitized operations. Thereby, providing economy of scale over operations execution on commodity products in wind industry. This study has focused only on bolts and fasteners associated commodity products and related supply chain. However, this provides a steppingstone foundation for future which can be scaled and mapped to any other commodity product and related supply chain in wind industry. Finally, this study also presents a demonstrator developed in a controlled lab environment to demystify the use of blockchain technology in related manufacturing and supply chain setups of the wind industry.

3.1 Introduction

The hype around blockchain technology in industry4.0 [1] is well established [2], [3], but its utility in a practical and operational settings is limited [4]. To better understand how the technology may be useful for specific industry sectors, projects which involve qualitative analysis of stakeholders for problems identification and realization of demo cases are typically running at national/European level in controlled environments. This paper presents the finding of such project, named as *"UnWind"*, focusing on wind industry in Denmark. As goals of this case study project, first an investigation of the wind turbine industry's value supply chain is conducted to identify and map processes with their potential benefits from blockchain technology. Second, an active collaboration has been done with companies/stakeholders of value chain who will get potential benefits from the findings of the first activity. This was done to consider their perspective and associated real world challenges to build a business use case and prototype demonstrator. Finally, the actual realization of a demonstrator with blockchain technology as a solution has been developed for the challenges identified during a series of workshops with the collaborated stakeholders/companies.

For this study, qualitative data was collected in the period January 2020 to March 2022 using an action research-inspired approach [5] with dual-focus on academia and practice. The qualitative data was collected through 10 unstructured interviews, 15 semi-structured interviews and 17 development workshops in which the industrial partners from the wind turbine industry have been interviewed/observed.

The Fastener project, which is an industry born initiative and closely connected to goals of UnWind project, seeks to standardize the procurement procedure for bolts and fasteners to simplify the buyer-supplier relationship between fastener/bolt supplier and wind turbine manufacturers (the buyer). From the bolt/fastener supplier's perspective this is intended to make the production setup simpler as they will no longer have to live up to individual demands from each of the turbine manufacturers. From the turbine manufacturer's perspective, the situation is improved due to reduction of the bull-whip effect as production errors at the supplier end are reduced and by extension the chance of delays in delivery is lessened as well. From both the buyer and supplier's perspective, there is also an added advantage of increasing batch sizes of fasteners, meaning economy of scale will be beneficial. In addition, there is less risk in storing backup components in warehouses, as there are more potential users of the safety stock if all bolts and fasteners are made to fit in with the standard schematics of all the wind turbine manufacturers production schemes. An important question arises here is that how these bolts and fasteners components are tied to blockchain context. The answer to that lies in the part of their procurement standardization process which also guides on how components in wind industry operations are traced across the value chain throughout the years. Practically, there is no need for high level traceability on all fasteners/bolts, as only larger fasteners are of critical importance for the wind turbine's operational performance. For this reason, the traceability standardization only applies to larger fasteners (8+ mm in thickness). To map the fasteners from the physical world to the blockchain in the digital world, a gateway technology in the form of Quick Response (QR) codes is utilized. Practically, each batch of fasteners will have a unique QR-code and through QR-code the information on the given batch of bolts is then digitally tied and stored to the blockchain. The way that each individual fastener becomes identifiable comes in the next steps of its life cycle.

3.1.1 The Fastener Lifecycle: Events and Data-points in the Blockchain

The figure 3.1 is illustrating the lifecycle of the fastener it goes through and the events that occur in its journey along with their data points that are logged and traced into the blockchain through the scanning of the QR-codes. The lifecycle consists of two major phases; 1) *Manufacturing and installation* of the fasteners (prior to the operational start of the turbine) and 2) *Service and decommissioning* of the fasteners (post operational start of the turbine).

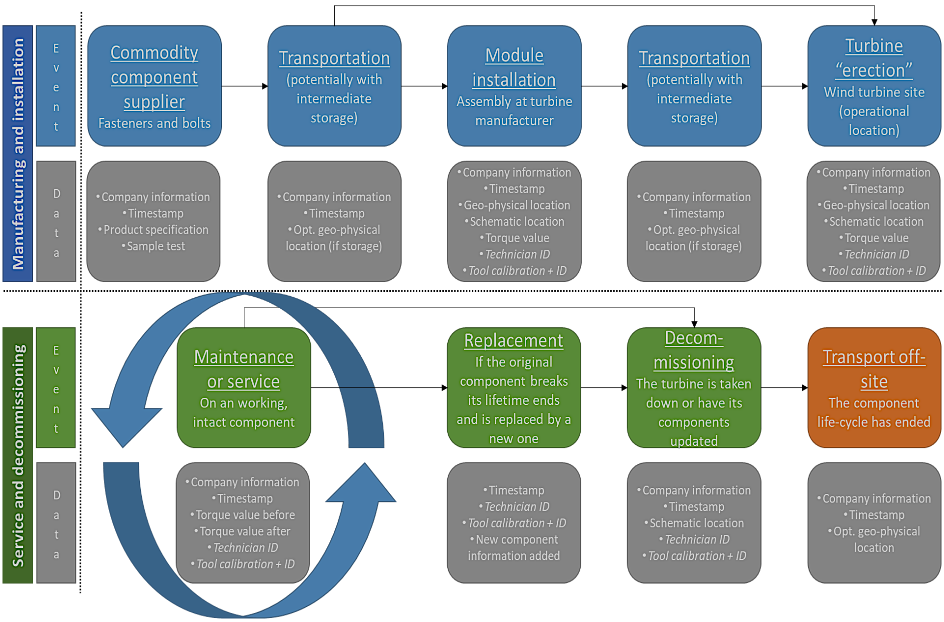


Figure 3.1: Overview of Events and Data Points in the Supply Chain and Life Cycle of Common Products (e.g., Bolts) in the Wind Industry

It should be noted that each event in the two phases can be considered a block of information (or transactions) in the chain of events making up the lifecycle of the fasteners thus forming the basis of representing these blocks in blockchain system. Two consistent data points that are recorded at every event, per transaction level, in the lifecycle, is a timestamp and a digital notation of the organization uploading the new event/operations. The first phase, where elements are marked blue (top row) on Figure 1, starts with the supply chain associated events for each fastener and ends after the fastener is installed in the fully assembled turbine. As touched upon earlier, the first event ties to the manufacturing of the fasteners (also called *commodity component tier one supplier*) and includes data points containing information on product specifications and a sample (quality) test performed on each batch of bolts produced at the manufacturer end. Second comes the *transportation event* of the fasteners, which includes geo-physical data for the relocation sites. This relocation may either lead directly to a *turbine erection event* (i.e., where the turbine is put under operation) or the fasteners may be sent to the *module installation* which then subsequently be *transported* to the *turbine erection* site. Regardless of whether there is module installation event or not, new data is logged into the blockchain when the fastener is no longer an individual component, but a part of a larger schematic (as a module or in the fully assembled turbine). This new data consists of registering the schematic location of the fastener, the value of torque with which the fastener is tightened and potentially the identification (ID) of the service technician who performs the installation operation of the fastener into turbine.

The second phase consists of events in the fastener's lifecycle that occur during the turbine's operational period and is illustrated with green and orange (3rd row) boxes. As opposed to the events that occur in the first phase, this second phase includes repeated events of maintenance or service, in which similar information may be updated repeatedly over 20-30 years in which the turbine remains operational. When maintenance is performed on the fasteners, several data-points are logged and updated in the blockchain. The torque value during each service cycle is registered, the information of technician facilitating the maintenance is logged and in case, a smart tool is used such as a digital wrench, the ID and calibration value of the tool is also logged. If everything occurs as intended, this event will occur regularly until the end of the turbine's life after which the fastener will be decommissioned along with the turbine, resulting in the logging and update of several data points again. In the case of a *replacement event,* which occurs between the *maintenance* and *decommissioning phase, there is a need to replace the broken fastener with a new one in the physical world.* This results in addition of new lifecycle to the blockchain based on the new fastener's lifecycle events (all the information logged as per phase 1). In principle, the replacement protocol can also occur multiple times and this has not been included in the figure 1 to avoid complexity. After this, the final event, which is marked as orange color (last box in 3rd row), occurs in the fastener lifecycle wherein the component is transported away and (ideally) sent to recycling, at a location which is logged into the blockchain as the final entry.

3.2 Business and Sustainability Implications of Blockchain Technology in Large Scale Wind Turbine Setups

Economies of scale: One of the primary points of the associated business case of the UnWind blockchain solution is to create the ideal conditions for turbine manufacturers to take advantage of economies of scale, so that turbines are cheaper to build as such, making the renewable energy production cheaper and more desirable for investments. Bolt/fastener suppliers also take advantage in the economy of scale as developing larger batches are cheaper, less time consuming and with fewer production risks in comparison to produce smaller batches of varying bolts/fasteners. This initial use case of the wind industry leveraging blockchain exclusively focused on bolts and fasteners which are considered commodity (non-compete) items amongst the turbine manufacturers. However, the intention is to expand the principle for all other critically important components of the turbine involved in the supply chain to improve their operational functionality in terms of quality, ownership, transparency, and monitoring.

The key selling point of the blockchain is tied to how the technology enables traceability across organizational bounds in the value chain [3]. By registering each event in commodity components lifecycle, it becomes easier to pinpoint where malfunctions occur or errors are made on the turbine's components – and most importantly, it enables actions to be taken faster and more accurately. For example, if a technician while performing maintenance on a turbine finds a fastener is broken, while it is still valid for its lifetime span, the technician can report this to be a potential issue with a replacement event in the component lifecycle. Via the blockchain this latest event can then be identified, and the responsible company can be informed, and other tasks can also be double-checked at other locations where fasteners are being used thereby enabling a more proactive approach to perform maintenance and service. Similarly, it may be found that a component issue tracks all the way back to a batch at the supplier level in which case the blockchain data can be utilized to identify other locations of fasteners from that batch and test if they also need replacement or maintenance service.

In extension of the traceability argument, the recording of events and product details also enables transparency [2] in what is inside the turbines meaning it is easier for service organizations to bring the right tools and equipment, as well as the rightfully certified personnel to perform tasks. One key issue in the wind turbine industry today (that is particularly expensive and time-consuming for offshore turbines) is that technicians are sent to perform tasks to turbine sites, only to realize upon arrival that they are either not properly certified or equipped to be able to perform their job. Furthermore, service contracts are rarely withheld by one company over the turbine's lifetime, meaning it is critical to pass reliable, accurate information to the companies taking over. However, since the original service provider is done with their commitment, they currently have little incentive to pass on information properly to their replacement. Even more critical is the fact that service contracts will often be taken over by competitors of the original service providers, meaning there may be reluctance to provide the competitor with accurate information. The blockchain, however, enables transparency through its data point registration, removing (or at least minimizing) the risk of inaccurate or lackluster documentation for former events.

Sustainability: Overall, the characteristics of a blockchain can enable better maintenance and monitoring for wind turbines, reducing production downtime which in turn enables more energy production per turbine over its lifetime, meaning each turbine has a larger amount of value that can help pay back the initial investment of the turbine. This translates to a lower levelized cost of energy for wind-based power generation, thereby improving the conditions of the industry. In other words, blockchain is an opportunity for wind turbine manufacturers to collaborate with their immediate competitors (other wind turbine manufacturers) to better compete with other energy producing industries like solar [6], fossil or nuclear fuels. In other words, blockchain technology holds promise to improve sustainability aspects for the service part of the wind industry [7], in addition to other ways this technology is perceived to be an enabler for more sustainable practices [8].

3.3 Realization of Demonstrator in Controlled Environment

To realize the system context of blockchain contribution in wind turbine supply chain industry, a demo has been designed and developed in a controlled lab environment known as *DigiMicroFactory* Lab at department of Business Development and Technology, Aarhus University in Herning, Denmark. Typically, as discussed earlier, wind turbine industry is based on supply chain consisting of multiple stakeholders, workflows, events, data points, and their supplied services or commodities. The different stakeholders that are involved, but are not limited to - wind manufacturers, first tier vendors, services operations staff. There are different system operations associated to the supply chain workflow, such as transportation, assembly, installation, service (replacement/modifications) maintenance and de-commissioning of wind turbine components [9]. This needs to be performed by the relevant service personnel or engineer at specific value chain lifecycle events either at the onsite or remote location. Each of these operations are associated with lots of data which is usually stored in traditional internet web (including cloud) enabled IT systems based on the inputs of service engineer [10] . Some of these inputs are collected based on output/outcome of mechanical operations directly or via digital systems placed in support to mechanical operations. The service engineer records the output/outcome and enters them in the system interface manually. This human intervention is often prone to errors. For example, consider a scenario wherein a bolt needs to be fastened by wrench (normal or digital) and engineer applies a force on the bolt to be tightened up for the target torque. The service engineer interprets the outcome (i.e., success or failure) of the operation, based on his cognitive skills and updates the system related user interface qualitatively. Therefore, many such mechanical operations where end-to-end digital intervention is not present are often prone to human errors. Additionally, such small errors could cost heavily in energy production running phase of the wind turbine, in terms of down time due to failed operations, quality compromised, missing proactive handling of incorrect operations, and incorrect interpreted information flow to different stakeholder that may leads to conflicting situations and inappropriate action flows. As explained earlier in Figure 1, the service and maintenance aspects of the wind turbine case are really where the complexity comes to grow and are also the potential areas of issue in the value chain.

For the service technicians to both gain and give value to the blockchain solution, they must be able to interact with the solution, and this requires a user interface and backbone technologies in which the technicians can read and update the blockchain data without necessarily understanding the technology. Practically, the technicians will interact with the blockchain by scanning the Bar/QR-code with a tablet/mobile, which will give them access to either a website or an application through which the relevant information can be accessed and updated. Furthermore, there are multiple IT web systems (referred to as traditional/legacy systems/applications) involved and active from many years, such as delivery tracking, inventory management, service management, quality metrics etc. These systems run in the organizational boundaries, as they belong to different stakeholders, thus creating multiple data silos as well. To overcome this situation, their information or data sharing needs inter-organizational related cross functional system interfaces development. These interfaces should be abstracted and converged into a unified data format among all stakeholders in a trusted, traceable, and transparent manner. At the same time, their organization boundaries need to be protected and defended with complete trust and security. These objectives pose data integration, security, integrity and ownership challenges among different systems and their stakeholders. Therefore, there is need to build a novel framework in place, which can address these challenges and fulfill the relevant objectives at their design level. This is exactly what has been achieved through the developed blockchain enabled unwind demonstrator system concept presented in this paper.

3.3.1 Overall High Level Architecture

The high-level architecture of the unwind demo concept is illustrated in Figure 3.2. The assumption for the design architecture is the single commodity component operation support i.e., fastener or bolt related value chain in wind industry. The operational use case is a scenario of bolt fastening in wind turbine assembly with capabilities of digital recording its fastening operations with trust to support quality. In the architecture, the following functional layers are shown from the operations point of view:

a) **Wind Turbines at onsite location** – These are the target entities to perform operations upon in the wind turbine supply chain. To build a turbine at onsite, it needs transportation of components like bolts, blades, rotors, motors, electronics equipment, poles etc. Once they transported, then there is a assembly and installation workflow of components needs to be executed. Here, it is assumed that all operations are performed onsite with end-to-end digital traceability in support of mechanical operations. This wind turbine related operational data at onsite acts as a data source.

b) **IoT Edge Device (Bolt/fasten control)** – This is the system interface available at onsite to support turbine operations digitally. In our architecture, it is the kind of bolt fasten control system which monitors the bolt screw operations and related data and then send this data towards systems in upstream flow in real time. These types of systems can communicate with wind turbine systems using internet (e.g., via 5G/LTE/Wi-Fi) and radio (Bluetooth, LoraWan [11], etc.) interfaces and are available close to the wind turbine [12]. That is why it has been referred as IoT edge device in the given architecture. This also gives the scope of improving operations, in terms of low latency, aligned to edge computing goals in future [13], [14].

c) **Traditional organizational level centralized systems:** These are the enterprise level IT web enabled systems (usually enterprise resource planning – ERP - applications) which belong to different stakeholders. These systems record data coming from onsite locations for various events in value chain, process and store them in their databases. It is important to mention here, that the architecture supports the traditional systems flow as is, since they are functional and can save on investments made already and offer ease of work over familiar system. Therefore, this architecture does not suggest replacing the existing systems, rather advocates complement the existing system by adding a new system/net work of blockchain.

d) **Blockchain Network:** This is the new system in place which aims at supporting complete supply chain of wind turbine in a distributed sharing, trustable, immutable, transparent, and secure manner by design. Blockchain systems perfectly fit in this context as they offer the same by design and that is the reason it has been chosen as an integral part of this architecture. Additionally, with the advent of modern technologies like Blockchain, IoT and edge computing and their integration is expected to benefit [15] the traditional way of managing the wind turbine supply chain. These technologies integration induce lots of capabilities which includes real time data generation, collection, and processing through IoT, distributed storage, transparency, trust, quality, and traceability among different stakeholders using blockchain and low latency driven decision near the data sources (i.e. wind turbine sites) using edge computing and many more [16].

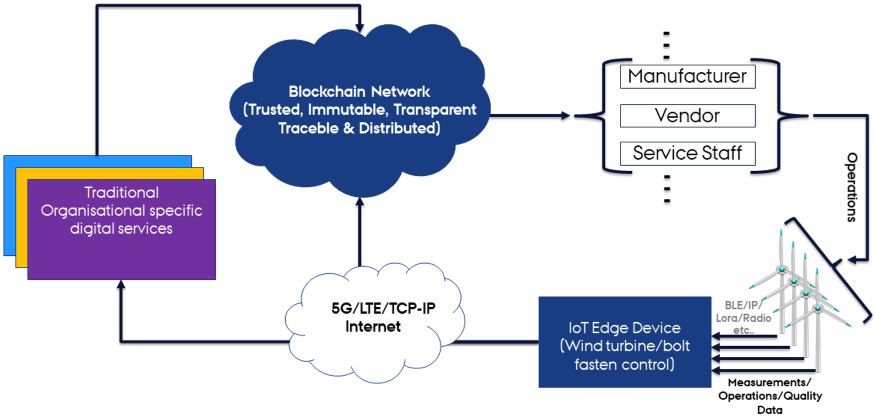


Figure 3.2: High Level Digital Traceability System Architecture

3.3.2 Demonstrator Realization and Code Snippet

For the demonstrator, different digital equipments were used as illustrated in Table 1.

Table 1: Equipment and Services Used For The Realization of Blockchain Demonstrator

Item	Purpose
Wind Turbine	A small demo turbine installed in lab upon which the Bolt operations to be performed to replicate the real-world scenario.
Bolt	12 mm Bolt that needs to be fasten to the wind turbine.
Digital Wrench	To fasten the bolt and digital monitoring of the torque readings.
Raspberry Pi+ LCD display	Acted as an edge device to collect the data from different tools during installation or maintenance operation. LCD is used on top of RaspberryPi to display real time data from real time operations.
BAR/QR Code Scanner	To scan the QR and Bar codes present on the different components of Wind turbine assembly and to identify them uniquely.
QR/BAR Code printers	This is used to print the QR/BAR code for digitalize tagging of the physical assets.

Item	**Purpose**
Blockchain as a Service	Public Blockchain service (Ethereum [17] based) from Unwind Project that has REST APIs offering over internet to ingest and query data related to supply chain operations.
Local blockchain enabled Edge server	Private blockchain service (Hyperledger Fabric [18] - HFabric - based) running at edge server to simulate permissioned blockchain service in a controlled and constrained lab environment.
Programming Flow	Node-red based programming flow developed to control/manage devices and related data processing towards blockchain.

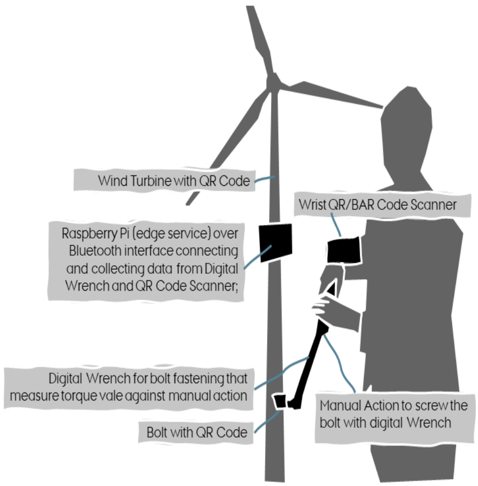


Figure 3.3: The Digital Traceability with HFabric or Ethereum Blockchain Demonstrator within the DigiMicrofactory Lab at Aarhus University BTECH

Following the architecture (Figure 2), a demonstrator has been developed while focusing only on bolt fastening operation at wind turbine is shown in Figure 3.3. The physical components such as wind turbine and the bolt to be fastened is tagged with QR code. These codes are generated by registering the turbine and batch of bolts details in blockchain system beforehand. To read these QR codes, a digital bar code scanner has been used in our demo. This tagging and scanning of QR code is one of the methods to represent identifications of physical systems in digital world in unique manner. In other words, it identifies the specific operation being performed at specific wind turbine with specific component at relevant geolocation. In our case, it establishes a *semantic relationship (one to many type) [11] of specific bolt from a specific vendor/supplier associated with the digital wrench (tool/device) operation performed by specific service engineer on specific wind turbine which belongs to a specific owner/operator at specific onsite/offshore location.* An information tied in such semantic relationship model offers semantic search and traceability capabilities with multiple dimensions at ease.

The demonstrator consists of digital wrench device which fastens the bolt to the predefined torque threshold. There is a raspberry Pi, which acts as an edge gateway/device providing internet connectivity to upstream systems in order to send them recorded data and as well as radio interfaces connectivity (based on Bluetooth) towards digital wrench and bar code scanner, to monitor onsite component operations. Additionally, the same edge device in the demo has also been used to install and implement a custom programming service flow (based on node-red), as shown in Figure 3.4. This service is used to control and manage the bolt screw devices, collect, display, and send data to upstream traditional systems as well as blockchain systems (Hyperledger fabric or Ethereum based) in real time via REST based application programming interfaces (APIs) [19]. The complete operational flow is given as follows:

a) Register the turbine and related batch of bolts with blockchain application. The output of registration event generates a QR code.

b) Tag the turbine and related components with QR code labels.

c) Using programming flow running at edge device, register all the devices (scanner and digital wrench) over Bluetooth interface.

d) Scan all the QR codes (wind turbine and bolt related).

e) Perform bolt fasten operation using digital wrench. This will send torque specific reading in real time to the programming flow running at edge.

f) Metadata (data about wind turbine and bolt identification) and data from specific fasten operation, in semantic relationship manner, is merged in programming flow and send into two different flows upstream traditional systems and blockchain systems via REST APIs.

Note: The APIs serves as the interface for storage and retrieval of data related to bolt, wind turbine, and technical operations to/from Hyperledger fabric (used as a private) or Ethereum (used as public) based blockchain. The registration of the bolt batches, wind turbine, documented technical service and recycling data invokes corresponding solidity based smart contracts in the blockchain that validates and stores the respective information. The APIs endpoints process and forward the requests to the smart contract by means of the web3 library. Different stakeholders, choose to be part of the blockchain network, can access the data of their interest based on their privileges (as per the smart contract agreements

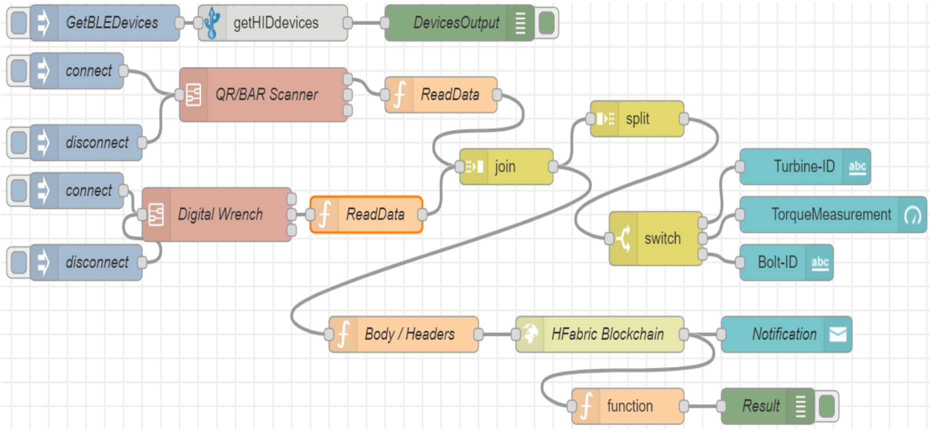


Figure 3.4: Code Snippet for Connecting All Reader Devices and Enabling Dataflow and Registration with Blockchain

among the stakeholders) in near real time, shareable, reliable, traceable, secure, and transparent manner. Thus, able to perform the required actions (proactive or course of correction) as per the need of the situational event.

3.4 Conclusions

Throughout this industry 4.0 case study, we examined with qualitative methods the challenges which wind industry is facing during manufacturing and operational phases for the entire wind turbine life cycle of commodity components. We analyzed the findings and identified the area of the supply chain for commodity products where blockchain technology can contribute significantly as solution to enhance the economy of scale for commodity products in terms of digital traceability, quality, operations with trust and transparency as well as to establish a more sustainable supply chain. Furthermore, we developed a working demo in a controlled lab environment to demonstrate the feasibility of using blockchain in the digital traceability of commodity products such like bolts/fasteners adding value to the entire life cycle of wind turbines in this industry.

Acknowledgements. The authors would like to thank many actors from the wind industry at APQP4wind, Vestas and Simens Gamesa for valuable discussions on various subjects. The study partially funded from the Danish Industries Fund under the grant name “UnWind” (https://industriensfond.dk/project/unwind-blockchain-i-vindmolleindustrien/).

References

[1] Beliatis, M. J., Jensen, K., Ellegaard, L., Aagaard, A., and Presser, M., “Next generation industrial iot digitalization for traceability in metal

manufacturing industry: A case study of industry 4.0," *Electronics*, vol. 10, no. 5, 2021.

[2] Treiblmaier, H., "The impact of the blockchain on the supply chain: A theory-based research framework and a call for action," *Supply Chain Management*, vol. 23, no. 6, pp. 545–559, 2018.

[3] Cole, R., Stevenson, M., and Aitken, J., "Blockchain technology: Implications for operations and supply chain management," *Supply Chain Management*, vol. 24, no. 4, pp. 469–483, Jun. 2019.

[4] Holm, K. and Goduscheit, R. C., "Assessing the technology readiness level of current blockchain use cases," in *2020 IEEE Technology & Engineering Management Conference (TEMSCON)*, IEEE, 2020, pp. 1–6.

[5] Coughlan, P. and Coghlan, D., "Action research for operations management," *International Journal of Operations & Production Management*, vol. 22, no. 2, pp. 220–240, 2002.

[6] Beliatis, M. J. *et al.*, "Slot-die-coated v2o5 as hole transport layer for flexible organic solar cells and optoelectronic devices," *Advanced Engineering Materials*, vol. 18, no. 8, pp. 1494–1503, Aug. 2016.

[7] Holm, K., "Blockchain as a sustainable service-enabler: A case of wind turbine traceability," in *Lecture Notes in Mechanical Engineering*, 2022, pp. 516–523.

[8] Saberi, S., Kouhizadeh, M., Sarkis, J., and Shen, L., "Blockchain technology and its relationships to sustainable supply chain management," *International Journal of Production Research*, vol. 57, no. 7, pp. 2117–2135, Apr. 2019.

[9] Nejad, A. R. *et al.*, "Wind turbine drivetrains: State-of-the-art technologies and future development trends," *Wind Energy Science*, vol. 7, no. 1, pp. 387–411, 2022.

[10] Zhao, Y., Xia, S., Zhang, J., Hu, Y., and Wu, M., "Effect of the digital transformation of power system on renewable energy utilization in china," *Ieee Access*, vol. 9, pp. 96 201–96 209, 2021.

[11] Singh, P., Acharya, K. S., Beliatis, M. J., and Presser, M., "Semantic search system for real time occupancy," in *2021 IEEE International Conference on Internet of Things and Intelligence Systems (IoTaIS)*, IEEE, 2021, pp. 49–55.

[12] Hossain, M. L., Abu-Siada, A., Muyeen, S., Hasan, M. M., and Rahman, M. M., "Industrial iot based condition monitoring for wind energy conversion system," *CSEE Journal of Power and Energy Systems*, vol. 7, no. 3, pp. 654–664, 2020.

[13] Olaniyan, R., Fadahunsi, O., Maheswaran, M., and Zhani, M. F., "Opportunistic edge computing: Concepts, opportunities and research challenges," *Future Generation Computer Systems*, vol. 89, pp. 633–645, 2018.

[14] Tziouvaras, A. and Foukalas, F., "Edge ai for industry 4.0: An internet of things approach," in *Proceedings of the 24th Pan-Hellenic Conference on Informatics*, 2020, pp. 121–126.

[15] Pan, J., Wang, J., Hester, A., Alqerm, I., Liu, Y., and Zhao, Y., "Edgechain: An edge-iot framework and prototype based on blockchain and smart contracts," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4719–4732, 2018.

[16] Seitz, A., Henze, D., Miehle, D., Bruegge, B., Nickles, J., and Sauer, M., "Fog computing as enabler for blockchain-based iiot app marketplaces-a case study," in *2018 Fifth international conference on internet of things: systems, management and security*, IEEE, 2018, pp. 182–188.

[17] Alrubei, S., Rigelsford, J., Willis, C., and Ball, E., "Ethereum blockchain for securing the internet of things: Practical implementation and performance evaluation," in *2019 International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*, IEEE, 2019, pp. 1–5.

[18] L. Foschini A. Gavagna, G. M. and Montanari, R., "Hyperledger fabric blockchain: Chaincode performance analysis," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, 2020.

[19] M. S. Alam U. D. Atmojo, J. O. B. and Lastra, J. L. M., "A rest and http-based service architecture for industrial facilities," in *2020 IEEE Conference on Industrial Cyberphysical Systems (ICPS)*, 2020.

Chapter 4

Paper II - Digital Dataspace and Business Ecosystem Growth for Industrial Roll-to-Roll Label Printing Manufacturing: A Case Study

Parwinder Singh, Nidhi, Justina Karpavice, Michail J. Beliatis, Mirko Presser

Originally published in *SENSORCOMM 2023 : The Seventeenth International Conference on Sensor Technologies and Applications.* September 2023, ISBN 978-1-68558-090-2, pp. 7–13. DOI:ISBN 978-1-68558-090-2

Abstract

Manufacturing industries across the globe are adopting modern technologies rapidly and are now moving towards Industry 5.0. However, many manufacturing units still need to yield their Industry 4.0 level amid reasons including heavy investments in upgrading existing infrastructure and scrapping existing machinery in the name of modern digitalization. These Small and Medium-sized Enterprises (SMEs) need help finding the best-fit technology to meet their business requirement and impact along with the capital to support digitization. In this study, we present one such case of a Danish manufacturing industry, a label printing SME, stranded between conventional technology and a race towards modern digitization. On the one hand, it has old machinery with a large volume of heavy printing and lamination operations with an established customer base. On the other hand, it is willing to digitalize its manufacturing processes (with minimal upgradation in its existing mechanical infrastructure) to enhance its efficiency and sustainability and catch up with the pace of digitalization for further expansion. This can be achieved by nurturing the benefits of digitalization through the latest technologies, such as Enterprise Resource Planning (ERP), Internet of Things (IoT), Edge, Blockchain, Cloud computing, etc. We studied and analyzed the case of the SME in consideration to understand the core requirement and anticipated bottleneck in the process. This paper has presented our findings through related business process flow mappings, challenges, vision, and possible digital architectural solutions using modern-day technologies, scientific approaches and realization tools. These findings and methods will also be value-added and applicable to other SMEs in similar situations. Thus, it enables them to save their Return on Investment (ROI) while adapting to modern technologies with minimal risk, impact and investments.

4.1 Introduction

The manufacturing industry has been the backbone of any economy across the globe for decades. With the advent of Industry 4.0 [1], the manufacturing industry has focused towards adaptation, automation and data-driven manufacturing operations leveraging modern technologies such as the Internet of Things (IoT) [2], Big Data [3], Enterprise Resource Planning (ERP), Platform Solutions, etc. [4]. The industry is slowly moving towards Industry 5.0, wherein human and machine interaction is a prime focus. However, the Industry 5.0 evolution can only be achieved once Industry 4.0 is adopted, as it addresses the socio-economic impacts of Industry 4.0 on humans [5]. Thus, the manufacturing industry must first achieve the Industry 4.0 goals, which seems way easier for large manufacturing companies because of the available workforce, resources and investments than SMEs. Adopting Industry 4.0 seems strenuous for SMEs as they have limitations with finances, knowledge resources, workforce, and trending technical advances [6].

To deal with similar challenges of SMEs in Denmark, the Danish government has started a growth technology project called Manufactory to boost collaborations between industry and knowledge institutions at a regional level under the supervision of the Danish Industry Foundation [7]. This project has several consortium partners from academia and industry to develop a common platform where both can collaborate and co-create solutions, leveraging modern cutting-edge technologies, to real-world industrial problems (particularly for SMEs). This study results from one such initiative among several taken under this Manufactory project at our Department of Business Development and Technology (BTECH), Aarhus University, - a key academic partner in this project. We collaborated with a Roll-to-Roll (R2R) Label Printing SME with a worldwide clientele base in the Midtjylland area of Denmark. The company got its latest machinery and other infrastructure installed in 2007; thus, the upgradation to modern mechanical infrastructure is not straightforward. However, the SME still has a desire to nurture benefits and cope with modern digitalization to create efficient and value-added expansions, such as data sharing with their customers in real-time, making the production environment smarter and augmenting digitalization slowly and steadily to come up with a differentiator and transformed business model in the Label Printing domain. In this context, for example, IoT technology and related edge network sensing capabilities, such as environmental, print quality, ink toning, bar-code, energy efficiency, predictive maintenance, production line sensing, etc., can streamline operations, improve quality control, enhance supply chain visibility, and reduce costs for the SME. Thus ultimately benefiting both the SME and its customers.

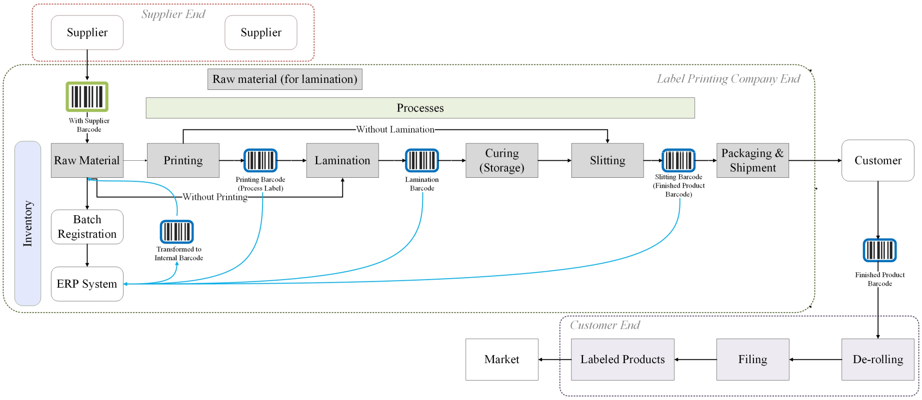


Figure 4.1: Business Process Model for the SME Workflow

This study summarizes the insights and knowledge acquired while collaborating with the R2R label printing SME through workshops and one-on-one interactions with the company's top-level management to understand their business process, perspectives, associated challenges and vision. The study contributes with both business and technical standpoints wherein a business of SME is investigated, which leads to the identification of challenges and process level understanding and results in a system architecture that supports the vision and mission towards digitalization and sustainability.

The rest of the paper is organised as follows: Section II will highlight the background of the label printing industries with associated challenges and our approach to the investigation and analysis. Section III will elaborate on our findings and business propositions, and highlight BPM. Section IV will present the proposed architecture to mitigate and digitize the concern challenge. Finally, Section V will conclude the paper with future opportunities and research scope.

4.2 Background and Methodologies

The label printing industry is essential in manufacturing for creating labels for varied purposes such as brand promotion, product packaging, information leaflets, etc. These industries are further classified based on the print type, material, and adopted technology. For instance, the labels can be of various types with materials ranging from paper to specialized films, such as pressure-sensitive, shrink sleeves, cut-and-stack, in-mould, etc., employing different print technologies like flexographic, digital, and offset printing [8]. With such exhaustive and complex processes, it is evident that errors in label print manufacturing can disrupt production and affect label quality. Common errors in the print industry include misalignment, colour variations, ink smudging/fading, inconsistent print quality, registration, paper jams, spooling, tearing, adhesive problems, barcode errors, missing labels, material compatibility, poor winding quality, etc. These errors can cascade on SMEs, impacting their operations, customer relationships, costs, and overall business performance [9]. To address these challenges, manufacturers employ quality control, equipment maintenance, and staff training to ensure labels consistently meet quality and accuracy standards. In this realm, the convergence of technologies (e.g., IoT, AI, and Blockchain) and dedicated sensors orchestrates precision and efficiency, from monitoring temperature, humidity and pressure to tracking motion, light and colour functions of label printing. The application of sensors in line with the manufacturing process enables real-time data generation, monitoring and collection to provide efficient label printing and quality control in industrial automation settings.

We adopted Business Process Modeling and Notation (BPMN) models and tools [10] to realize efficiently the SME's process flow. It helped us to identify bottlenecks, challenges, gaps and opportunities to enable digital technologies in SME's environment. In addition, a mix of case study and action research methodologies have been used to conduct our research for this study. We also conducted onsite and offsite workshops with the company's top-level management and researchers from the university as a method of qualitative research to understand, observe, collect, map and analyse the information about the relevant processes, operational flows, challenges and future perspectives. The workshops were conducted openly, focusing on transparent communication, setting clear agenda and objectives, practical use cases and scenarios of the SME, and the participant's presentation, followed by in-person visits to the SME facility, discussions, feedback loop and lots of brainstorming sessions.

4.3 Business Analysis

The Danish Label Printing SME for our study is a pioneer in flexible packaging employing the Flexo printing technique followed by solvent-free and water-based laminations. The SME also provide customer-ready foils for customers to use on their packaging machines after cutting and adding customized functionalities to the foil, such as micro-perforation, embossing, etc., if requested. Based on the data gathered through the workshops, the process flow was developed to map all the operations and inter-departmental process flows as illustrated in Figure 4.1. It represents the entities in the supply chain, from raw materials to market-ready finished products, primarily categorized into three clusters: Supplier End, Label Printing Company (i.e., SME) End, and Customer End, as explained below:

- **Supplier End**: Once the raw materials are delivered from the suppliers, the supplier barcode placed on each item is uploaded to the ERP system and then transferred to an internal barcode.

- **Label Printing Company End**: Label manufacturing consists of four main processes: printing, lamination, curing, and slitting. After each process, the new barcode is created, printed and attached to the printed/laminated/slitted reel, except for the curing operation, which needs to be processed separately. Then, the reels, a.k.a bundles, are slitted into smaller reels labelled with the finished product barcode, packaged, and shipped to customers.

- **Customer End**: The received reels of labels are de-rolled and applied to the customer products in distinct techniques concerning the type of the products that must be packaged, i.e., labels can be attached, wrapped, laminated on the product, or filled with the product.

4.3.1 Identified Challenges

Flexography, used by the SME, is a well-known printing technique for producing high-quality images and graphics at high printing speed in a versatile and cost-effective way [11]. Although flexography is based on a simple concept of ink transfer, multiple variables affect the final quality of the production, including properties of the plates, anilox rollers, printing pressure, and printing substrates [12]. As a result, the investigated SME faces several challenges related to print defects requiring frequent process stoppages for quality control. The following challenges have been identified as outcomes of our workshops and visits to the SME's facility.

- **Error Management**: The main challenge for the SME, which it wants to address immediately, is error management. During the printing or lamination operation, errors occur and need to be communicated internally and to the customer. For internal communication, two (green and red) labels are used manually to mark the start and end of the material to be scrapped during the slitting processes, shown in Figure 4.2. Similarly, the external error marking allows SME's customers to manually adjust their filling and packaging operations based on the position of the red label, thus acting as passive markers of errors.
- **Automation of Processes**: The processes at the production line are primarily manual. For instance, the real-time information flow between different production stages and waste management is logged manually. The company's ERP system does not cover all the operational aspects, as upgrading and integrating old machinery infrastructure is challenging, impacting overall productivity efficiency.
- **Reduction in Waste Material**: During printing and lamination operations, the occurred errors result in waste. But no data or logs are maintained for the errors' cause, occurrence or frequency. Therefore, a semantic context awareness-driven digital method enablement must be there to enhance productivity and reduce scraps.
- **Data-driven Compliances and Decision Enablement**: Abiding the Climate Act, Denmark, like other EU countries, aims to significantly reduce greenhouse gas emissions by 50-54% in 2025 and 70% in 2030 compared to 1990 levels [13]. Thus, this necessitates stringent compliance

Figure 4.2: Error in Bundle (Red/Green Label Indicators)

from manufacturing industries, demanding comprehensive data on their operations. The SME is found to be struggling to comply with these requirements due to a lack of data-driven operations.

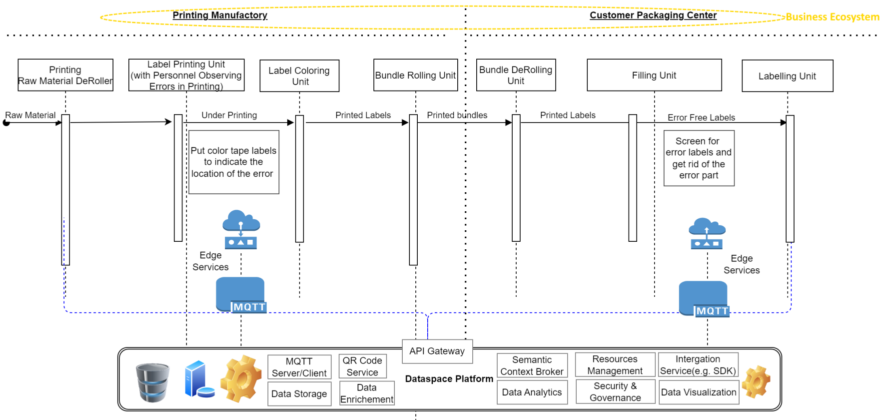


Figure 4.3: Digital Intervention Architecture

4.3.2 Vision

Our workshops and brainstorming sessions with the SME management highlighted their vision behind the initiative to incorporate and facilitate the convergence of technologies into their existing manufacturing and production lines.

The SME expects to digitalize the operations to make their products smarter. For example, the printed bundles have associated a lot of operational data during their printing journey, but it is not captured and thus adds no value. The SME's immediate focus is on error management, a low-hanging fruit where digital operations can be induced as a starting point and scaled incrementally for other challenges. Additionally, they want to develop a data-driven servicing model in future, which can sustain and expand the current business horizon for the SME. Therefore, we proposed a platform model based on which the servitization business model [14] will provide value-chain enablement across customers and also enable them to make their production environment smarter.

4.4 Architecture

This section will primarily provide potential solutions and an architecture to address the identified challenges, focusing on Error Management as the starting point for inducting digital tools in the SME environment. In the present scenario, the entire error identification at its occurrence and relevant communication/action process is handled manually without any digital intervention, giving rise to our problem statement:
"How can we make error management and dependent operations smarter to enhance the overall productivity?"

Based on our interactions and analysis, we proposed a solution to tackle the challenge of managing errors through the entire workflow and having a log of occurrences. A digital intervention architecture is shown in Figure 4.3 to address this question. This consists of digital enablement of services at SME and relevant customer on-premises environments, and finally, the Dataspace platform-enabled services provide real-time distributed data services to all the stakeholders. The high-level operational flow of the label printing process is given in the following three steps:

- **Step A**: The Manufacturing Unit receives raw material in huge bundles, each of which goes on the de-rolled roller.
- **Step B**: The de-rolled material goes through the design printing operation, where errors are manually observed during the printing operation and marked with tapes for identification, which is removed later. The error could occur for many reasons, as explained earlier. For instance, when one bundle is finished, it is spliced into a new bundle, and the point of occurrence is notified using tapes. Apart from splicing, errors occur due to printing operations, such as misalignment, colour variations, ink smudging, inconsistent print quality, machine malfunction, etc. To address such errors, we proposed a digital intervention to detect and record the error (explained in the next section).
- **Step C**: At the customer end, the reverse process is repeated to de-roll the bundle and make adjustments on a machine for the error flag in the

bundle, and then the error-free labels are split from the bundle and pasted onto the final product.

The above steps have associated data labels as barcodes or manual receipts, but they lack real-time linkage as they flow through the printing operations. Thus, to automate this stage, we propose IoT edge-enabled Quick Response (QR) code-based data labels at all stages of printing operations. We also propose to generate a *"Context-Aware Operational Environment"* and link the information in different functional flows in future during the entire "Order and Error Management" process for the relevant customer. The semantic model can easily capture the business level flow relevant operational context and provides technical implementation grounding at the system level, such as using semantic RDF [15], NGSI-LD [16] or JSON-LD [17] standards. In the following subsections, we have described the applicable process and system model using the semantics ontology approach that can be implemented by the supporting digital platform to induce digitization in the processes.

4.4.1 Digital Processing Model for Printings Operations

Figure 4.4 illustrates the semantic data model for managing printing errors. The *Customer* places an *Order*, which has to be processed with delivery target *Date* by assigning an order identification - *ID*. Similarly, the *ID* property can be used for every operational entity in printing processing. To process this order, there is a need for *Bundle* that contains the raw material for printing with a certain *Length* in meters provided by the customer in his order to print the labels. During the printing operation, there might be *Error* associated, which is identified by an *ID*, defined by its *Metadata*, occurred due to a *Reason* that in turn belongs to a specific *Type* of an *Error*. The printing bundle is finally prepared on specific *Date* for delivery to the customer that receives the final bundle on relevant *Date*.

4.4.2 Semantic System Model

To initiate digitization, we mapped the SME processes against the digital processes. The system is perceived as a semantic system that provides operational and linked data context awareness [18]. Figure 4.5 shows the semantic system model that captures the system's relevant operational entity level information. The system focuses on the activities inside a label printing manufacturing unit, that starts with the *Raw Material Bundle*, a type of *Bundle* used for *Printing Label*. During the printing operation, there can be an occurrence of *Error*, which is explained by its metadata that is digitally captured in a *QR Code Label* by the error detection method that also sends the related error information to the *Dataspace* via *MQTT* protocol. In addition, the error method can also query the information from *Dataspace* for various purposes, such as to generate a QR code for encountered errors. At present, the error is monitored manually by personnel of the SME, but in the future, we propose to automate and replace it through

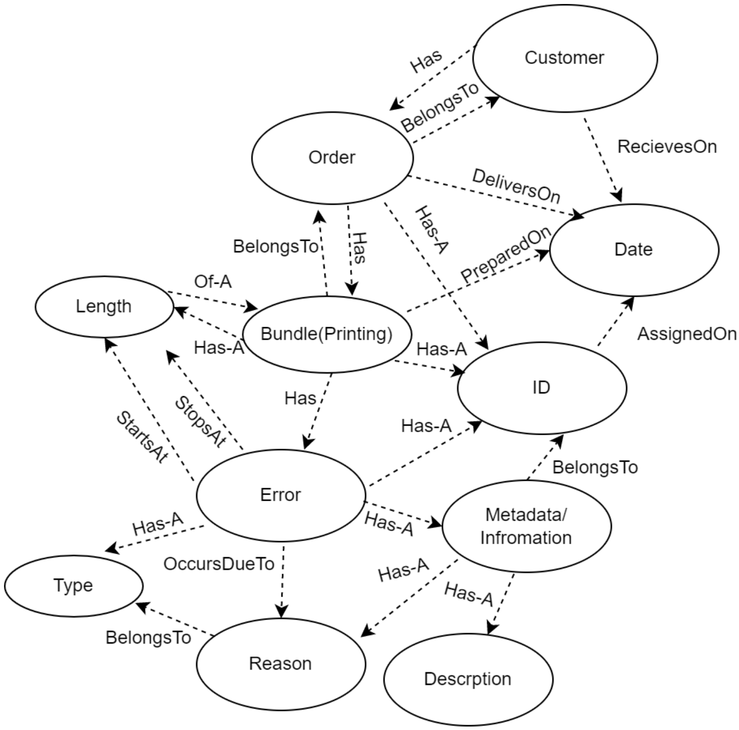


Figure 4.4: Semantic Processing Model for Label Printing Operations

edge-enabled (on-premises) IoT sensing and computer vision capabilities. The data received by *Dataspace* platform is persisted in the *Database*, which can be queried later.

4.4.3 Dataspace Platform

We propose a platform with Dataspace capabilities [19], currently implemented at the university facility, for the SME to prototype and validate its expectations towards initializing digitalization for its error management challenge. The platform is implemented with dataspace capabilities for many reasons: first, the SME management wants to enable digitalization for its processes to augment data-driven decisions over cross-organizational boundaries and thus exploit it as a value addition in their products, i.e., printing bundles, for their business value-chain. For example, the transition of dummy bundles into smart bundles having associated data at the customer end in real-time that can explain the bundle digitally and enable customers to fine-tune their operations and planning accordingly, thus enhancing operational productivity at the customer end. Second, it allows the SME to bring innovation in their business value chain that spans and has an impact at the customer end. This will enable the expansion of SME's portfolio using the platform services model based on smart printing products. Finally, the significant objective for any business is to create

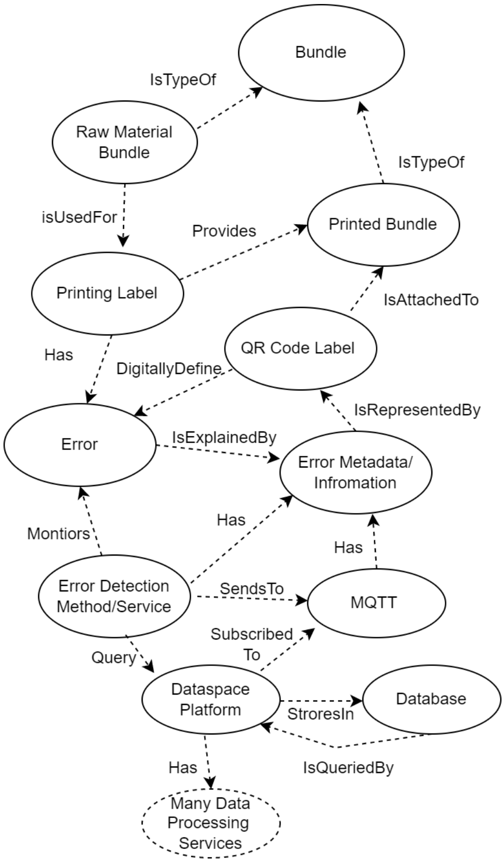


Figure 4.5: Semantic System Model

monetary opportunities provided by this Dataspace platform, based on the data associated with printing operations. As shown in Figure 4.3, the Dataspace platform consists of a wide range of data processing services, as explained below.

- **MQTT**: This is an industrial protocol standard [20] that provides standardized push, pull, subscribe or notify data operations. It is widely deployed in the IoT domain that needs lower resources and power to transmit data over the internet. Though the data model is not standardized at the application level, the communication protocol is standardised. Therefore, context-aware semantic models such as NGSI-LD can be used at the application level.

- **QR Code Service**: This service creates, generates and manages QR codes compatible with various programming languages such as python-qrcode, js-QR in javascript, etc. through open-source libraries. QR codes can be generated from JSON data models, derived from the earlier semantic process model, improving error tagging in manufacturing. This will digitalize the printing operations at each step we explained earlier. In addition, these QR tags can be pushed to the blockchain network to make them digitally traceable and transparent for all stakeholders [21].

- **Data Enrichment**: Inside the platform, this service will allow the modification of received data from the manufacturing unit, e.g., to add a timestamp when the bundle gets its error or when it gets delivered to the customer or error reference modifications or semantic adaptation.

- **Data Storage**: This service allows the data to persist for history and real-time operations among different stakeholders within or across customers of the SME.

- **Semantic Context Broker**: This provides context awareness over the data defined as per the semantic model—for example, the error, length, bundle, customer and delivery relationship. The stakeholder can query the data with a specific context. For example, a customer can ask - *How many and where the errors were when the ordered bundle was printed?*. Under such a scenario, the knowledge base created by a semantic context broker will yield the corresponding query result in a much simplified manner.

- **Data Analytics**: These AI/ML-driven data analytical services can assist SME and customers in their decision-making process. For example, questions like *how many errors have occurred in the last 6 months and how much material has been wasted due to different error types?* can be answered through these services efficiently, and estimated predictive analysis can also be performed.

- **Security and Governance**: This aspect is very important, especially when data and multiple stakeholders are involved. On top of that, GDPR compliances are stringent in Europe for manufacturing units to comply with to avoid hefty fines. This can be implemented using smart contracts among stakeholders, leveraging Blockchain technology such as HyperledgerFabric [22]. This will enable trust through transparency, immutability, digital traceability and tamperproof printing operations. In addition, this service will also provide identity, authentication and authorization management functionality.

- **Resource Management**: This typical ERP-related service manages resources at different operational stages. In addition, this service can also host the responsibility to manage the edge and cloud-level resources consumed by instantiated or orchestrated services. Here, oper-source ERP systems such as Odoo [23] can be very helpful for SMEs to start with.

- **Data Visualization**: This service will provide data visualisation in different formats such as bar charts, line charts, pie charts, heatmaps, 3D charts, Scatter plots, Gantt Charts, etc. Here, open-source tools like

Grafana, Elastic Search Logstach Kibana (ELK) stack and industrial tools like Power-BI are heavily used.

- **Integration Service**: This service will provide the required interfaces or Software Development Kit (SDK) to integrate with the Dataspace platform. This will be needed when the SME, under its customer service initiative for its products, say smart bundle, wants to share its operational data with the customer at the customer end so that the customer can align its processing or planning operations accordingly or for other purposes.

- **API Gateway**: This acts as a proxy gateway for the backend microservices running in a distributed environment and thus provides a single point of interaction for the SME or the external world, including customers. This can be implemented using production-grade open-source such as Nginx.

This dataspace platform is recommended to implement following the distributed microservices architectural design approach. This allows the services to be realized at the edge with the IoT sensing capabilities for data generation and processing to optimize bandwidth usage and reduce latency at the edge, i.e., the on-premises environment of the SME. In addition, this can be extended further to integrate the digital twin functionality to achieve operations validation during command and control, higher levels of efficiency, quality, and flexibility. This platform can be deployed on-premises as a private edge cloud and can leverage resources at the cloud level, following the Cloud-Edge continuum hybrid approach [24]. We suggested starting with the edge level deployment to re-use the existing infrastructure to deploy edge cloud leveraging open source Kubernetes-based bare metal microservices oriented containerized approach [25]. Later, as required, it can expand to the public cloud, e.g., AWS, MicrosoftAzure, Google Cloud Platform, etc. This platform enables value-added services across the value chain of the SME by enhancing the label printing process through optimized data-driven real-time operations, ensuring data integrity, and providing powerful analytics and visualization capabilities. This will empower the SME and its customers to make informed decisions, transparency, trust and drive efficiency in label printing operations.

4.5 Conclusion and Future Work

In this study, we have provided insights on business analysis and digitizing approach for Industrial R2R Label Printing Manufacturing SME in Denmark as part of the Manufactory project. We observed that the SME, operating on a global scale with substantial printing operations and recently installed heavy machinery, faces significant challenges in upgrading its existing infrastructure due to various constraints such as capital investment, resources and affordability. Our approach focused on understanding the intricacies of the SME's operational processes, leading to valuable insights. As a result, we did the business process mapping that provides a better and more precise understanding of enabling data and digitalization points in the existing environment. We identified numerous challenges, including productivity enhancement, error management, communication inefficiencies, waste reduction, compliance requirements, and the potential for smart products through platform-based data augmentation etc. To address these challenges, we narrowed our focus to error management within printing operations and presented a semantic process and system model. We addressed the error management challenge specific to printing operations that can occur for many reasons, e.g., inconsistent material, colour, speed, alignment, etc.) through the induction of digital processes around the same. In addition, we have also proposed a process-mapped architectural solution based on a semantic model that captures the business and technical level aspects concurrently. This proposed solution implementation approach is also explained based on well-defined standards, protocols and an open-source tools ecosystem that requires minimal investments and risks. As an outlook, we would like to measure the impact and value addition of the proposed solution on SME's printing operations to enable digital transformation in business productivity enhancements, reduction in waste material, and expansions through data-driven sustainable growth.

References

[1] Rojko, A., "Industry 4.0 concept: Background and overview.," *International journal of interactive mobile technologies*, vol. 11, no. 5, 2017.

[2] Ghasempour, A., "Internet of things in smart grid: Architecture, applications, services, key technologies, and challenges," *Inventions*, vol. 4, no. 1, p. 22, 2019.

[3] Xu, L. D. and Duan, L., "Big data for cyber physical systems in industry 4.0: A survey," *Enterprise Information Systems*, vol. 13, no. 2, pp. 148–169, 2019.

[4] Qin, J., Liu, Y., and Grosvenor, R., "A categorical framework of manufacturing for industry 4.0 and beyond," *Procedia cirp*, vol. 52, pp. 173–178, 2016.

[5] Xu, X., Lu, Y., Vogel-Heuser, B., and Wang, L., "Industry 4.0 and industry 5.0—inception, conception and perception," *Journal of Manufacturing Systems*, vol. 61, pp. 530–535, 2021.

[6] Masood, T. and Sonntag, P., "Industry 4.0: Adoption challenges and benefits for smes," *Computers in Industry*, vol. 121, p. 103 261, 2020.

[7] Thomsen, S., *The Danish industrial foundations*. Djøf Forlag, 2017.

[8] Morris, B. A., *The science and technology of flexible packaging: multilayer films from resin and process to end use*. William Andrew, 2022.

[9] Rao, C. H., Avinash, K., Varaprasad, B., and Goel, S., "A review on printed electronics with digital 3d printing: Fabrication techniques, materials, challenges and future opportunities," *Journal of Electronic Materials*, vol. 51, no. 6, pp. 2747–2765, 2022.

[10] Odeh, Y., "Bpmn in engineering software requirements: An introductory brief guide," in *Proceedings of the 9th International Conference on Information Management and Engineering*, 2017, pp. 11–16.

[11] Abusaq, Z., Zahoor, S., Habib, M. S., Rehman, M., Mahmood, J., Kanan, M., and Mushtaq, R. T., "Improving energy performance in flexographic printing process through lean and ai techniques: A case study," *Energies*, vol. 16, no. 4, p. 1972, 2023.

[12] Valdec, D., Hajdek, K., Vragović, L., and Geček, R., "Determining the print quality due to deformation of the halftone dots in flexography," *Applied Sciences*, vol. 11, no. 22, p. 10 601, 2021.

[13] Agency, D. E., "Denmark's climate status and outlook," Danish Energy Agency, Tech. Rep., 2022.

[14] Lamperti, S., Cavallo, A., and Sassanelli, C., "Digital servitization and business model innovation in smes: A model to escape from market disruption," *IEEE Transactions on Engineering Management*, pp. 1–15, 2023, Times cited: 3 event-title: IEEE Transactions on Engineering Management, ISSN: 1558-0040. DOI: 10.1109/TEM.2022.3233132.

[15] Angles, R., Thakkar, H., and Tomaszuk, D., "Rdf and property graphs interoperability: Status and issues.," *AMW*, vol. 2369, 2019.

[16] CIM, E., "Ngsi-ld cim 009," ETSI, Tech. Rep., 2019, Times cited: 3.

[17] Sporny, M., Longley, D., Kellogg, G., Lanthaler, M., and Lindström, N., "Json-ld 1.1," *W3C Recommendation, Jul*, 2020.

[18] Singh, P., Acharya, K. S., Beliatis, M. J., and Presser, M., "Semantic search system for real time occupancy," in *2021 IEEE International Conference on Internet of Things and Intelligence Systems (IoTaIS)*, IEEE, 2021, pp. 49–55.

[19] Solmaz, G., Cirillo, F., Fürst, J., Jacobs, T., Bauer, M., Kovacs, E., Santana, J. R., and Sánchez, L., "Enabling data spaces: Existing developments and challenges," in *Proceedings of the 1st International Workshop on Data Economy*, 2022, pp. 42–48.

[20] Standard, O., "Mqtt version 5.0," *Retrieved June*, vol. 22, p. 2020, 2019.

[21] Singh, P., Holm, K., Beliatis, M. J., Ionita, A., Presser, M., Wolfgang, P., and Goduscheit, R. C., "Blockchain for economy of scale in wind industry: A demo case," in *Global IoT Summit*, Springer, 2022, pp. 175–186.

[22] Srivastava, A. and Desai, Y., "Performance analysis of hyperledger fabric based blockchain for traceability in food supply chain," in *2021 IEEE 2nd International Conference on Technology, Engineering, Management for Societal impact using Marketing, Entrepreneurship and Talent (TEMSMET)*, IEEE, 2021, pp. 1–5.

[23] Li, C., Mantravadi, S., and Møller, C., "Aau open source mes architecture for smart factories–exploiting isa 95," in *2020 IEEE 18th international conference on industrial informatics (INDIN)*, IEEE, vol. 1, 2020, pp. 369–373.

[24] Lordan, F., Lezzi, D., and Badia, R. M., "Colony: Parallel functions as a service on the cloud-edge continuum," in *European Conference on Parallel Processing*, Springer, 2021, pp. 269–284.

[25] Orive, A., Agirre, A., Truong, H.-L., Sarachaga, I., and Marcos, M., "Quality of service aware orchestration for cloud–edge continuum applications," *Sensors*, vol. 22, no. 5, p. 1755, 2022.

Chapter 5

Paper III - Meta Standard Requirements for Harmonizing Dataspace Integration at the Edge

III

Parwinder Singh, Nidhi, Asim Ul Haq, Michail J. Beliatis

Originally published in *IEEE Conference on Standards for Communications and Networking (CSCN).* Nov 2023, pp. 130–135. DOI: 10.1109/CSCN60443.2023.10453211

Abstract

Emerging technologies such as the Industrial Internet of Things (IIoT), hyper-connected societies, and cyber-physical systems in Industry 5.0 (I5.0) have increased data volume substantially. Conventionally, the data process and analysis are carried out within the boundaries of organizations that can be perceived as edges closer to the data sources, resulting in the generated values stuck within the boundaries of organizations. However, this data can enable a cross-organizational data pool, based on which lots of semantics context-driven services can be developed and monetized to harness the collective power of data. This is precisely what a Dataspace (DS) leveraging distributed edge-coordinated ecosystem can offer to build a value chain on shared services and reusable data across organizations, domains, systems and devices. However, various challenges need to be addressed to achieve the same, such as lack of participatory motivation, heterogeneity due to the integration interfaces at different levels of data management operations and diverse standards available to perform the same target operation. Therefore, this study proposes a model for motivating entities to open up for data integration and build a services-driven cross-boundary value-chain system through a coordinated edge-driven DS ecosystem. In addition, it introduces the modified data life cycle, emphasising the integration phase and the integration harmonization ecosystem focusing on standards-driven heterogeneity harmonization, which demands Meta-Standard development and related guidelines.

5.1 Introduction

The synergy of data and enabling technologies has resulted in an abundance of products and related services with unprecedented speed. Diverse organizations, including service providers, manufacturing, digital streaming platforms, etc., rely on data analysis [1] to gain insights into the modern digitalized business landscape. Data collected upfront for an active business can create numerous passive market opportunities. However, the data belonging to a particular entity is often trapped within its boundary. The entity, here, represents any system, platform, process or organization with digital data management capabilities that handles all aspects of the Data Life Cycle (DLC), i.e., generation, identification, acquisition, analysis, usage, access, storage, and destruction [2]. In this study, the terms - *entity* and *organization* have been interchangeably used. Holding on to the in-house generated data to secure an entity's interest limits its growth graph and makes it unsustainable in the long run. Alternatively, data, if shared and reused, can unfold boundless market opportunities to enable new products and services besides enhancing existing businesses and processes [3]. Thus, whenever two or more entities desire to share, exchange and reuse their data and associated enabling services, a vital Research Question (RQ) is triggered:

Why, What, and How - data and associated services to be opened up for an entity?

Opening up the data beyond its organizational boundary unlocks huge possibilities for co-innovation and co-creation of values through sharing and reusability. However, making data shareable among other counterparts is still restricted or subject to closed solutions at organizational levels. There are government policies to protect and regulate data usage, like the Data Governance Act (DGA) [4] and General Data Protection Regulation (GDPR) [5], to set data availability and shareability guidelines to induce reusability and safeguards personal data privacy, respectively. The challenge in facilitating data integration lies in defining what data to share, which formats or standards to follow, and how to generate monetary benefits from sharing. Often, this gap results in the adoption of closed solutions involving significant investments in time, resources, and capital by the engaged entities [6]. Thus, there is a dire need to address integrating entities' challenges to agree upon common grounds for data sharing and reusability to explore cross-business value chains [7]. This can also be seen as the requirement to integrate data bodies among multiple entities with a standardized method or mechanism, which currently seems missing.

In this study, we propose to address the standard integration challenges by realizing the Dataspace (DS) concept [8] at the edge. DS provides a conceptual or virtual contextual environment where data can be accumulated, processed, transformed, or searched by various participating entities in varying contexts. Thus, this can be perceived as a "DS" created over a distributed and coordinated edges (i.e. representing entities) environment wherein it enables users to perform dynamic data operations in a controlled manner. However, building such a DS ecosystem has many challenges, such as a lack of motivation to participate, heterogeneity at different integration levels of data operations, and diverse standards available to perform the same target operation. To facilitate cross-domain data reusability and to mitigate integration challenges, this study mainly contributes to the following:

- Proposes data sharing motivation model for entities towards seamless DS integration development.
- Explains briefly the role of distributed coordinated edges in building DS in the local context.
- Proposes enhanced data life-cycle to enable the DS integration in a cross-boundary value-chain system.
- Demonstrates the need for Meta-standard and required guidelines and proposes an Integrated Harmonization Ecosystem to highlight heterogeneity due to the diverse use of standards.

The rest of the paper is organized as follows: Section II will briefly cover the background and adopted methodology, while Section III will elaborate on our depiction of data lifecycle and DS. Further, Section IV presents the system model, Section V establishes the essentials of harmonization in a DS integration ecosystem, and Section VI lays down the roadmap for standardization efforts. Finally, Section VII will conclude the paper with future research directions.

5.2 Background and Methodology

In reusing and sharing data, the integration aspect plays in power and controls the market [9]. Thus, the DS ecosystem emerges as a promising paradigm for harnessing the collective power of data by enabling organizations to transcend traditional data silos and fostering horizontal data sharing based on contextual semantics [8], [10]. The International Data Space (IDS) and GAIA-X initiatives (by EU) have developed reference architectural frameworks and guidelines for fostering a robust data economy through cross-domain data sharing via building DS ecosystem [9]. However, there is still a void to develop such DS in a local or regional context, e.g. at the edge, and much work still needs to be done on harmonizing the standards. Furthermore, data heterogeneity, interoperability and generated value are major hurdles in the development of DS [10]. Emerging techniques like semantic modelling, preprocessing, Natural Language Processing (NLP), etc. [11] can mitigate the issue of data interoperability and interface heterogeneity among integrating entities. However, there are still unaddressed challenges concerning data heterogeneity due to the diverse use of standards to implement the same target.

The possibility of technology convergence at the edge (close to the data source) has opened up a new market based on constantly evolving user requirements [12]. The fusion of technologies brings significant integration challenges across system, interface, and application levels, such as ensuring heterogeneity, interoperability, data privacy and security [13]. Additionally, establishing a distributed DS ecosystem, leveraging coordinated edges as the first integration point requires careful consideration [14] and addresses edge-related critical questions like computing resources, data protection, standards, technology convergence, and the distinct advantages of edge over conventional cloud solutions [15].

To build DS with an open-up approach to facilitate data-driven value chains, it's crucial to have motivated entities willing to collaborate [10]. We have proposed a system model in section 5.4 that addresses the "WHY" aspect of the RQ for stating DS entities' motivation. In this context, each entity can act as an edge to offer the necessary computing infrastructure and connectivity at different integration levels for the DS. These entities own and control their data and associated services, aligning with DS goals. By integrating multiple such edges, we can enable a coordinated edge-driven ecosystem, advancing the DS landscape [16].

Moreover, for an efficient DS integration, it is paramount to define data-sharing norms, interfaces, and protocols in addition to addressing challenges in heterogeneity, interoperability, service integration, and trust. Thus creating a void that can be filled with the notion of *Meta-Standards* aiming to harmonize diverse standards, as explained further in section 5.6. This can offer a promising solution for facilitating interoperability and seamless integration for cross-domain DS applications.

Our overarching approach aligns with Design Science Research methodology [17], focusing on practical problem-solving and iterative development with feedback. Our methodology combines various approaches, including digital business development methods and tools [3], St. Gallen magic triangle and the Value Chain model to define our research context. A thorough literature review established the foundation of our proposals in line with the related research, followed by probabilistic mathematical modelling to develop the system model to address the participating entity's motivation. Finally, we adopted the architectural design approach to propose an integrated and harmonized DS ecosystem.

5.3 Data Life Cycle and Dataspace

Data has the potential to generate and discover boundless opportunities to enable new products and services together with enhancing existing businesses and processes. Data traverses through its life cycle, i.e., generation, acquisition, processing, analysis, visualization, storage, and destruction [2]. The DLC typically illustrates the scope of data from its generation to destruction. Thus, to incorporate data reusability, we are enhancing the definition of the DLC journey by including the integrating phase supported by data sharing and reusability. As illustrated in Figure 5.1, data from one entity (Org 1) can enter the DLC of another entity (Org 2) under agreed norms and consensus [18]. Thus, this enhanced DLC enables all entities to innovate value chains on top of the shared resources instead of reinventing the whole wheel. For example, a product-based company manages data during the life cycle journey that initiates at the customer end and terminates at the company end, providing many data insights for the customers, company and stakeholders.

Multiple entities willing to share can open up their data leveraging this enhanced

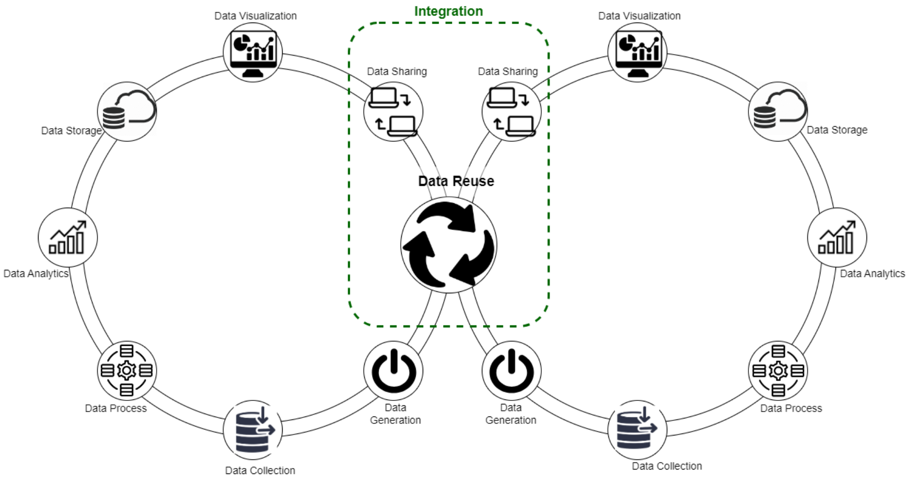


Figure 5.1: Data Lifecycle Adaptation to Sharing and Reusability in Cross-Organization domain

DLC. However, opening up the data also brings forth many challenges, such as sharing or exchanging data beyond organizational confines with granular access control and generating mutual or exclusive service-driven values on top of the shared data along with related privacy, governance, and ownership. As a solution, building a DS ecosystem at the edge can allow swift data interaction with tailored guidelines among the entities involved. In this context, the edge can be defined as an entity that holds and provides access to the data and associated services. Therefore, if the organization holds the data within its organizational boundary, it will be considered an edge or an organizational edge. Each edge has the power to generate, collect, process and compute data. Each edge owns its data and manages it through Create, Read, Update and Delete (CRUD) operations in the DS ecosystem, enabling the development of a DS combining edges. This edge-enabled DS can support reusing and sharing converged resources (i.e., systems, infra, technologies stack), data and services to generate a value chain to get mutual gains by integrating cross-domain resources. It also emphasises that the data will be stored by its owner (an edge) permanently, but for cross-entity value gain, the data can be shared temporarily or for longer run through different consensus (e.g., only read operation) and business models (e.g., the pay-as-you-go).

It is essential to address challenges like trust, contextual processing, semantic data search and relevant services enablement. For example, Blockchain and contextual deployment can provide trust for resource sharing among value chains and using semantic standards and technologies, e.g., RDF, NGSI-LD, etc., can provide semantic search [19]. Overall, DS at the edge, built on resource convergence, tech-enablers, co-creation, and mutual interests, answers the RQ's

HOW part to enable data sharing and reusability to produce mutual gains based on relevant values. The following section will explain the motivation behind such mutual gains through the system model emphasizing the entities' value chain development to understand the RQ's *WHY* part to adopt an open-up approach.

5.4 System Model - Participatory Motivation in Dataspace Development

The system model is assumed to have two organizational entities, i.e., O_1 and O_2, with values X_1 and X_2, respectively, that are interested in developing value chain by sharing values based on their data and associated services, if they have mutual gains from the mutual values Z, as illustrated in Figure 5.2. Without loss of generality, Z can be represented mathematically as;

$$Z = f(X_1, X_2, ..., X_n), \tag{5.1}$$

Where n represents the number of organizations interested in DS integration and X_i for $i \in \{1, 2, ..., n\}$ can be defined as the function of the underlying value set of the respective organization. Mathematically it can be represented as;

$$X_i = f(D_i), \forall i \in \{1, 2, 3, ..., n\}, \tag{5.2}$$

Where D_i represents the underlying data set of the O_i organization wherein the nature and quality of data influence the value. These respective underlying data sets are further defined as the binomial function of services S_i within the organization O_i, and their boundaries B_i where D_i and S_i are trapped. Therefore, it can be represented as;

$$D_i = f(B_i, S_i) \forall i \in \{1, 2, 3, ..., n\}, \tag{5.3}$$

Hence, from the analysis of equation 5.2 and 5.3, it can be concluded that the values X_i of organization O_i are dependent on the boundary B_i, services S_i, and underlying data set D_i of the *ith* organization. By the stated definition, equation 5.2 and 5.1 can be reformulated as;

$$X_i = g(B_i, Si) \forall i \in \{1, 2, 3, ..., n\}, \tag{5.4}$$

$$Z = h(B_i, Si) \forall i \in \{1, 2, 3, ..., n\}, \tag{5.5}$$

Where the function $g(.)$ is defined as the function of the function of boundary and services, i.e., $f(f(.))$, and $h(.)$ is defined as $f(g(.))$. These equations signify that the values X_i are generated based on the relevant underlying data D_i using the services S_i within their respective organizational boundaries B_i. This emphasizes that each organization has exclusive access to the data and services within its boundaries, and the values X_i generated by organization O_i, are the results of their specific resources and capabilities. In addition, this also highlights the idea that the value generated by each organization is intricately tied to its unique set

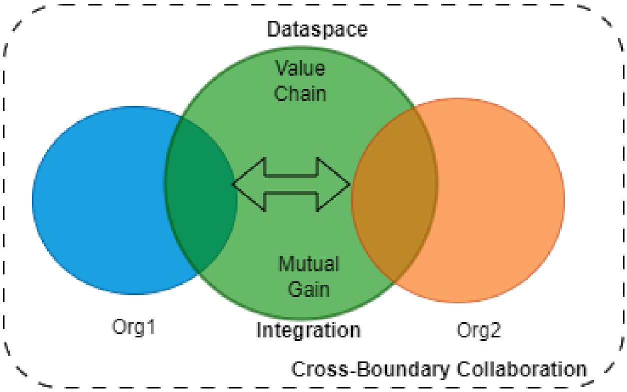


Figure 5.2: DS Integrated Boundaries Driven Value Chain

of data and services, providing an opportunity and emphasizing the importance of data integration for achieving mutual benefits Z through collaboration, sharing and re-usability by adopting a data opening up approach as highlighted in this article, keeping the GDPR rule in consideration [5]. Therefore, this motivates us to focus on data integration and space creation based on the proposed methodology, which makes the DS more realistic. Therefore, we also need to define the integration function.

To understand the integration function of the data, let us say that the values X_i generated by the organization O_i, are statically independent but not mutually exclusive within the shared DS. Hence, it can be stated that knowing values X_1 generated by organization O_1 can reveal some information regarding X_2 generated by O_2. Generally, the random values from any organization can be mathematically represented with the help of the chain rule;

$$\begin{aligned} P(\mathbf{X_n} = x_n, \ldots, \mathbf{X_1} = x_1) = \\ P(\mathbf{X_n} = x_n | \mathbf{X_{n-1}} = x_{n-1}, \ldots, \mathbf{X_1} = x_1) \\ P(\mathbf{X_{n-1}} = x_{n-1}, \ldots, \mathbf{X_1} = x_1), \quad (5.6) \end{aligned}$$

Where the $\mathbf{X_i}$ represents the random variable (RV) for the value X_i generated by organization O_i, and x_i is the event occurred or observed of RV $\mathbf{X_i}$ for all $i \in \{1, 2, 3, \ldots, n\}$.

In the proposed system model, we assume two organizations, O_1 and O_2, each generating the Gaussian Distributed values X_1 and X_2 with mean μ=0 and variance σ=1, respectively. This distribution model is ubiquitous in literature due to the finite σ of Dataset and the increase in probabilities as getting closer to μ. It is observed that the integration function on which both organizations agreed upon to produce mutual gain is 37.4%, showing that the relevant services between O_1 and O_2 are more likely to be integrated, as shown in Figure 5.3. This explains that both organizations contribute towards generating mutual value from each other's datasets. The assumption of the mutual exclusiveness

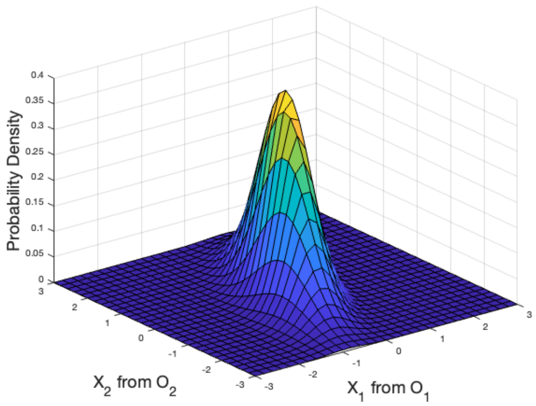


Figure 5.3: Probability Density Functions of X_1 and X_2

of the data of O_1 and O_2 shows that the contribution in mutual benefits is always higher than the individual gain operating in silo. Finally, we can say that the integration function provides an opportunity to develop a win-win business model among entities through the DS ecosystem that accumulates and abstracts data and relevant services across the organizations.

This approach motivates organizations to combine their strengths to generate a mutual value chain through data sharing and reusability across partner organizations. There can be many other reasons based on which an entity can decide to open up its data and relevant services, some of which are as follows:

- Opportunities to find new values throughout the collaborative value chain.
- Gain knowledge and insights from shared resources and enabling services.
- Seeks contribution through open-source collaboration.
- Enable customers to participate in their respective product or service provider insights actively.
- Intentions to extend the market base by offering open data and services to yield new business revenue models.

There is no one-fit answer to this question; ultimately, it hinges on entity management's discretion. We strongly believe the decision to open up is a step towards business expansion and sustainable growth that can be driven out of shared values in the value chain via innovation, co-creation, and collaboration efforts. This is significant, especially in modern times, to sustain, grow and get an edge in the competitive market.

5.5 Dataspace Integration Harmonization Ecosystem

Integrating data and associated services across organizations poses significant challenges due to multiple processes, stakeholders, systems and services involved. Figure 5.4 proposes an integration harmonization ecosystem to offer the possible integration mechanisms among organizations. The vertical column indicates an organizational boundary consisting of the entity's system, services, data, application and business model relevant to its operations. However, the horizontal row shows the cross-entity integration at different levels, creating opportunities for developing a value chain to generate mutual gain through the DS ecosystem. Each organizational entity typically handles a data system wherein different services run and provide a way to perform different operations with the data. The data is processed in its life cycle using its defined model with the relevant context and delivers service-driven values leveraging relevant application interfaces. The context can be served in local (i.e., intra-organizational context) or semantics-driven linked data context (i.e., inter-organizational context), which presents cross-domain data integration leveraging a wide range of semantic standards such as NGSI-LD [19].

The application delivers specific values to the end customer based on the data and accessible services. For example, we could see three use cases in this scenario serving different applications in different domains. First, the wind turbine industry domain use case aimed at integrating the supply chain information for different events in the turbine operation's life cycle [19]. Second, the manufacturing domain use case where label printing error operations are digitized through real-time IoT-based digital intervention [20]. Third, the smart campus environment in spatial infrastructure domain based real-time room booking system [21]. The typical data functions in each use case follow the DLC journey explained already in section 5.3. An important point from these three use cases and the proposed ecosystem is that data is typically trapped in vertical columns following specific organizational boundaries. So, when we say to open up the data, it means to either expand this boundary or allow data to go beyond its organizational boundary. Therefore, the horizontal slicing of an organizational boundary is shown wherein data and service stack offer integration at different levels wherever it is deemed fit.

Generally, the trend is to follow the integration at the service level, where Application Programming Interfaces (APIs) are defined for integration. However, the problem is that each entity in its environment will define the service/API structure (methods and endpoints) following its own needs and naming conventions. Thus, creates a heterogeneous environment, though it follows well-defined standards at different levels. It can be observed from Figure 5.4 that many standards can be defined for the same purpose at different integration levels. This gives birth to standards-driven heterogeneity. Hence, emphasizing the need to harmonize the usage context of standards, i.e., define the standard of standards (i.e., Meta-Standards), explained further in the next section.

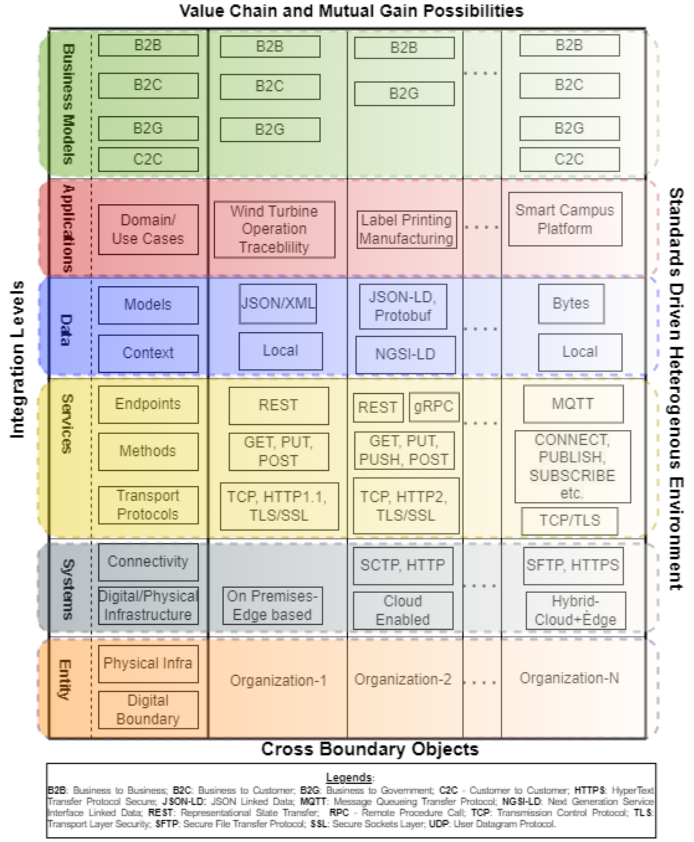


Figure 5.4: Integration Harmonization Ecosystem

5.6 Standardization Focus For Meta-Standards

Data sharing under integration to generate value-chain for mutual gain can be handled through its life cycle-associated events based on defined usage context and consensus between the organizations that will eventually become a DS ecosystem. Each organization has a layered stack of enabling technologies for data processing, as shown in Figure 5.4. However, they all can be heterogeneous, though they follow well-defined standards. The question is, *Why is this heterogeneity still?* The reason is that there are a lot of standards, and there are different ways and choices to implement the same thing. For example, an API can be implemented using REST, RPC, gRPC, and MQTT as shown in Figure 5.4, and each can have different data models such as JSON, JSON-LD, NGSI-LD, etc. and access methods. Therefore, this also points toward harmonization of standard usage requirements. For example, to define standard guidelines that guide on what suits best under what scenario.

5.6.1 Objectives

As explained above, the requirement of standardization efforts has emerged to deal with heterogeneity that arises due to the usage context of standards. Perhaps a Meta Standard approach can be defined where the integration-focused handshake processes, protocols, methods, payload, data models, semantic contexts, and interfaces can be put on the development agenda. We strongly believe the RQ's Why, What and How parts are significant in addressing heterogeneity among standards at different integration levels to develop the DS with local or global boundaries to generate a value chain with mutual gain. This Meta-Standard approach or guidelines are expected to provide seamless integration and interoperability between different data integration entities. Based on our study, we could see that this can be addressed using the edge-enabled DS concept. However, this needs further in-depth investigation. We have already identified high-level guidelines in this direction that can be worked out for harmonization requirements. These are given below:

- Defining DS ecosystem in the modern edge context and related integration interfaces that entity can use for different CRUD operations with different access controls at granularity (e.g. API, method or attribute) level.
- Address concerns of data sharing specific questions such as - what to share, how to share, how much to share and how to reach a consensus level for sharing in a digital traceable manner. Here, Blockchain-defined contracts [19] can be very handy to define the resources sharing access, ownership, secure exchange of information, transparency, building trust and traceability in a multi-stakeholder environment.
- Identification of challenges and possibilities at different integration levels with relevant cost-benefit analysis.
- To provide implementation guidelines for building DS integration ecosystem at a small scale, e.g., this study provided insights on building DS at the edge and in a distributed manner.
- To identify dynamicity-supported integration methods which can accommodate the heterogeneity neutrally.
- Define the DS integration's relevant data, ontology and semantics model.
- Address concerns related to data protection and secure operations.
- Define interface usage guidelines in dynamic contexts for different integration stakeholders.
- Build exemplary scenarios to build value chains based on data and associated service sharing.
- Define the role and interfaces for semantic context, standards, and linked data awareness operations involved.

- Define the role of convergence of different technologies supporting integration at the edge to achieve seamless DS integration.
  These are tentative lists of items that can grow further during the discussions around standardization efforts.

5.7 Conclusion

This study has focused on opening up, sharing and reusability of cross-organizational data and services by building an edge-driven DS Integration Ecosystem. Multiple decision factors for an organization need to be considered while opening up its resources for sharing based on answering Why, What and How factors as defined in the main research question. We have presented a system model to answer the Why and What part, and a Harmonized Integration Ecosystem to answer the HOW part. During this investigation, it was found that integration is possible at multiple levels with the possibility of multiple standards for an organization. This gives rise to a standards-driven heterogeneous environment, which needs to be addressed through Standardization efforts and guidelines in future. As an outlook, we would like to expand this study to include further details on different approaches to harmonize the use of heterogeneous standards, implementation prototypes and developing decision support methods that can assist in building a cross-business value-chain based on the presented system model. Together, we have contributed a knowledge base to develop a cross-boundary integrated value-chain system through the DS ecosystem in Industry 5.0.

References

[1] Aalst, W. van der, Hinz, O., and Weinhardt, C., *Big digital platforms: Growth, impact, and challenges*, 2019.

[2] Banerjee, R. and Hofmann, B., “Corporate zombies: Anatomy and life cycle,” *Economic Policy*, vol. 37, no. 112, pp. 757–803, 2022.

[3] Singh, P., Beliatis, M., Jakobsen, E. M., and Presser, M., “Digital business IoT maturity patterns from eu-IoT ecosystem,” in *Shaping the Future of IoT with Edge Intelligence: How Edge Computing Enables the Next Generation of IoT Applications* (River Publishers Series in Communications and Networking), Sofia, R. C. and Soldatos, J., Eds., River Publishers Series in Communications and Networking. Germany: River Publishers, Oct. 2023, ch. 18, pp. 377–403, Available. DOI: 10.13052/rp-9788770040266.

[4] Baloup, J., Bayamlıoğlu, E., Benmayor, A., Ducuing, C., Dutkiewicz, L., Lalova-Spinks, T., Miadzvetskaya, Y., and Peeters, B., “White paper on the data governance act,” 2021.

[5] Goddard, M., "The eu general data protection regulation (gdpr): European regulation that has a global impact," *International Journal of Market Research*, vol. 59, no. 6, pp. 703–705, 2017.

[6] Bressanelli, G., Adrodegari, F., Perona, M., and Saccani, N., "Exploring how usage-focused business models enable circular economy through digital technologies," *Sustainability*, vol. 10, no. 3, p. 639, 2018.

[7] Orlova, A., Nogueira, R., and Chimenti, P., "The present and future of the space sector: A business ecosystem approach," *Space Policy*, vol. 52, p. 101 374, 2020.

[8] Curry, E., Derguech, W., Hasan, S., Kouroupetroglou, C., and Hassan, U. ul, "A real-time linked dataspace for the internet of things: Enabling "pay-as-you-go" data management in smart environments," *Future Generation Computer Systems*, vol. 90, pp. 405–422, 2019.

[9] Association, I. D. S., *Gaia-x and ids. position paper, version 1.0 01*, 2021.

[10] Solmaz, G., Cirillo, F., Fürst, J., Jacobs, T., Bauer, M., Kovacs, E., Santana, J. R., and Sánchez, L., "Enabling data spaces: Existing developments and challenges," in *Proceedings of the 1st International Workshop on Data Economy*, 2022, pp. 42–48.

[11] Kumar, G., Basri, S., Imam, A. A., Khowaja, S. A., Capretz, L. F., and Balogun, A. O., "Data harmonization for heterogeneous datasets: A systematic literature review," *Applied Sciences*, vol. 11, no. 17, p. 8275, 2021.

[12] Baldwin, R., *The great convergence: Information technology and the new globalization*. Harvard University Press, 2017.

[13] Farahani, B., Firouzi, F., and Luecking, M., "The convergence of iot and distributed ledger technologies (dlt): Opportunities, challenges, and solutions," *Journal of Network and Computer Applications*, vol. 177, p. 102 936, 2021.

[14] Kalogeropoulos, I., Vlontzou, M. E., Psaromanolakis, N., Zarogianni, E., and Theodorou, V., "Edgeds: Data spaces enabled multi-access edge computing," *arXiv preprint arXiv:2304.05966*, 2023.

[15] Martella, C., Longo, A., Zappatore, M., and Ficarella, A., "Dataspaces in urban digital twins: A case study in the photovoltaics," 2023.

[16] Hernandez, J., McKenna, L., and Brennan, R., "Tikd: A trusted integrated knowledge dataspace for sensitive healthcare data sharing," in *2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC)*, IEEE, 2021, pp. 1855–1860.

[17] Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S., "A design science research methodology for information systems research," *Journal of management information systems*, vol. 24, no. 3, pp. 45–77, 2007.

[18] Mertens, C., Alonso, J., Lázaro, O., Palansuriya, C., Böge, G., Nizamis, A., Rousopoulou, V., Ioannidis, D., Tzovaras, D., Touma, R., *et al.*, "A framework for big data sovereignty: The european industrial data space (eids)," in *Data Spaces: Design, Deployment and Future Directions*, Springer International Publishing Cham, 2022, pp. 201–226.

[19] Singh, P., Holm, K., Beliatis, M. J., Ionita, A., Presser, M., Wolfgang, P., and Goduscheit, R. C., "Blockchain for economy of scale in wind industry: A demo case," in *Global IoT Summit*, Springer, 2022, pp. 175–186.

[20] Singh, P. and et al., "Digital dataspace and business ecosystem growth for industrial roll-to-roll label printing manufacturing: A case study," in *SENSORCOMM 2023, The Seventeenth International Conference on Sensor Technologies and Applications*, Thinkmind Digital Library, Porto, Portugal, Sep. 2023, ISBN: 978-1-68558-090-2.

[21] Singh, P., Acharya, K. S., Beliatis, M. J., and Presser, M., "Semantic search system for real time occupancy," in *2021 IEEE International Conference on Internet of Things and Intelligence Systems (IoTaIS)*, IEEE, 2021, pp. 49–55.

Chapter 6

Paper IV - Enabling Edge-Driven Dataspace Integration Through Convergence of Distributed Technologies

Parwinder Singh, Michail J. Beliatis, Mirko Presser

Originally published in *Internet of Things Journal.* Nov 2023, pp. 130–135. DOI: 10.1016/j.iot.2024.101087

IV

Abstract

Dataspace and emerging technologies play a key role in developing value chain systems using cross-domain data, services and systems integration. Therefore, this study has conducted a comprehensive literature review for six years (2017-2022) on the convergence of Internet of Things (IoT), Artificial Intelligence (AI) and Distributed Ledger (Blockchain) technologies for supporting Dataspace integration efforts at the Edge. As an outcome, this study has identified relevant challenges that include heterogeneity, integration and interoperability, distributed security, trust, scalability, and resource management. It has also been found that very limited research covers the architectural aspects of distributed edge in the context of the convergence of technologies for Dataspace integration purposes. Therefore, this study has proposed an architectural framework - Distributed Edge Network Operations-oriented Semantic (DENOS) model that extends the traditional Cloud-Edge-Device architecture with three new layers - Semantic, Convergence, and Dataspace integration. In addition, the model leverages the power of semantic modelling (i.e., Processing, Service, and Data) context, which enables the model to have a dynamic implementation context to suit the diverse needs of target use cases. To showcase the validation of the model, a use case related to the digital traceable operation of the wind energy domain has been presented. The objective of the DENOS model is to enable Dataspace integration to build edge-enabled value chain networks. Thus, it contributes to secure and semantic integration using the convergence of resources and technologies, cross-domain collaboration, reusability and data-driven decision-making of resources.

[0]Department of Business Development and Technology, Aarhus University, Herning, 7400, Denamrk, parwinder@btech.au.dk

6.1 Introduction

At present, centralized cloud computing (also referred to as cloud layer) plays an important role in Industry 4.0 (I4.0) realization by delivering digital services in different sectors and domains [1]. The surge in Internet of Things (IoT) devices [2], which are going to be in billions, is expected to increase demand for resources at the cloud level significantly [3]. This will, in turn, raise the demand for scalability of services and computational resources such as CPU, storage, energy, security, maintenance, and operations management at cloud [3], [4]. In addition, it is evident (or inevitable/imminent) that it adds up to the carbon footprint, latency and reduced network efficiency due to large resources and energy consumption at cloud data centres [5] to handle the upstream flow of IoT data (devices to the cloud), at centralized location [6], [7]. Therefore, this poses the need to have decentralized and distributed architectural designs, such as edge computing, that are capable of addressing the raised concerns from a future perspective. However, an important question arises whether the industry or market is ready for the paradigm shift from cloud to edge computing.

In a recent qualitative survey [7], it has been revealed that more than 90% of respondents are interested in migrating their performance-sensitive applications from the cloud to metro edge computing locations. By 2025, it is estimated that the edge computing market in Europe will be worth $500 billion, with approximately 41.6 billion connected devices that will generate and process about 75% of enterprise data outside central data centres [8]. However, this does not mean that edge computing will be perceived as a replacement or alternative for cloud computing. Instead, edge computing can complement cloud computing by providing additional and distributed computing capabilities closer to the data sources to gain the benefits of enhanced throughput, real-time decision enabling, and latency saving [9]. For example, in many use cases, such as smart-city [10], or smart-home [11] related security or public safety applications, the size and volume of data (e.g. video format) are very high in order to send it to the cloud, but latency is very crucial at the same time to perform quick decisions of the incident. Therefore, in such use cases, the role of edge becomes more significant than the cloud. Thus, it is critical to have edge computing to complement/aggregate with cloud computing to offer sustainability through its distributed edge architecture [12], saving carbon footprints by offloading and efficient utilization of edge resources along with low latency and bandwidth saving [13].

Considering the future trend towards edge computing, transforming business models from cloud to edge computing is highly likely. Therefore, an edge-driven ecosystem can enable disruptive business models wherein stakeholders can collaborate directly, i.e. organization-to-organization represented by edge-to-edge ecosystem to develop win-win business models through shared value chains instead of competing [14]. The value chain in this study context means value driven out of shared data, resources, services and related integrated digital operation enablers. Dataspace can be very beneficial to realize such cross-organizational value chains. Dataspace represents a concept in information management that focuses on integrating and harmonizing data across various domains (within or cross-organization), systems, and applications [15]. It provides a unified view of data that can be used in multifaceted areas. It enables efficient sharing, reusability, co-creation, co-innovation, and monetization opportunities with provenance and governance controls over data within distributed environments. In the Dataspace context, the edge can be perceived as an organizational entity that offers the required computing infrastructure with real-time processing capabilities closer to the data sources beyond latency and bandwidth-saving factors. In addition, using edge also fulfils the core requirement of Dataspace, wherein data management and ownership control stay with the organization itself. [16].

The emergence of edge computing presents a significant opportunity to enable the edge-driven Dataspace ecosystem, leveraging the convergence of modern technologies to build new services and values among interested organizations and related stakeholders. In this context, an edge can serve as a crucial data integration link across distributed data sources (residing in silos or across organizational boundaries) for Dataspace integration applications. In addition, many other digital technologies can be combined at the edge to exploit relevant benefits. These technologies include Artificial Intelligence (AI) for automated decision-making, smart and quick connectivity based on 5G/6G Telco advancements, Next Generation Internet-driven smart IoT, and Distributed Ledger Technologies (DLT) for seamless security [4]. From an edge perspective, the advancement in technologies, networks and devices is also expected to present higher resource availability (usually distributed in nature) at the edge [17]. Such infrastructures are already available that have been used under Mobile Edge Computing (MEC) and ultrafast distributed networks such as 5G and beyond networks [18], [19]. Therefore, this allows resource-hectic technologies to function in real-time at the edge. Apart from that, technologies such as DLT and AI-enabled services enable secure distributed communication and intelligence on data, respectively, which are highly stable and beneficial in certain contexts [20].

The integration of these technologies can be utilized to provide smooth evolution towards distributed edge architecture from conventional centralized architectures such as cloud computing [21]. Therefore, the integration and convergence of next-generation infrastructure and technologies, such as IoT, DLT, Edge and AI-based Machine Learning (ML) (hereafter referred to as IDEAL), need to be explored to design optimized distributed edge-oriented solutions such as:

- Cloud-to-Edge evolution by bringing processing, and analytical capabilities closer to the data sources to offer sustainability (towards Industry 5.0) and lower operational costs in the long run.
- Semantic integration and interoperability among diverse data sources for Dataspace integration [22].
- Distributed security, trust and governance for sharing and reusability of resources for collaborative value chains.

- Reduction in the demand for high bandwidth and resources in central cloud networks.
- Low latency for real-time critical applications by establishing an edge-enabled service ecosystem.
- Low environmental carbon footprint by optimum utilisation of local edge resources through resource pooling.

In order to nurture the combined benefits of IDEAL integration and convergence, this study has comprehensively reviewed the literature on the convergence of IoT, Blockchain and AI technologies to support the Dataspace ecosystem at the edge. This study will provide an exhaustive literature review emphasising recent trends, architectural aspects, associated challenges, and gaps. We propose a Distributed Edge Network Operations-oriented Semantic (DENOS) framework to address the challenges at the design level through the integrated capabilities of IDEAL and semantics-driven system contexts. With the DENOS framework, we also intend to address the void in the existing architecture to back the convergence of technologies to develop an edge-enabled Dataspace value chain ecosystem.

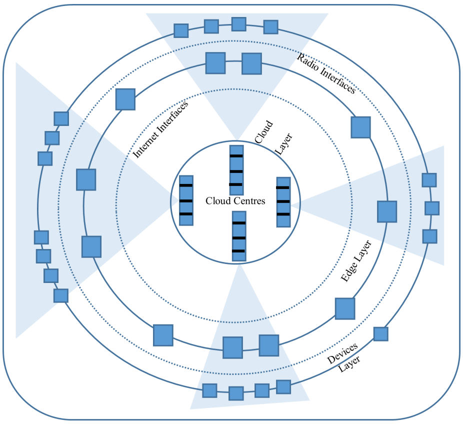


Figure 6.1: Cloud-Edge-Device Continuum

DENOS leverages semantics-driven target system contexts to deal with different challenges. The objective of the DENOS model is to enable Dataspace integration and efficient management of the resources in the Cloud-Edge-Device (CED) continuum and facilitate the sharing of services, data and resources across the edge-to-edge horizontal continuum for diversified domains and applications. Thus, it contributes to the building of edge-enabled value chain networks that can create novel business opportunities and promote innovation.

Figure 6.1 depicts a typical CED continuum environment, but here it also presents multi-relational entities (layers, systems, devices, stakeholders etc.) in a multi-tenancy-driven (e.g. multi-vendor or multi-operator) Dataspace environment. This environment includes multiple clouds (provided by various cloud providers, such as AWS, GCP, Azure, and OpenStack), multiple edge servers or gateways (provided by various vendors), and multiple devices (provided by various manufacturers or vendors) operating in a geo-spatially distributed environment. The CED continuum presents multiple relationships both horizontally and vertically with various stakeholders (business or service operators) in diverse domains, devices, edge systems, and cloud layers. This provides many opportunities to build Dataspace integration applications leveraging the CED continuum where resources (e.g. infrastructure), data, and services endpoints reside. Therefore, this CED continuum, shown in Figure 6.1, serves as the basis which we have extended it for our DENOS model.

This paper has been structured as follows: The first section has covered the introduction and motivation part. The second section will cover the literature review and findings on the convergence context of IDEAL technologies, relevant challenges and identified gaps. Section three presents a DENOS model, related system semantics context, and methodologies to tackle identified challenges, such as semantic context, optimized utilization of distributed resources (pooling and scheduling) and integration of edge network with Blockchain/DLT network to address edge-enabled Dataspace distributed security concerns. Finally, the paper provides a conclusion in section four.

6.2 Literature Reveiw

6.2.1 Methodology

Our literature review methodology has been shown in Figure 6.2 and consists of four phases, which include planning, quantitative, qualitative and analysis. In the planning phase, the research context is defined, which is expected to be achieved from the literature review. The research context for the literature review is shown in Table 6.1. The main goal was to identify the most prominent use of Blockchain, an AI-supporting IoT ecosystem at the edge and how they are offering values from a business and/or Dataspace integration point of view. In the quantitative phase, the literature has been searched as per the research context set in the planning phase. In the qualitative phase, papers were screened manually, based on inclusion, exclusion and quality criteria, to find the closest match for the given research context. Finally, based on the outcome of the quantitative and qualitative phase, relevant findings are analysed and summarized respectively. As an outcome of the quantitative phase, literature review data has been collected, analysed, and visualized in the form of trends illustrated in Figure 6.3 and Figure 6.4. To observe and analyze the trends, we referred to different scientific databases, such as Google Scholar, web-of-science and Scopus. We have compiled the findings over the last six years (Jan 2017 to December 2022). The data is pre-processed to find relevant, unique entries, and the data is augmented with additional columns (named IoT, AI, DLT and Edge) based on their relevant usage found in each of the publication entries. The usage context of technology has been analyzed based on the title abstract, and conclusion content basis.

Table 6.1: Literature Review Context

Research Question	How are Blockchain/DLT, IoT, Artificial Intelligence, and Edge computing being used in combinations, and What are the relevant challenges to integrating these technologies at Edge for Dataspace applications?
Search String	(("Blockchain" OR "DLT" OR "Distributed Ledger") AND ("edge computing" OR "edge") AND ("Internet of Things" OR "IoT") AND ("artificial intelligence" OR "AI" OR "machine learning" OR "ML") AND ("convergence" OR "integration") AND ("dataspace" OR "dataspaces integration"))
Inclusion Criteria	The selected studies shall include the following aspects: • the combinational use of DLT, IoT, and AI technologies at the edge. • the use of Blockchain or DLT as a security method in edge computing and IoT. • the role of AI/ML in edge computing and IoT. • Highlight the challenges and advantages of the integration of IDEAL technologies in different combinations, including general and relevant Dataspace integration aspects. • Proposal or discussion on relevant distributed edge architectures and methodologies for integration of these technologies.
Exclusion Criteria	Studies shall not be selected with the following aspects: • The focus is solely on the individual use of technology such as Blockchain/DLT without considering IoT, edge and/or AI. • The focus is solely on IoT and AI without any application of edge computing or Blockchain involved. • The studies found to be published before the year 2017.
Quality Criteria	Following two quality criteria involved during the filtering process in the qualitative phase. • Q1—Are the research objectives, methodology, experiments/-data and results/analysis defined and correlated? • Q2— Is the study matching to our research context?
Search Engine	Scopus, Web of Science, Google and Semantic Scholar
Results	Total 482; 44 (based on inclusive and exclusive criteria))

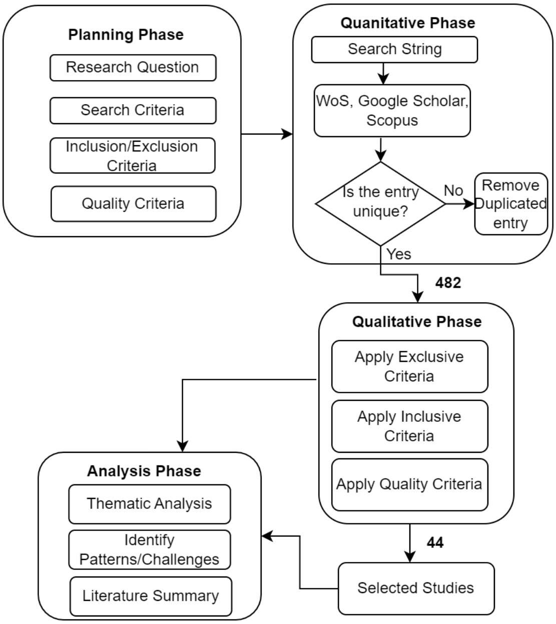


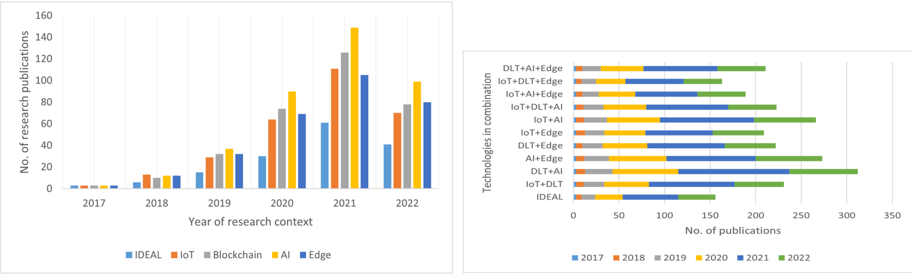
Figure 6.2: Systematic Literature Review Methodology

6.2.2 Technology Trends

Based on the outcomes of the quantitative phase, Figure 6.3a presents the usage trends of Blockchain/DLT, IoT, Edge, AI and IDEAL (all together) under the given research context. It is showing, as per the given research context, that AI is the dominating factor, followed by Blockchain, edge computing and IDEAL in terms of their usage trends. Here, the IDEAL reflects the use of all four technologies together. In addition, the research trend (for the last six years) in Figure 6.3b shows the emergence of DLT (Blockchain), AI, Edge and IoT in the combination of the doublet, triplet and all four together. The usage of technologies in combination has been increasing significantly, especially the trend of IoT+AI, DLT+AI and AI+Edge are growing. Additionally, Figure 6.4 shows the correlation among technologies but in different combinations with each other. This also shows that the research in recent years is using technologies in combination to create certain values by solving relevant research problems in diverse domains. This also supports our research context basis, i.e., *convergence/integration of IoT, AI and DLT technologies to support Dataspace at the edge.*

6.2.3 Summary of Literature Review

Modern technologies such as IoT, AI, Edge and DLT are contributing significantly to transforming businesses toward sustainability. These technologies bring digitalization in I4.0 [23]. Digital technologies in I4.0 play an important role in



(a) Technologies Trend in Recent Years (b) Technologies Usage in Combination

Figure 6.3: Research Trend of Technologies

manufacturing industries to enhance their operational efficiency and profitability through its digitized processes and operations [24]. IoT ecosystem leveraging AI capabilities in I4.0 is adopting new distributed edge architectures to nurture the benefits of computing closer to the data sources or device layer [25]. Edge computing has the advantages of consuming low bandwidth capacity, low latency, instant decision enablement, reduced storage costs, and improved connectivity [26]. From the business perspective, distributed edge computing is transforming relevant business ecosystems and providing benefits in various aspects, such as supporting profitability, scalability, low operational cost, and enriching the user experience by delivering QoS applications [27]. Besides benefits, it has certain challenges, such as a lack of efficient service/resource discovery, managing distributed resources, monitoring, fault tolerance and high availability methods [28]. Additionally, security over distributed edge networks is a significant challenge that needs noble methodologies to protect relevant infra, including gateways, transmission media, devices, and their interconnectivity [29].

6.2.3.1 Edge AI and DLT Integration

DLT provides robust security, data and operational integrity, transparency, reliability, immutability, and digital traceability that can address many of the edge-related security issues [30]. Therefore, recently the trend of edge and DLT convergence (as in Figure 6.3) is emerging [29], [31]. In addition, DLT-enabled edge intelligence [32] and federated machine learning leveraging Blockchain [33] associated areas are also being explored to bring trust and intelligence closer to data sources via edge [34] to satisfy sophisticated requirements (e.g., latency sensitive) of the industrial applications [32]. The advanced areas of ML, which is Federated Learning (FL), have already been seen as a potential feature for bringing industrial IoT analytics to the edge. FL enables distributed learning models that can be deployed at the edge. FL-related privacy and security concerns, such as securely transferring the models to distributed infrastructure, can be achieved through Blockchain [34]. This brings a new paradigm called Federated Edge Learning (FEL), wherein various models are proposed. One such model, named "lightfed," is proposed in [35], that addresses the issues of communication efficiency and data security vulnerabilities for big-size models

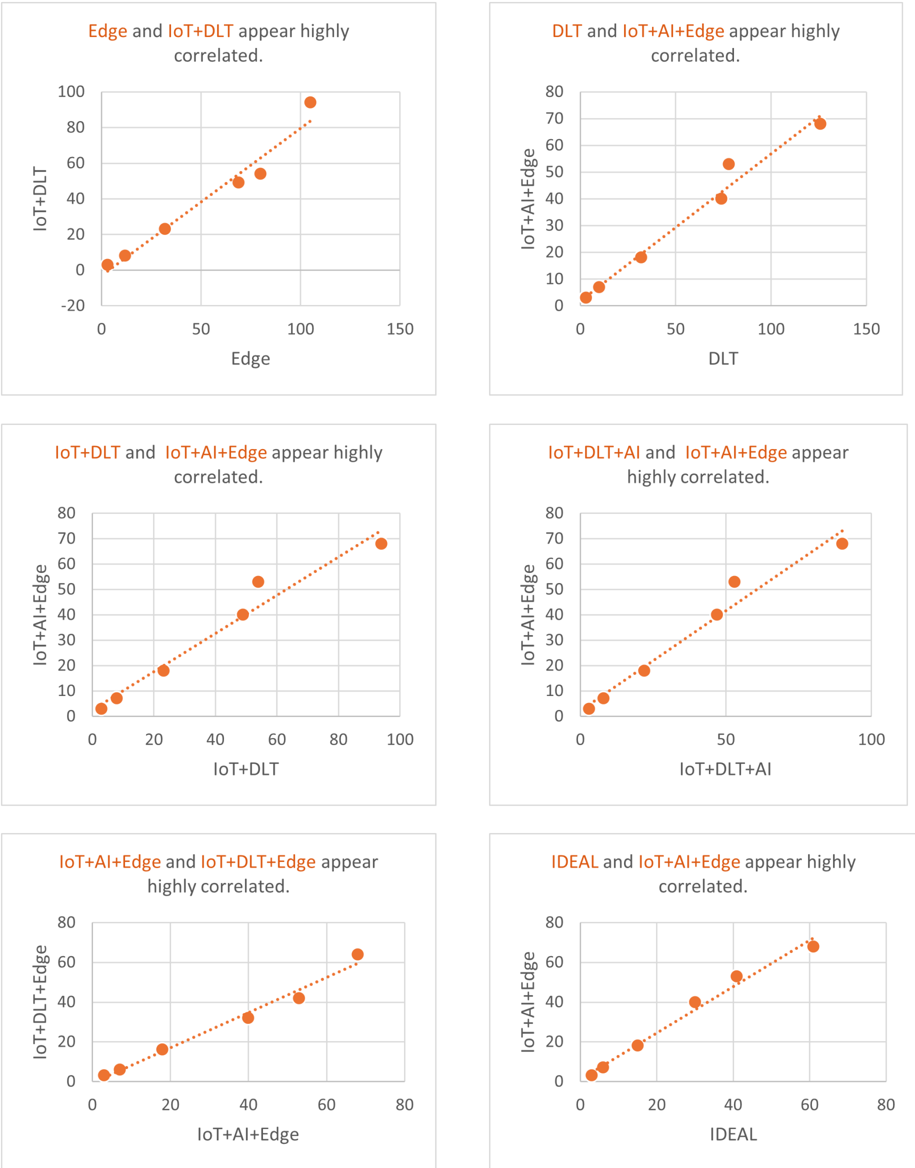


Figure 6.4: Correlation of Different Technologies for Literature Data (2017-2022)

over fluctuating transmission links. However, strategies to aggregate wastage resources at the edge layer are still missing.

6.2.3.2 Exploiting Software-Defined Networking

In [36], an interesting idea is explored by combining public-private Blockchain with Software-Defined Networking (SDN) mechanisms to manage the cluster structure of IoT networks while justifying energy and security issues by focusing on reducing associated Proof of Work (PoW) overhead by inserting blocks in the chain without any PoW consensus. However, managing the SDN controller itself in simple IoT networks appears to be complex and has a significant overhead. Similarly, in [37], the improvement in energy consumption is shown by selecting a single miner approach for PoW consensus for edge-Blockchain distribution. Motivated by SDN, IoT network flows can be segregated (such as control or data plane flows) to enforce dynamic policies to achieve optimal processing at the edge [38]. This has been used as a methodology in our proposed model to process different types of data planes at the edge.

6.2.3.3 Relevant Distributed Edge Architectures

The distributed architecture presented in [39] highlights the exploitation of security and data sharing by tweaking the consensus mechanism of Blockchain and combining it with the edge for AI-enabled IoT applications. However, the study is unclear about the challenges addressed or considered as the basis for the proposed architecture.

Similarly, [40] proposes a fog-cloud-Blockchain architecture based on different agents having different roles and using the concept of Blockchain. The architecture suggests placing a service request on the cloud if it cannot process at the fog node. However, there is no emphasis on the pooling of resources at the edge layer itself instead of opting for the placement of the service request on the cloud in case one edge node cannot accommodate it.

A practical integration of a distributed edge platform (called edgeX) and centralized Blockchain architecture through a Blockchain agent running at the edge is presented in [41]. The architecture focused on providing secure access to IoT resources and having a cloud-level repository for scalable and secure transactions. However, the corresponding result shows it suffers latency issues (not mentioned why, but probably due to centralized components in the architecture) [41]. Therefore, it indicates the practical challenges of Blockchain integration at the edge and, hence, emphasizes the need for a robust distributed architecture design for the edge ecosystem.

6.2.4 Thematic Analysis of Identified Challenges

Based on thematic analysis of the literature review and research context, we have identified challenges related patterns/labels and classified them into relevant groups or themes. This can be found under Table 6.2. The commonly identified challenges have been classified as per the following labels:

- Heterogeneity,
- Interoperability and Integration
- Quality, Trust and Privacy
- Resource/Infrastructure constraints
- Resources (Service, Data, MetaData, Infra) Discovery and Access
- Resources Management and Orchestration
- Analytics and Knowledge Extraction
- Data Lifecycle Management
- Offloading and Optimal Resource Utilization at the Edge
- Distributed Resource Pooling
- Distributed Edge Security

Similarly, the above challenges have been mapped to one or more groups/themes that are given as follows:

- Dataspaces integration context
- Convergence of technologies
- Distributed edge ecosystem
- Edge-DLT integration

The above-identified themes with relevant challenges mapping have been discussed further in subsections.

6.2.4.1 Challenges Specific to the Dataspace Context

The challenges specific to the context of Dataspace are given as follows:

Heterogeneity: It is all about managing heterogeneous conditions, events, or elements in a dynamic environment. This mostly includes integration and harmonization of data models, interfaces, and systems from various domains which is a challenging task since this involves diverse data sources, formats, structures, and processing [16].

Quality and trustworthiness of Data: Applications operating under diverse Dataspace pose a threat to the quality assurance, governance, accuracy, and reliability of data due to involved heterogeneity, inconsistencies, lack of traceability, errors, and integrity options availability [16],[42].

Data Discovery and Access: Large-scale and distributed data sources make the semantic search and access of relevant data within a Dataspace environment challenging [43]. Thus, efficient mechanisms for semantic data search, retrieval, and access are required for traversing multiple cross-domain data sources in the given semantic context.

Analytics and Knowledge Extraction : Data converged from Dataspace needs to be analyzed to generate actionable insights and knowledge. Therefore, this requires advanced analytics and knowledge extraction techniques [44]. However, developing scalable and efficient methods for data analysis and knowledge extraction is a challenge.

Interoperability and Integration: Dataspace includes data convergence from different devices, systems, platforms, and applications. Therefore, achieving interoperability and seamless integration among these disparate entities is challenging [45]. In addition, integrating and aggregating data from multiple sources in Dataspace can be complex due to variations in data formats, semantics, and data interfaces [46]. Therefore, there is a need to develop effective methods for data integration and interoperation.

Data Security: In Dataspace, a pool of heterogeneous data from varied sources can be created. Thus, ensuring security is paramount [47]. It is mandated to have efficient mechanisms to enable data protection against unauthorized access and to ensure data CIA (Confidentiality, Availability, and Integrity) triad [48].

Data Lifecycle Management: Every data that flows from its source to its destination is associated with its life cycle events. For e.g. these events could be associated with data produced, collected, transformed, processed, enriched, stored or deleted. Therefore, the lifecycle management of data flowing in Dataspace integration is an important aspect and poses challenges [49] at different phases, such as during acquisition, retention, versioning, and governance etc.

6.2.4.2 Challenges related to the Convergence of Technologies

The challenges related to the convergence or integration of technologies for IoT, AI, and DLT at the edge have been summarized as follows:

Security, Trust and Privacy : Establishing trust and ensuring the security of edge computing systems that leverage the convergence of technologies like IoT, AI, and Blockchain is crucial [50]. Protection of trustworthy data sources, reliable communication channels, and robust consensus mechanisms are essential in edge environments [51]. Building a secure ecosystem around the edge will promote trusted environments by verifying the CIA triad of data and ensuring the reliable execution of relevant services derived from the convergence of technology.

Resource Constraints: Edge devices often have limited resources in terms of storage, computing power, and network connectivity. Optimizing resource utilization and managing resource constraints while maintaining performance and scalability is a challenge and should be taken care of simultaneously [17]. Techniques such as edge caching, lightweight algorithms, and efficient resource pooling and allocation strategies [52] can help address these constraints.

Interoperability and Integration: Integration of different technologies to develop valuable services among diverse IoT devices, systems, and data platforms is a significant challenge [53]. This is due to the factor of heterogeneity involved in technologies, systems, operations, interfaces, and data models [45]. Therefore, there is a need to develop standardized protocols, compliant interfaces and frameworks that enable seamless integration and interoperable operations. Interoperability standards-compliant data models and interfaces (such as open APIs) leveraging semantics contexts can handle heterogeneity, interoperability, and integration in a simplified manner [54]. Therefore, there is a need to design architectures that seamlessly integrate these technologies in relevant semantic contexts and ensure their compatibility and synergies.

Service Discovery: The convergence of technologies can be used to develop a wide range of distributed service offerings at edge leveraging Dataspace integration. For example, analytics-as-a-service, infrastructure-as-a-service or Blockchain-as-a-service at edge [33]. However, the discovery of services in a distributed edge ecosystem will always remain challenging due to integration, interoperability, and heterogeneity environments [6]. Therefore, effective mechanisms for service registry, discovery, and publishing are required at the edge.

Scalability: - The convergence of technologies is meant to use multiple technologies to nurture relevant benefits [29], which may often lead to the consumption of resources at the edge that has already a scarcity of resources. However, this can be handled by resource pooling and partitioning of services at the edge layer [74]. Therefore, scalable distributed architecture designs are needed to address relevant scalability concerns.

Management and Orchestration: Distributed management of resource scheduling and orchestration of services developed from convergence of technology poses other significant challenges [67] in terms of dynamic instantiation, configuration, monitoring, fault tolerance, and relevant software updates. This can be achieved with dynamic deployment, monitoring, and provisioning methods that automate tasks [68], provide centralized control, and ensure system resilience.

6.2.4.3 Challenges related to the Distributed Edge Ecosystem

Interoperability & Heterogeneity: The edge ecosystems involve heterogeneous workloads and edge entities such as gateways and devices that belong to different vendors and are compliant with diverse standards [55], [56]. These distributed networked diversified entities are becoming active components of a broader yet synergetic and linked web of systems-of-systems [6]. However, a major challenge for this transformation is to support interoperability with legacy systems and multi-stakeholder (vendors/standards) environments due to proprietary or different data formats, models and interfaces [75]. Therefore, it is important to define interoperability mechanisms at the edge that allow legacy systems to become part of the next-generation edge ecosystem, enabling a smooth transition towards the edge ecosystem. This can be empowered further by semi-automated semantic adaptation [46], and application partitioning to

Table 6.2: Thematic Analysis of Related Challenges

Challenge	**Applicability of Challenges in Related Themes**	**Reference**
Heterogeneity	Edge-DLT integration; Dataspace integration context; Distributed edge ecosystem; Convergence of technologies	[55], [56], [6], [16], [46], [53], [57], [58]
Interoperability and Integration	Edge-DLT integration; Dataspace integration context; Distributed edge ecosystem; Convergence of technologies	[16], [46], [45], [53], [45], [54]
Quality, Trust and Privacy	Edge-DLT integration; Dataspace integration context; Distributed edge ecosystem; Convergence of technologies	[16],[42], [47], [48], [50], [51]
Resource or Infrastructure constraints	Edge-DLT integration; Dataspace integration context; Distributed edge ecosystem; Convergence of technologies	[17], [59], [58], [60], [61], [62]
Resources (Service, Data, MetaData, Infra) Discovery and Access	Edge-DLT integration; Dataspace integration context; Distributed edge ecosystem; Convergence of technologies	[43], [63], [64], [65], [66], [6], [33]
Resources Management and Orchestration	Edge-DLT integration; Distributed edge ecosystem; Convergence of technologies	[67], [68], [34], [58]
Analytics and Knowledge Extraction	Edge-DLT integration; Dataspace integration context	[44], [29], [31], [32], [33], [35]
Data Lifecycle Management	Dataspace integration context	[49]
Offloading and optimal resource utilization at edge	Edge-DLT integration; Distributed edge ecosystem	[4], [3], [69]
Distributed resource pooling	Edge-DLT integration; Distributed edge ecosystem; Convergence of technologies	[70], [62], [6]
Distributed edge security	Edge-DLT integration; Distributed edge ecosystem	[47], [48], [50], [71], [72], [73]
Scalability	Edge-DLT integration; Distributed edge ecosystem; Convergence of technologies	[29], [74], [71], [34]

support large-scale cross-platform processes across the distributed computing continuum.

Distributed Resource Pooling: Emerging smart edge-oriented IoT services can be effectively implemented by virtually abstracting and discovering edge nodes to determine and build a pool of available resources at the edge layer. This can also be optimized further based on inferenced situation awareness [70] that crawled for the availability of local and global edge resources (data, devices or network). The semantic discovery and sharing of resources with each other across edge networks can be used to build resource pools that can fulfil the required QoS demands of an application [6]. This requires discovery enabling and integration of information (e.g., resources and application demands) across functional hierarchy with situation awareness at distributed decision-making levels [62]. This will help to reduce the requirements for highly expensive centralized computing, governance, and silo effects.

Offloading and Optimal Resource Utilization at Edge: The huge surge in IoT-related services is preferred to be hosted in the cloud which increases resource demand at cloud [4]. This impacts operational and capital expenditures, scalability, latency, security, and other unknown aspects. In addition, modern applications requiring streaming of high volume data in an ultra-low latency [69] (close to real-time) manner need to effectively utilize their edge network resources [3]. Therefore, edge infrastructure (wherever possible) needs to be optimally utilized to offload centralized computing, save bandwidth and avoid latency.

Management and Orchestration: Traffic generated over the network is characterized by the 4Vs of data which include volume, variety, veracity, and velocity. These 4Vs are expected to significantly affect the QoS of next-generation edge networks and their resources [34]. The question it raises is how to effectively manage and orchestrate the resources at the edge network for different types of flows that derive from different streams (fused) of data, especially when they belong to heterogeneous services and devices and have different data models [58]. This requires semantic context-driven dynamic provisioning and collaborative processing (e.g., pooling of resources from a nearby edge) at the edge. In the collaborative edge context, services, data and infrastructure resources from diverse stakeholders (under Dataspace integration) can be pooled and managed to develop a value chain network such as cross-domain infrastructure or resource reusability-driven servitization models [76].

Dsitributed Edge Security: The rapid advancement of I4.0 brings forth the crucial need to safeguard cyber security, privacy, and trust across users, applications, data, infrastructure, and related processing [73]. The inherent challenges faced in the edge computing landscape encompass information sharing, user privacy, granular access control, vulnerability management, and defence against various cyber threats and attacks. These challenges necessitate the implementation of distributed security, privacy, and trust mechanisms to fortify the edge infrastructure and associated resources. Consequently, decentralized governance and resource protection becomes imperative. In this context, using DLT and traditional security methods can offer promising solutions [6]. By leveraging DLT, decentralized security measures can be established, ensuring transparency in the flow of information, traceability, permissioned control over resources, and trustworthiness within the networks. This empowers the efficient production and consumption of services among distributed edge resources.

6.2.4.4 Challenges of Edge-Blockchain Integration

The integration of Blockchain with edge computing gives rise to different challenges. This includes security vulnerabilities due to tweaking of Blockchain consensus mechanisms, resource constraints at edge [59] for executing Blockchain-related heavy tasks and data scalability issues of Blockchain. In addition, relevant devices, services, or data discovery are also significant challenges in IoT [63], [64]. To address the edge infrastructure constraints, MEC-enabled IoT has been recognized as a promising paradigm under 5G [58], [60] and 6G to improve network performance [61]. Therefore, the integration of MEC with Blockchain has been observed as a significant architectural component to support autonomy services for robotics automation in the manufacturing domain under 5G and beyond networks [62]. The emerging DLT and smart contracts, along with security features for discovery, privacy, and integrity, can also enable an incentivized marketplace for edge computing [65], [66]. Similarly, AI-enabled federated learning at the edge offers huge analytical capabilities closer to the data sources, such as low latency decision-making.

Therefore, IDEAL convergence and integration can complement each other and could bring significant transformations across several industries to offer novel distributed architectures and applications [57], along with Dataspace integration possibilities. In addition, it will also unlock new IoT business models wherein actors can engage directly at the edge, e.g., via an edge-based servitization model [76], to develop a transformed edge-oriented business value chain networks [71], [72].

6.2.5 Gaps and Necessity

Although there have been various technological combinations utilized in different industries, there is a notable lack of practical distributed architectures that specifically address the challenges at the edge in the Dataspace integration context. Furthermore, research on models for the integration and convergence of IoT, AI, and DLT (Blockchain) at the edge is limited, despite its potential to disrupt or complement traditional centralized cloud computing-driven business ecosystems. This lack of research hampers the development of Dataspace-oriented edge-driven business ecosystems where consumers can directly engage at the edge to access diverse services, resources and data of their interest. There are numerous scenarios where direct engagement at the edge can be a viable option under Dataspace integration scenarios, such as managing drone operations [77], autonomous car operations [78], or monitoring assembly/supply-chain quality of wind turbines in the energy domain [79]. Finally, addressing the aforementioned challenges requires a holistic approach that considers the unique characteristics and convergence of IDEAL and semantic technologies that can be realized effectively and unlock their full potential in developing innovative solutions.

Therefore, to bridge these gaps, this article proposes the DENOS model, which leverages the convergence and integration of IoT, AI, and DLT capabilities at the edge through semantics-driven adaptable contexts for Dataspace integration purposes. The next section covers the model in depth.

6.3 Proposed Model

In this article, we propose a distributed architectural framework that aims to enhance the capabilities and efficiency of edge computing systems through semantic modelling of resources and convergence of technologies. This has been named and referred to as DENOS, i.e., Distributed Edge Network Operations oriented Semantic model. The model has been shown in Figure 6.5. DENOS has a hierarchical and distributed layered architecture comprising Cloud, Edge, Devices, Semantics, Convergence, and Dataspace integration layers. In addition, each of the layers also offers relevant interfaces for operational integration and interoperation. This model extends the traditional concept of edge computing by introducing three new layers that are *Semantics, Convergence and Dataspace integration* layers. It leverages semantic modelling context, convergence and integration of IDEAL technologies, and distributed computing principles to address many of the aforementioned challenges, such as heterogeneity, interoperability, integration, resource orchestration management, security, and scalability at the design level. This integrated approach enables efficient and effective utilization of resources, seamless data sharing, and the development of cross-domain Dataspace solutions in a distributed edge network.

In DENOS, there are two types of operation planes identified, namely the Control Plane (CP) and Data Plane (DP) operations. CP is associated with the controlling or managing operations of resources, services, systems, devices and related processing flows over CED continuum infrastructure along with the non-functional and deployment aspects such as discovery, monitoring, scalability, provisioning, coordinated operation in a distributed environment, etc. DP operations are associated with data-specific operations such as data processing functions (e.g. transformation, storage, adaptation, validation, analysing, sharing, etc.) while the data flows upstream (coming from a device or data source) or downstream (coming from application or cloud or edge towards device or data source) direction following CED or Edge-Edge continuum.

6.3.1 Principal Features

At core, the DENOS model is designed to have five main features, i.e., Discovery, Execution, Negotiation (Dataspace), Optimization, and Security. Each of these has been explained below:

Discovery: DENOS facilitates the semantic search of distributed edge resources, including devices, gateways, services, and data sources. Based on the discovery, this feature expects to build a knowledge graph of available resources in terms of their proximity, service, and computing availability across the CED continuum. To achieve this, various semantic modelling techniques and standards can be utilized, such as RDF (Resource Description Framework) [80], NGSI-LD (Next Generation Service Interface - Linked Data) [22], OWL (Web Ontology Language) [81], and SKOS (Simple Knowledge Organization System) [82]. These techniques are used to model and implement such semantic search context in DENOS to develop knowledge graphs of resources, services and data availability along the CED continuum [83]. The Knowledge graph of available resources with a global and local view of utilization resources will enable the optimum utilization ((re)scheduling, deployment, scalability, etc.) of infrastructural resources for scheduling tasks and related processing. This feature enables the processing context implementation using identification, registration, and retrieval of available resources to be used in deploying diverse service contexts to process different data flows. Thus making it easier to locate, utilize and extract knowledge about available resources within the distributed edge network. Discovery of resources is associated with the CP operations.

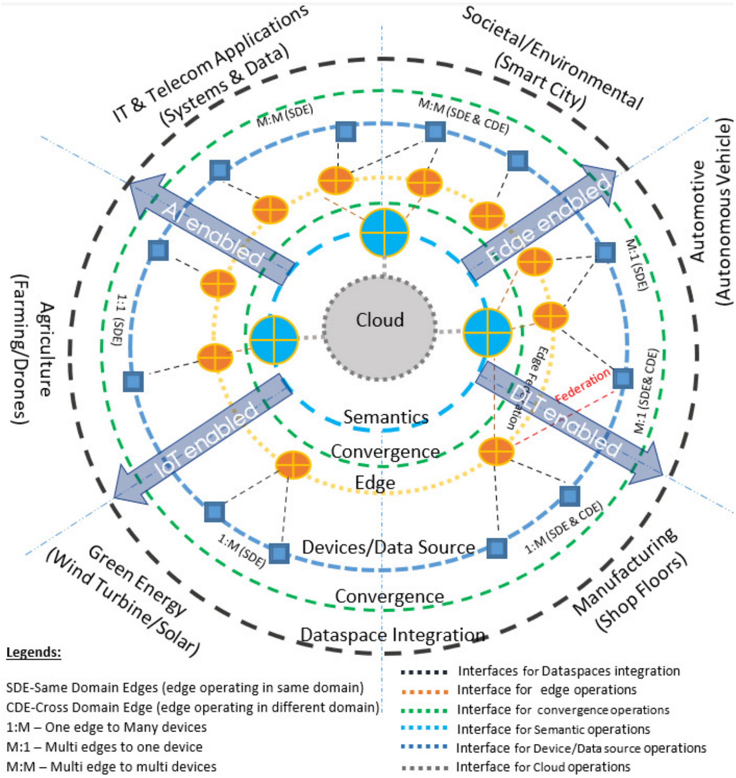


Figure 6.5: DENOS Model

Execution: Once processing contexts are discovered and defined, then the execution for these contexts is started. This includes the provisioning, configuration, allocation, and scheduling of tasks based on discovered resource availability, computational capabilities, QoS requirements, etc. This feature ensures efficient and optimal execution of processing context by utilizing relevant edge resources. This feature is associated with the convergence layer (to support DP and CP operations) that aims to converge resources and technologies to execute a certain semantic context over the discovered infrastructure and resources.

Negotiation (Dataspace): Within the business modelling and Dataspace integration context, DENOS facilitates the exchange of information to enable data-driven decision-making and collaboration among cross-domain edge resources for Discovery and Execution purposes. It empowers different domains to negotiate resource usage, data sharing, service reusability, and other operational aspects in a decentralized and secure manner. This negotiation capability not only fosters collaboration but also creates monetization opportunities for various stakeholders collaborating within the distributed edge network, aligning with the servitization model [76]. This is precisely what can be achieved through the Dataspace concept. For example, a stakeholder can offer its edge resources in Infrastructure-as-a-Service(IaaS), Platform-as-a-Service (PaaS), Software-as-a-Service (Saas) or Functions-as-a-Service (FaaS) business models to host services and associated data operations at the edge from another stakeholder based on agreed norms and conditions. The norms and conditions among different stakeholders can be digitally realized through DLT-driven smart contracts for such scenarios. Thus, by enabling decentralized negotiations, DENOS promotes efficient resource reusability, sharing, and integration and supports the realization of cross-domain collaborative and edge-enabled Dataspace value chain networks.

Optimization: DENOS aims for optimization through distributed resource pooling at the edge layer to enhance resource utilization and overall system performance along the CED continuum. It considers factors such as resource reusability, monitoring of computational load, network bandwidth, energy consumption, and application-specific QoS requirements to optimize resource allocation and task scheduling at the edge. This results in improved edge computing system scalability, efficiency, and responsiveness.

Security: DENOS leverages convergence and integration of IDEAL technologies to offer seamless, secure operations. DLT out of IDEAL is converged with a perspective to provide distributed security for seamless edge operations that can be executed directly at the edge. The security in the DENOS context is meant to cover the distributed trust, traceability, identity access and management, integrity and confidentiality aspects for both the CP and DP operations. For example, CP operations, which include infrastructure-specific discovery, provisioning (configuration or service instantiation), resource availability information, etc., can be shared through the DLT-enabled private network that consists of peers/entities (edge or cloud nodes) along the CED continuum. Similarly, DP operations can also be chosen to process through the DLT networks to avail DLT benefits if related latency compared to the traditional security methods is accepted. DLT makes the CP/DP operations immutable, tamper-proof, transparent and traceable to the CED continuum stakeholders.

During implementation, DLT-driven smart contracts can be used to build a consensus of allowed CP or DP operations and combine them with the traditional IoT-based identity and authentication management using OAUTH2, RBAC or ABAC methods (Role/Attribute Based Access Control) among services, systems and stakeholders [6].

6.3.2 Distributed Layered Architecture

As shown in Figure 6.5, DENOS has a distributed layered architecture comprising six layers. Let's explore how each layer of the DENOS model is defined and linked in addressing the identified challenges:

Cloud Layer: In the DENOS model, the cloud layer still plays a diverse yet key role in the architecture. Besides other functions, it is responsible for aggregated analytics and centralized control functions across the CED continuum [84]. The cloud layer presents a centralized cloud infrastructure that supports the edge layer by providing additional computational resources, storage, and advanced analytics capabilities. It addresses scalability challenges posed by the edge layer by accommodating resource-intensive tasks that may exceed the capabilities of edge devices. The cloud layer also addresses data volume challenges by providing long-term storage and enabling large-scale data processing and analysis.

Edge Layer: The edge layer comprises edge devices, gateways, and local processing capabilities. It serves the data source as the first and the closest level of processing. Entities at the edge layer can have different semantic relationships (e.g., based on cardinality one-to-one or one-to-many) with other entities placed in the same or in a different layer, such as edge-edge or edge-device layer entities. The edge layer addresses multiple challenges, including resource pooling, orchestration and management, heterogeneity, interoperability, scalability, and latency. It leverages the semantic layer to perform semantic resource discovery to build resource pools among distributed edge infrastructure and semantic context-driven provisioning and processing to deploy workloads. Under the DENOS model, one edge collaborates with other nearby edges (horizontal scalability) for the pooling and discovery of resources at the edge layer. And if the resources are not sufficient to meet the QoS of the application then it offloads the processing to the cloud (i.e. vertical scalability). Therefore, it ensures the availability of resources to process a given task at the edge layer with its best-effort services.

Devices or Data Sources Layer: This layer represents the physical devices, sensors or available data sources (such as data platforms in Dataspace). The data sources can belong to various domains, such as smart homes, industrial settings, healthcare facilities, and transportation systems. The devices generate data and interact with the edge layer to process the data to perform relevant actions. The device layer, as a data generation or sourcing layer, addresses the challenges of data generation and collection from various sources. It is also responsible for the transmission of data to the edge layer for further upstream processing [85].

Semantic Layer (SL): The semantic layer forms the core of the DENOS model. This layer aims to provide semantic capabilities for different purposes, such as defining semantic context for resource discovery, service or resource execution, data processing flows, convergence of technologies, and operations optimization. Therefore, this layer defines three types of contexts given as follows:

- **Data Context (DC)**, defines metadata or schema of data for context-aware DP operations and can be achieved via semantic adaptation of data using certain data context, e.g., NGSI-LD core context.

- **Processing Context (PC)** defines the knowledge about life cycle management (LCM) events associated with CP operations and computing infrastructural requirements related to a certain process for a given service. The LCM consists of events such as discover/monitor, schedule, re-schedule, start, stop, update, scale, upgrade, etc. These events allow the operational management of workload covering the non-functional aspects over the CED continuum.

- **Service Context (SC)** defines the service(s) deployment artefact and mechanisms to access the service, such as port, IP, API/URL path, etc. It is associated with CP operations specific to service deployments and related LCM events. This establishes the processing path for DP operations processing by instantiating required services to process the data flow.

This layer enables the intelligent management and coordination of resources across the distributed edge network. The semantic layer incorporates semantic standards, techniques, and technologies to realize different semantic contexts, such as ontology-based web/data modelling OWL, RDF, NGSI-LD, etc. [83]. This layer uses semantic reasoning on top of semantic modelling to enhance the understanding and utilization of resources in a certain context which can address many challenges related to heterogeneity, integration, and interoperation. It provides common grounds to understand diverse data models, operations, services, deployment and resources context through semantic modelling and ontology-based approaches [83].

Convergence Layer (CL): This layer provides the actual execution or instantiation for different contexts, such as data context, processing context and service context. The data context provides the metadata based on which the data flow will be validated for its schema and matched against the processing labels in order to get processed further. The processing context provides the necessary infrastructure to execute the service context. The service context provides the service packages and related deployment information along with the information for accessing the service for serving the data flows. Thus, the convergence of technologies at different levels converges to develop certain data functions/services to realize specific service contexts.

The convergence layer combines the use of IDEAL, i.e., IoT, DLT, Edge, and AI/ML technologies, to work in synergy and to support the functionality of other layers of DENOS. Each technology brings its own unique strengths and capabilities. Along the CED continuum, IoT acts as an ecosystem for real-time data generation via environment sensing, collections, processing, and executing relevant actions. DLT can provide decentralized and secure channels for data/meta-data sharing and trust establishment among stakeholders for building the Dataspace value chain. The edge and cloud layers provide analytical processing through the deployment of AI/ML algorithms (e.g. Reinforcement Learning, Decision Trees, Semi/Un-Supervised anomaly detection, etc.) over the CP/DP data generated by the DENOS ecosystem. Thus enabling intelligent decision-making to support efficient CP/DP operations based on knowledge extraction from data. In DENOS, these technologies can be converged in different processing or combinational semantic contexts to develop dynamic data, processing and service contexts. For example, some data flows can be defined to be processed only in the cloud for their analysis, and others can be forced to use AI capabilities at the edge. Similarly, some data flows can be selected to process with Blockchain to track their digital traceability, and other flows do not need Blockchain service at the edge to avoid unnecessary delays. Therefore, DENOS can leverage the synergistic capabilities of the convergence layer in different semantical execution contexts to create powerful solutions that address complex challenges in edge networks.

Dataspace Integration Layer: The Dataspace integration layer plays a key role in the business modelling context of DENOS directly at the edge. This layer enables seamless integration and collaboration across multiple domains with the help of other layers. Leveraging other layers functions, it facilitates the efficient and secure sharing, reusability, governance, and provenance of data and services in diverse systems, processes and business domains. This layer leverages the capabilities of the convergence and semantic layers to address interoperability and integration challenges. By harmonizing data from different sources, it enables efficient discovery and utilization of resources across domains. Therefore, this layer is capable of supporting data integration for various use cases, such as smart cities, agriculture, healthcare monitoring, and supply chain management [60]. The different use cases are associated with data from multiple domains, which can be integrated and analyzed to develop business values or other innovative solutions. Thus, this layer enables the development of an innovative Dataspace value chain that spans multiple domains. By leveraging the power of a distributed edge ecosystem, the Dataspace integration layer enhances the efficiency, scalability, and effectiveness of data sharing and utilization. This will improve integrated decision-making processes and promote value-creation chains in diverse domains and industries.

Overall, the DENOS model has hierarchal and interconnected layers that work collaboratively. Three new layers (i.e., semantic, convergence, and dataspace

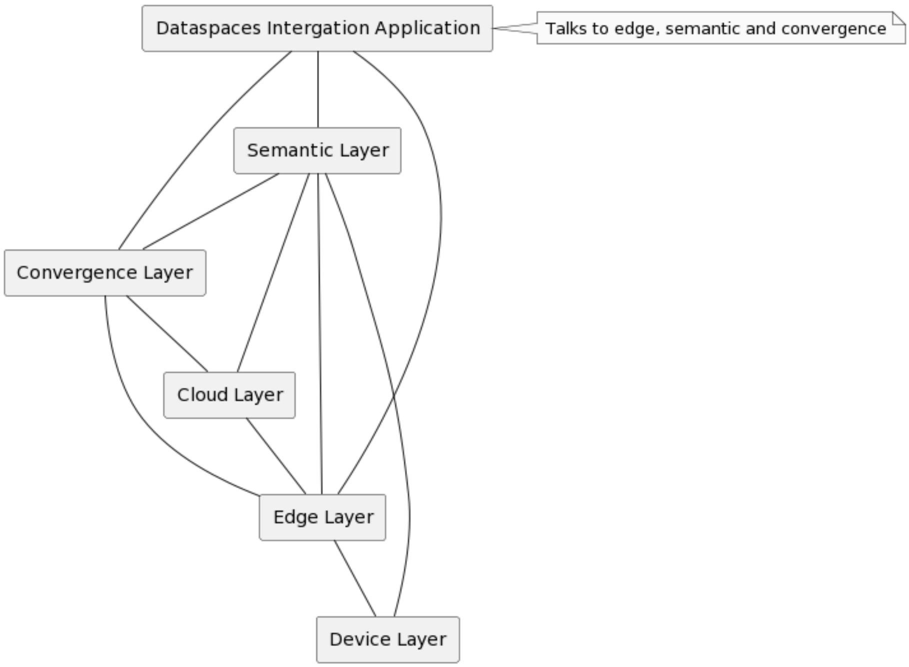


Figure 6.6: Denos Layers Interconnection

integration) are introduced to address the identified challenges. The semantic layer provides semantic capabilities for defining diverse data, processing and service semantic contexts, while the convergence layer offers execution of these contexts leveraging IDEAL technologies in different combinations. The cross-domain Dataspace integration layer leverages the capabilities of the underlying layers to develop innovative value-chain that span multiple domains. To enhance the understanding of the DENOS framework, Figure 6.6 illustrates the interplay and processing among the different layers. The Dataspace integration layer establishes communication with the edge through the semantics and convergence layer. Serving as the central functionality, the semantics layer connects to all other layers, providing diverse semantic contexts tailored to specific needs such as data processing, modelling, enrichment, and technology utilization. The CED continuum follows the traditional connection pattern but is intercepted by the semantic and convergence layers during processing, enabling seamless integration of these layers.

6.3.3 Semantics Driven Processing Model

The semantic context-driven processing model (or knowledge graph) for DENOS has been shown in Figure 6.7. It consists of Processing, Data and Service Context, which have already been explained. It has been developed using RDF modelling techniques based on turtle format coding [86]. RDF provides a general method for conceptual description or modelling of information used in web resources. Turtle provides the serialization formats to represent RDF data in a human-readable and compact manner. Here is a brief summary of the relationships between the entities presented in the RDF-driven processing model:

- d represents the DENOS contextual namespace.
- Semantic Context provides Processing, Data, and Service Contexts.
- Semantic Adaption processes the Data Context.
- Data Flow has Data, Selector, and Type properties and is processed by the Processing Context.
- Selector has a Matching Pattern and Type.
- Semantic Discovery has Type property and processes the Resources Discovery functions.
- Resources Discovery has Type property and provides "CPU/RAM/HDD-Resources" availability.
- DataContext has "@Context", Type, and Selector property and is processed by the Semantic Adaptation.
- Processing Context has Selector and Type property and processes Data Flow and Semantic Context.
- Service Context provides Service, Semantic Adaptation, and Semantic Discovery service functions and also has Selector and Type properties.
- Service has a Selector, Location, "Service-Package", Resources requirements, and Type properties.

The typical processing flow includes the incoming data from the devices at the edge, which will be treated as Data flow (DF) with certain matching patterns. Edge will check the associated Processing Context provided by the SL and preconfigured at the edge for processing the DF based on the matched pattern. However, this processing context can have distributed processing requirements across CED. Therefore, in such cases, the PC having the service context for the DF will guide the flow of related scheduling over the DENOS infrastructure (cloud, edge, or hybrid), where services associated with the service context need to be instantiated. This could be perceived as the distributed processing pipeline representing the service context (consisting of single or multiple services, each with a specific technological function) where a specific service in the pipeline will process the DF. The service represents implementing a specific function related to specific IDEAL technology. The Convergence Layer will provide the IaaS, PaaS, and SaaS to converge the processing for the data flow at the edge, cloud, or both. The Convergence layer implements the processing context provided by the SL for certain data flow matching requirements.

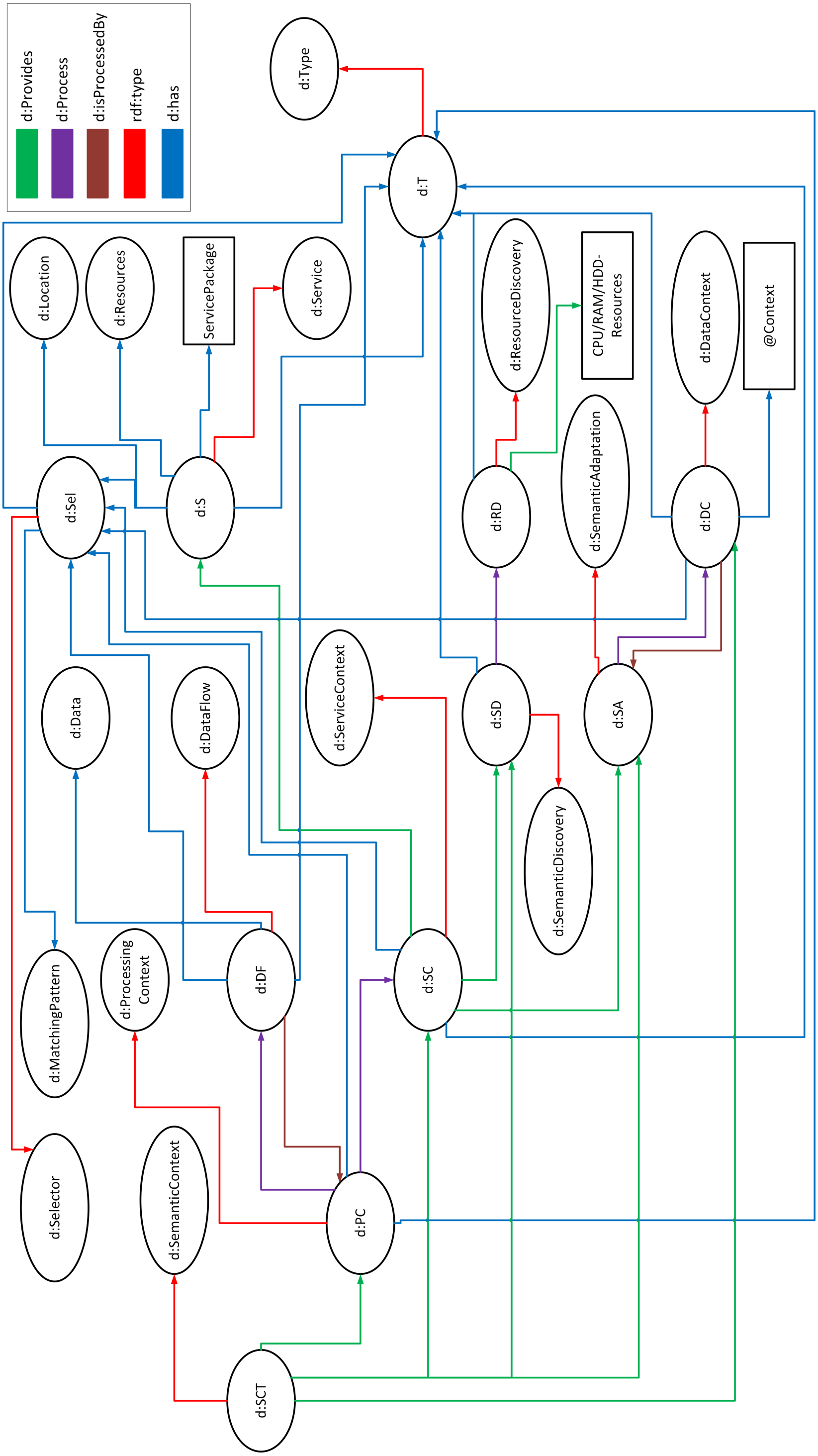


Figure 6.7: DENOS Semantic Processing Model

In summary:
– Processing context (Infrastructure requirements) tells how (i.e. service context) and where (location) to process the Data flow, i.e. offload, scale, (re)schedule and secure based on resources discovery and availability, i.e. to provide information about location and service that will process it.
– CL does the actual execution of the processing context for the data flow based on the Service Context.
– Service Context defines the artefacts, which includes the service package(e.g. docker), endpoint, and computing requirements. Service context consists of service(s) at the core, which uses different technologies such as IoT, AI or DLT to offer service functions in different processing contexts. For e.g. semantic adaption service can be defined under service context to augment NGSI-LD context (to make data context-aware at the Dataspace application layer) in incoming data flows at the edge.

6.3.4 System Implementation Context

DENOS provides an architectural framework. Therefore, the actual realization of the model depends on the target use case and related requirements. However, the below functional blocks guide the implementation of DENOS.
Semantic Context: The Semantic context consists of the definitions and implementation of the Processing, Service and Data contexts shown in Listing 6.1 to 6.3. These are NGSI-LD standard-based defined exemplary contexts just to show the realization of such context in the real environment. Therefore, it should not be treated as final specifications but as fundamental guidelines to build specifications in this direction. For example, the Processing Context refers to the Service and Data context shown in respective Listing 6.2 and Listing 6.3 to define the overall processing flow. The service context shown uses the Kubernetes-based deployment artefacts but can be changed to other deployment tools like Docker-compose. It is also important to note that Service Flows (SF) can be developed using the processing context to process the DF with a chain of services running over distributed edge infrastructure. Processing, Service, and Data contexts can be perceived as the IaaS, PaaS and DaaS (DataContext-as-a-Service) models at the edge layer. These contexts can be realised using IDEAL and other semantic technologies such as RDF and OWL (Web Ontology Language) to develop a Knowledge graph having local or global contextual information scope and to establish a common understanding of processing infrastructure/resources, services, data and metadata across CP/DP operations [87], [46]. The Semantic context enables seamless integration and interoperability over data and services from diverse sources, thereby addressing the issue of heterogeneity to a large extent. For example, as demonstrated in Listing 6.1 to 6.3, context awareness can be implemented at the infrastructure, data, service and application level using NGSI-LD [22] standard. NGSI-LD standard provides information about entities, their attributes, and relationships. In addition, it enables applications to become context-aware and make informed decisions based on real-time information. Apart from that, MQTT-like standards can also be used to achieve a

publish-subscription infrastructure type model to offer synchronization and QoS aware processing at DP level for different DF entities [88].

Data Integration: This functional block is about developing a semantics and context-aware data integration ecosystem at the edge using semantic data integration techniques that facilitate the aggregation and harmonization of data from multiple sources, which can be federated in a distributed environment [59], [22]. The context-aware data integration techniques provide cross-domain data and service integration opportunities to develop a Dataspace value-chain-driven ecosystem. This enables the sharing and reusability of data, services and systems among multi-stakeholders to create win-win business models using a unified view of data for decision-making and analysis [14].

Microservices Oriented: This is related to the realization of the Service context. Microservices' architectural design enables services to be implemented in an independent, self-contained, modular, and scalable manner. Therefore, relevant technological service contexts, such as analytical or DLT-enabled, should be orchestrated and packaged using modern microservices deployment tools and templates, e.g. Docker and Kubernetes. These tools and template packages deploy different services into independent and scalable units [89]. Deploying DENOS as microservices allows for lightweight, flexible, easy management, and efficient scaling of individual processing components.

Edge-to-Cloud Collaboration: This functional block ensures the infrastructure availability for the Processing context across the CED continuum using coordination between edge devices and cloud resources. This is required to offload computational tasks if the edge layer does not have sufficient computing capabilities [73]. Apart from that, backward compatibility for centralized cloud-enabled data processing and storage can also be utilized. The objective is to leverage the strengths of both edge and cloud environments to offer balancing workloads for seamless operations, optimize utilization of edge resources, and save on cost, latency, and bandwidth metrics.

Blockchain Integration: This functional block guides the implementation of robust data security and privacy mechanisms both in CP and DP operations, such as encryption, access control, governance and provenance techniques, to protect sensitive information within the Dataspace value chain network. This ensures compliance with privacy regulations and builds trust among users and stakeholders. For this, the integration of Blockchain/DLT technology as a distributed security is needed to support Dataspace stakeholders' operations at the edge. This will ensure trust, immutability, and tamper resistance in cross-domain interactions and resource sharing across value chain networks [90].

Machine Learning and AI Algorithms: Developing and deploying machine learning and AI algorithms such as Reinforcement Q-Learning model (to optimize control decisions), Deep learning driven autoencoders (for anomaly detection) or Swarm Intelligence based Artificial Bee Colony (ABC) algorithm (for energy-efficient optimization) within DENOS are needed to enable intelligent decision-making, predictive analytics and automated resource management. These algorithms can optimize resource management, dynamic task offloading, detect anomalies, and improve system performance. Therefore, there is a need

to explore the use of Federated Edge Learning (FEL) [77] techniques to train AI models collaboratively across distributed edge networks to support such AI-enabled operations. In traditional ML approaches, the data is collected and trained centrally, whereas the FEL enables the model to be trained locally at the edge, and only the model updates need to be sent back to the centralized locations. This approach ensures privacy and reduces the need for data transfer while still benefiting from collective intelligence and improved model accuracy.
Edge Computing Frameworks: The implementation of DENOS should leverage open-source edge computing frameworks instead of reinventing the wheel. Therefore, it is important to leverage already available open-source edge computing frameworks and tools such as Kubernetes [89], OpenFog [91], TensorFlow Lite [92], IDS Connector [47] for the edge to enable efficient data processing, real-time analytics, Dataspace integration operations and low-latency interactions within DENOS. These frameworks provide the necessary infrastructure and tools for deploying AI, IoT and Dataspace applications at the edge. These open-source tools offer open APIs and source code that can be extended further as per the target environment. These off-the-shelf solutions readily offer an extensive range of data processing, real-time analytical services, as well as essential infrastructure provisioning tools, facilitating the deployment of AI, DLT, and IoT enabled applications at the edge.
Dataspace Integration Application: Dataspace at the edge, where each edge acts as an organizational entity that wants to share and re-use its data and computational power source, can be developed to create cross-domain value chain applications [14]. For cross-domain value chain development, there is a need to capture the relevant concepts, relationships, and rules for the specific application domain using domain-specific ontologies and semantic models [81]. Afterwards, semantic reasoning and inference can be applied over cross-domain informational models to derive or extract knowledge insights to determine correlations and make intelligent decisions based on the integrated data. In addition, the role of IDEAL technologies is crucial, as each technology offers unique features with respect to data processing and related management. For example, AI can provide data analysis capabilities, and DLT can solve the issues of building digital trust, traceability, transparency, and governance related to shared data and associated services among stakeholders in the Dataspace ecosystem. Here, cross-domain operational interoperability can be achieved by utilizing standard complaint APIs such as NGSI-LD, MQTT [88], and defining common data models to exchange information between diverse data sources, services, systems and related applications.

Listing 6.1: Processing Context

```
{
"@context":  "https://uri.etsi.org/ngsi-ld/v1/ngsi-ld-core-context.
    jsonld",
"id": "urn:ngsi-ld:ProcessingContext:001",
"type": "ProcessingContext",
"label": {
```

```
    "type": "Property",
    "value": "urn:ngsi-ld:ProcessingLabel:001"
  },
  "infrastructure": {
    "type": "Property",
    "location": {
      "type": "Property",
      "value": ["urn:ngsi-ld:Edge:001"]
    },
    "matchPattern": {
      "type": "Property",
      "value": ["DF matching pattern"]
    },
    "orchestrator":{
     "type": "Property",
     "value": {
         "id": "urn:ngsi-ld:orchestrator:001",
        "type": "Kubernetes|Docker-Compose",
        "interface": "URL:to:orchestrator"
        }
    },
    "resources": {
      "type": "Property",
      "value" :{
            "allocation": "static|orchestrator",
            "strategy": "LoadBased",
            "Scaling":{
            "Threshold": 0.7
            }
      }
    }
  },
  "operation_type": {
    "type": "Property",
    "value": "CP",
    "description": "Type of operation (CP or DP) associated to LCM events"
  },
  "lcm_events": ["discover", "monitor", "schedule", "re-schedule", "start", "stop", "update", "scale", "upgrade"],
  "distributed_capabilities": {
    "type": "Property",
    "value": {
      "availability": "High",
      "resource_types": ["Cloud", "Edge"]
    }
  },
  "serviceFlow": [
    {
      "type": "Relation",
```

```
      "child": {"type": "DataContext", "id": "urn:ngsi-ld:DataContext:001"}
    },
    {
      "type": "Relation",
      "child": {"type": "ServiceContext", "id": "urn:ngsi-ld:ServiceContext:001"}
    }
  ]
}
```

Listing 6.2: Service Context

```
  {
  "@context":  "https://uri.etsi.org/ngsi-ld/v1/ngsi-ld-core-context.jsonld",
  "id": "urn:ngsi-ld:ServiceContext:001",
  "type": "ServiceContext",
  "label": {
    "type": "Property",
    "value": "Label for the Service Context"
  },
   "matchPattern": {
      "type": "Property",
      "value": ["DF matching pattern"]
    },
  "deployment_artefact": {
    "type": "Property",
    "value": {
      "kubernetes_service": {
        "apiVersion": "v1",
        "kind": "Service",
        "metadata": {
          "name": "SemanticAdaption-service"
        },
        "spec": {
          "selector": {
            "app": "SemanticAdaption"
          },
          "ports": [
            {
              "protocol": "TCP",
              "port": 80,
              "targetPort": 80
            }
          ],
          "type": "NodePort|External"
        }
      }
    }
```

```
  },
  "operation_type": {
    "type": "Property",
    "value": "DP",
    "description": "Type of operation specific to service processing of DF"
  },
  "lcm_events": ["discover/monitor", "schedule", "re-schedule", "start", "stop", "update", "scale", "upgrade"]
}
```

Listing 6.3: Data Context

```
{
  "@context":  "https://uri.etsi.org/ngsi-ld/v1/ngsi-ld-core-context.jsonld",
  "id": "urn:ngsi-ld:DataContext:001",
  "type": "DataContext",
  "context": {
    "type": "Property",
    "value": "https://schema.lab.fiware.org/ld/context.jsonld",
    "description": "Data Context to be used for semantic adaptation of the DF"
  },
  "label": {
    "type": "Property",
    "value": "DF matching pattern"
  }
}
```

6.3.4.1 Operational Flow

Figure 6.8 presents exemplary (as DENOS offers dynamic implementation context) CP/DP operations in action and interactions across the various layers, showcasing the seamless integration and dynamic processing flow capabilities. The Device layer generates the DF and sends it to the nearby edge for processing. This could be the second event trigger (upstream trigger as data moves upstream) point to initiate the processing flow. The Edge layer (assuming convergence and semantics layer functions hosted in the Edge layer infra) activates the Processing and Service context if the local context matches the DF pattern. If no local context is found for the DF, then it will request the same from the global context (i.e., a knowledge graph) at the Semantic layer through the Convergence layer. Here, the DENOS will always ensure the default context if no matches are found locally or globally to remain backwards compatible with the traditional way of processing DF under the CED continuum. The Processing context then instantiates the Service Context, i.e., service LCM operation (deployment, start, stop, scale, upgrade, delete, etc.).

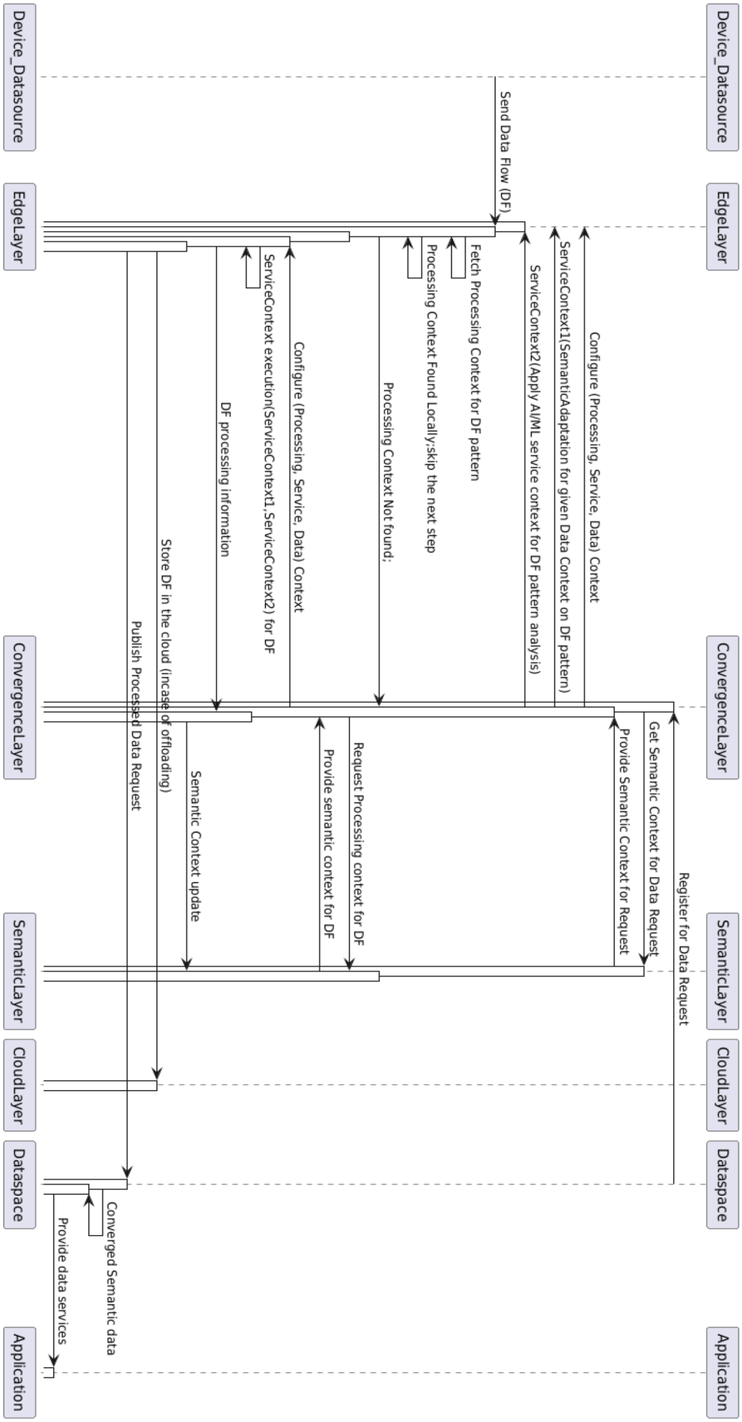


Figure 6.8: Sequence Diagram for Dataspace Data request

Different services can be instantiated in the pipeline to create a Service Flow (SF) to process the DF with different sets of functions at the edge. The SF information can come as part of the initial Data request or be configured statically with a pre-defined set of services in SF for certain DF processing. In the example, the SF contains two service contexts. The first service context is related to the semantic adaptation processing of the DF to augment it with the provided Data Context (e.g., NGSI-LD context). The semantic adaptation function over data can enable data context awareness (e.g., to transform or enrich data with metadata dynamically or based on some predefined pattern). The second service context is related to extracting patterns from DF based on AI/ML (e.g. Random Forests) analysis services to learn about the DF features automatically [93]. Further service context can be added in the SF to process DF with different sets of functionality depending on the use case requirement. For e.g., in some cases, this DF can be put in a private Blockchain network through a DLT-enabled service context so that multi-stakeholders can have a transparent view of the data. One such use case has been explained in section 6.3.10.

The Cloud layer comes into play when data needs to be stored for a longer time or especially in cases of offloading. Finally, the Dataspace integration layer receives the processed data, which is then semantics-converged, if needed, as per the data request. For e.g. two different data flows can be converged semantically to create a new value-chain service to enable the efficient sharing and reuse of diverse data sources belonging to different stakeholders. Finally, Dataspace provides data services to the target Application, allowing access to relevant data for various applications.

This integration flow showcases dynamic processing capabilities and service contexts, allowing for flexibility in data processing. It efficiently utilizes resources and promotes distributed edge-driven service-chain enablement, leading to enhanced operational outcomes within the DENOS framework.

6.3.5 Semantics Driven Deployment

The DENOS model can be represented as a hierarchically layered stack, including the Cloud, Metaedge, Edge, and Device layers. This is shown in Figure 6.9 which illustrates the semantics-driven deployment model context for DENOS, where different instances (with dynamic processing context) of an entity (Cloud, Metaedge, Edge or Device) can be created based on a semantic-driven template. The Metaedge represents the semantics layer in the DENOS model and supports the edge layer, facilitating the integration and interaction of data and service contexts from various sources. The convergence layer, which is considered part of the edge layer in this deployment context, plays a crucial role in data and service integration across the cloud, Metaedge, edge, and device layers. Communication and data exchange among these layers are facilitated through interfaces such as C-ME (convergence and Metaedge for CED continuum-related coordinated actions), ME-E (between Metaedge and edge for edge-edge coordinated actions), and E-D (between edge and device). These interfaces can be based on diverse protocols like REST, MQTT, AMQP, and others [94].

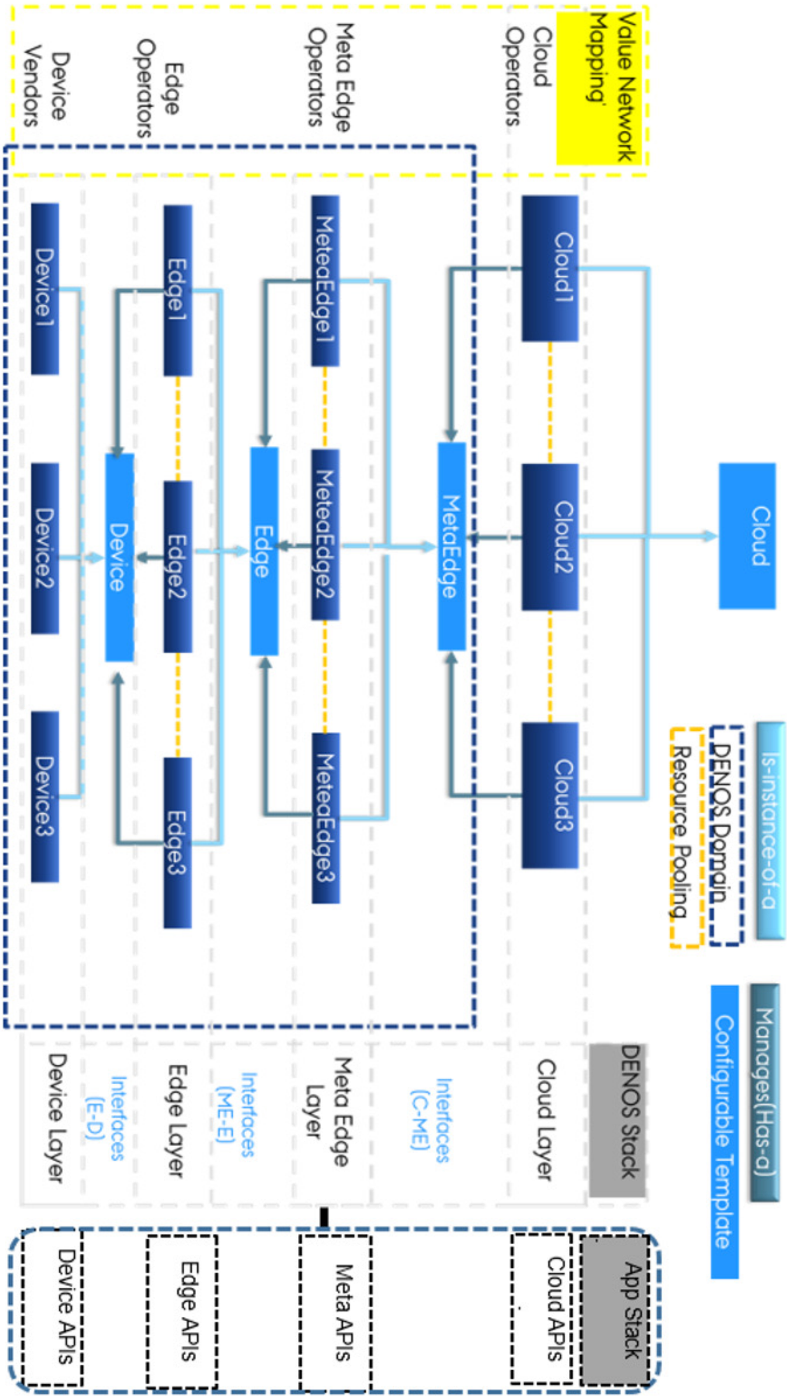


Figure 6.9: DENOS Dynamic Deployment Model

DENOS dynamic deployment model leverages semantic methodologies, such as NGSI-LD, RDF, or OWL, to define semantic contexts and to address heterogeneity, integration, orchestration, and dynamic processing flow challenges across layers during the deployment process. It recognizes the importance of distinct semantic contexts for different layers and focuses on integrating and harmonizing these contexts effectively. The model is designed to incorporate configurable templates or processing contexts to facilitate the semantic provisioning of entities (in diverse deployment contexts) in each layer [87]. Inspired by NGSI-LD, a deployment information model can be developed that encompasses core, domain, and cross-domain specific information, allowing for connection and federation with other information models using linked data approaches like JSON-LD [22]. This is shown in Listing 6.1 to 6.3, where the relevant NGSI-LD information model has been used to define the Processing, Service and Data Contexts. This semantic-driven approach promotes seamless integration, interoperability, and efficient deployment within the DENOS framework. The semantic model of DENOS also enables cross-domain (Dataspace integration) value chain mapping [95] and servitization [76], fostering collaboration and service offerings across organizations and layers. APIs corresponding to different layers support integration and access to services. DENOS ensures seamless distributed operations, integration, orchestration, and lifecycle management, incorporating semantics-related deployment context awareness.

Considering these dynamic aspects, we have provided a reference implementation context of DENOS in the next section to build a further understanding.

6.3.6 Denos Implementation Context Example

As DENOS is an architectural framework, it can have dynamic deployment architecture. Therefore, each layer should be implemented with careful consideration by selecting specific technologies, protocols, and tools for the realization of the DENOS model. The implementation approach may vary depending on the use case. By incorporating semantic context, DENOS can achieve a dynamic realization context (through defining different Processing, Service and Data contexts) of the same model that serves the specific needs of a use case. The DENOS implementation architecture, which consists of several functional blocks and utilizes various open-source tools and technologies, is depicted in Figure 6.10. These blocks include Cloud Hosted Services, Cloud-Edge Service Interfaces, Custom Programming Interfaces, Data Visualization, Semantic Context Interfaces, Distributed Service Chaining, Edge Analytics, Data Ingestion, Metadata Repository, Persistence, Data Interfaces, Transport Protocols, Data Sources, Edge Service Orchestration, and Distributed Security. These blocks use various open-source packages that are listed in their corresponding functional block. These functional blocks provide a range of functionalities in the DENOS framework. They are explained as follows:

- Cloud Hosted Services serve as IoT platforms, providing IoT-as-a-service and data lake services.

- Cloud-Edge Service Interfaces facilitate operational control and service context between cloud, edge, semantic layer, and Dataspace integration applications.

- Custom Programming Interface enables developers to create, share, and consume Dataspace services and data from different entities at the edge.

- Data Visualization provides a visual representation of data, resources and CP operations across DENOS layers.

- Semantic Context Interfaces ensure semantic context-related operations definitions for Processing, Service and Data contexts and their execution based on standards like NGSI-LD, RDF, and OWL. Thus enabling semantics modelling-driven resource discovery and pooling, processing/service/data flows, and related management.

- Distributed Service Chaining is about realizing the Service context across distributed edge infrastructure by aggregating and abstracting resources from multiple edges, facilitating microservices orchestration and collaborative data sharing.

- Edge Analytics leverages domain edge storage and federated learning for localized analytical tasks. Thus providing quick, real-time insights into the data.

- Data Ingestion handles the ingestion of data and semantic adaptation of the data (e.g. augment context or metadata) at the edge using semantic and convergence layers functionality.

- Metadata Repository stores (using Persistence interfaces) semantic contextual data and related framework CP configurations, keeping track of resource discovery, device registration, and availability.

- Persistence block offers the actual storage at the edge for metadata repository CP and DP operations (limited), with the option to offload to cloud storage using defined policies (e.g. Blockchain smart contracts).

- Data Interfaces and Transport Protocols provides the semantic data integration models and transport methods for transferring data to ensure interoperability between communication entities.

- Data Sources provides the source of data through registration of IoT devices, sensors, and other edges or IoT platforms acting as data sources, enabling seamless integration and access to data.

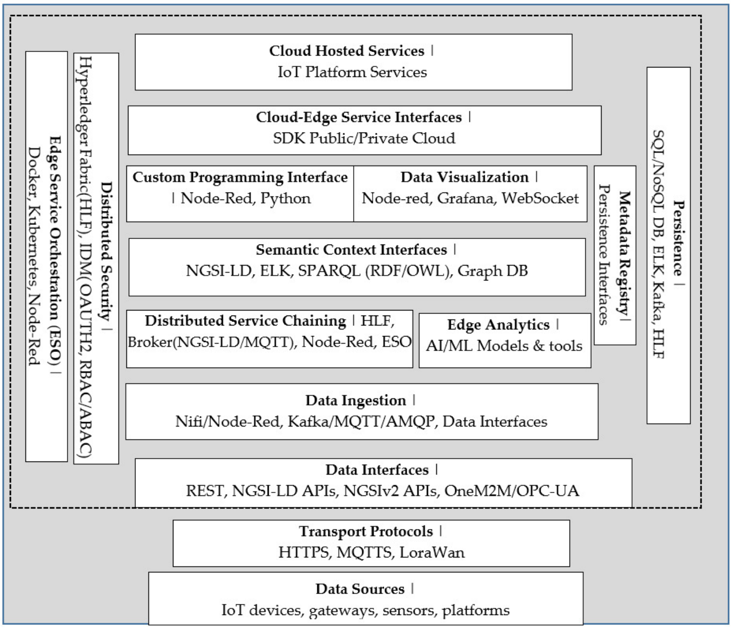


Figure 6.10: DENOS Exemplary System Architecture

- Edge Service Orchestration provides the Processing context implementation by managing the life-cycle of services, resources, tasks, and workload orchestration across multiple layers, optimizing resource allocation and coordination.
- Distributed Security, based on DLT (e.g. HyperLedger Fabric) and traditional security methods (e.g. OAUTH2, OpenID, RBAC/ABAC, etc.), ensures secure registration of devices, edges, and meta edges, with CP operations handled through DLT channels. Also, sensitive data where data governance, transparency, and traceability are of utmost importance, such as data sharing during cross-domain Dataspace value chain development under DP operations, can be handled through the DLT channels. Therefore, this provides distributed security and permissioned DLT segmentation of the edge network, enhancing trust and reliability in Dataspace integration.

By leveraging these functional blocks and technologies, the DENOS model can be realized to address challenges related to security, trust, resource constraints, interoperability, scalability, resource management, and discovery. Thus, it enables the building of a secure, scalable, and interoperable edge ecosystem.

6.3.7 Context Aware Edge Processing Algorithm

DENOS is designed to have a dynamic processing context in which upstream data flows at the edge are identified through matching patterns. These patterns play a crucial role in determining the service context guiding the processing of the relevant data flow. The services can be configured in various combinations to process the relevant data flow in sequence. Each data flow can be assigned to different service pipelines or combinations according to pre-configurations at the edge layer. In this context, a pseudo algorithm is developed to accommodate the dynamic data flow processing approach shown in Algorithm-1, which has used Kubernetes to deploy the DENOS processing context. This demonstrates how a dynamic processing context can be realized at the edge.

Algorithm 1 Dynamic Data and Service Flow Scheduling Using Kubernetes Scheduler

Require:
Set of incoming data flows DF at edge with defined characteristics (tuple attributes)
Configured Processing Context (PC)
Set of Service Context SC definitions i.e. service packages or templates; wherein $SC \leftarrow \{IoT, AI, DLT, SemanticAdaption - BasedServices\}$
Ensure:
Assignment of data flows to service pipelines on edge devices
1: Initialize Kubernetes cluster with controller and worker nodes.
2: Define service deployments:
- $Services(S) \leftarrow \{SC\}$
Define labels and selectors to schedule specific services at target locations:
- Label S with their selectors (Name, type).
- Label worker nodes (edge devices) with their selectors and matching capabilities (e.g., processing power, memory). ▷ This will allow services to be scheduled on target nodes with matching selectors.
3: Orchestrate/Deploy service deployments.
- Create pods P for defined service deployments S.
- Let the Kubernetes scheduler automatically select the node (based on the selector) for each deployed pod.
4: **while** DF is not empty **do**
5: f = next incoming data flow in DF
6: $characteristics$ = get_defined_characteristics(f) ▷ Based on PC
7: $selected\_service\_pipeline$ = select_service_pipeline($characteristics, SC$) ▷ Select service pipeline based on data flow characteristics and defined service context SC.
8: $deployed\_services\_pods$ = get_pods_for_service_pipeline($selected\_service\_pipeline$) ▷ Get pod endpoints for the selected service pipeline.
9: process_dataflow_setup($f, deployed\_services\_pods$) ▷ Process data flow using deployed services.
10: **end while**

6.3.8 Resource Pooling And Offloading

In addition to the proposed model, our objective is also to justify the need for a distributed resource pooling and offloading method that facilitates the aggregation of resources within the edge network and offloads to the cloud when necessary. This method involves discovering resources and allocating workloads based on their processing requirements to nearby edge servers, devices, and gateways. By doing so, we aim to minimize the reliance on centralized cloud facilities, leading to various advantages such as reduced latency, bandwidth consumption, optimized resource utilization at the edge, decreased energy consumption, and cost savings for cloud services [90]. The concept of pooling resources over distributed edges is based on [96], and we have extended it further by incorporating the idea of resource chaining among coordinated edges (see Figure 6.11).

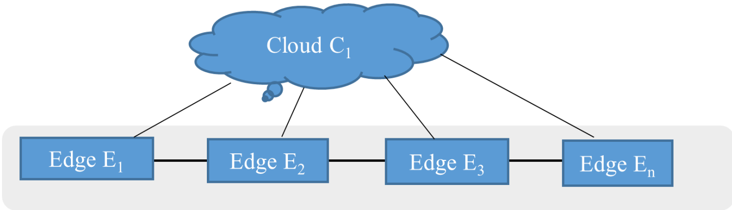


Figure 6.11: Cloud and Edge Interconnection

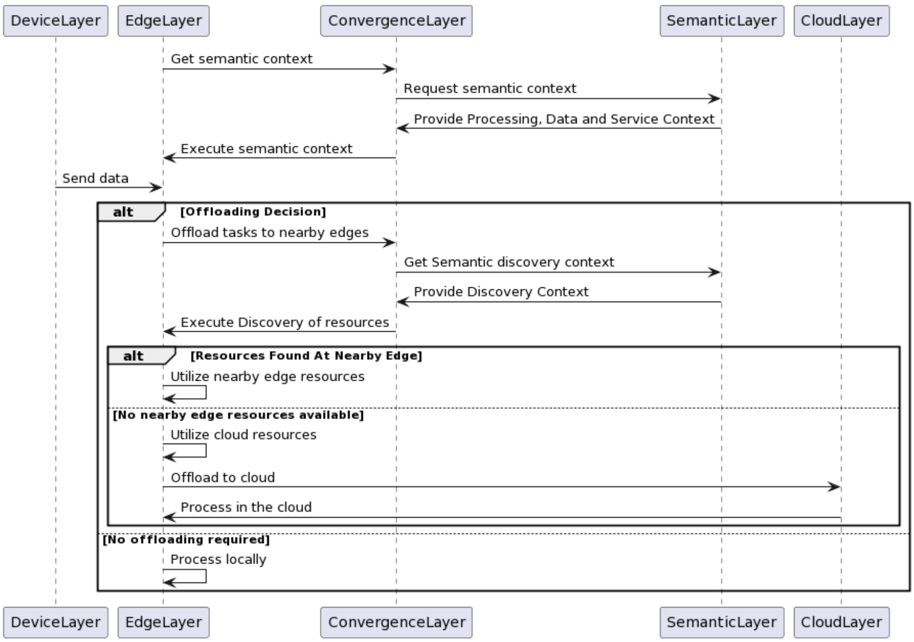


Figure 6.12: Semantic Discovery, Offloading and Utilization

In the described scenario, we have a network of edges represented by $E = \{e_1, e_2, e_3, ..., e_n\}$, where each edge e has resources available for processing workloads at a specific time τ. The resources at each edge include CPU (C), storage (S), and memory (M), which can be calculated using the accumulation function $f()$ as follows:

$$R^{\tau}_{e_i} = f^{\tau}_{e_i}(C, S, M)$$

The cumulative resources available at all edges, R^{τ}_{E}, can be obtained by summing up the resources of each individual edge:

$$R^{\tau}_{E} = \sum_{e=1}^{n} R^{\tau}_{e}$$

These resources are necessary for processing tasks from the task set $T = \{t_1, t_2, t_3, ..., t_n\}$. If a task t_n can be processed at the edge network without the need to send it to the cloud, it becomes an edge task T_E. Otherwise, if it fails to be processed at the edge, it needs to be scheduled in the cloud and becomes a cloud task T_C.

To process a task, the available resources that can be pooled across edges ($R^{\tau}_{p\in E}$) can be calculated as the sum of resources at each edge:

$$R^{\tau}_{p\in E} = \sum_{e=1}^{n} f^{\tau}_{p\in e}(C, S, M)$$

Similarly, the resources available in a given cloud account subscription A at time τ ($R^{\tau}_{c\in A}$) can be determined.

The processing time (λ_{T_E}) for an edge task T_E involves two components: the computation time ($\mu^{c}_{T_E}$) at the edge network and the data transmission time ($\mu^{t}_{T_E}$) from the source to the edge. It can be expressed as:

$$\lambda_{T_E} = \mu^{c}_{T_E} + \mu^{t}_{T_E} \quad ..(A)$$

If a task collects data at the edge but requires processing only in the cloud server, the processing time at the cloud (λ_{T_C}) can be calculated as the sum of the computation time ($\mu^{c}_{T_C}$), the computation time at the edge ($\mu^{c}_{T_E}$), and the data transmission time at both the edge and the cloud ($\mu^{t}_{T_{E,C}}$). In this case, the edge only handles the collection and forwarding of data for processing in the cloud. The relevant processing time at the cloud can be calculated as:

$$\lambda_{T_C} = \lambda_{T_E} + \mu^{c}_{T_C} + \mu^{t}_{T_{E,C}} \quad ..(B)$$

From equations (A) and (B), it can be observed that the processing time at the cloud is always greater than the processing time at the edge due to the additional transmission time from the edge to the cloud. Therefore, if the available resources at the edge network are sufficient to process a task, it should be scheduled at the edge to leverage the shorter processing time. This decision can be made by comparing the processing times for edge and cloud tasks.

To decide whether a task is to be processed at the edge or in the cloud, there is a need for the calculation of computation and transmission times [97]. The computation time can be determined by dividing the workload by the available computing power at the target computation point. The expected computation time at the cloud or edge for a workload task T with length L (in million instructions) and data volume V (to be transferred over the network with bandwidth NB in MBPS) can be calculated.

The expected computation time at the cloud is given by:

$$\mu^c_{T_C} = \frac{T_L}{f^\tau_{p\in A}(C, S, M)}$$

The expected computation time at the edge, considering the discovery of edge nodes and task scheduling is given by:

$$\mu^c_{T_E} = \frac{T_L}{f^\tau_{p\in e_n}(C, S, M)} + \mu^s_T$$

Here, μ^s_T represents the scheduling time. This consists of the time taken for the discovery of the edge node, task scheduling, and the time taken to transfer the task load ($\mu^t_{T_{S,D}}$) from the edge source to the destination edge node [98]. It can be calculated based on the function $f^\tau_{p\in e_n}(d, s, \mu^t_{T_{S,D}})$. This function is needed for the resource pooling scenario at the edge, where any nearby edge node can be selected to process the task.

If the arriving workload task can be accommodated in the same node where it arrives (which is the more likely case for nearby nodes), the scheduling time (μ^s_T) can be ignored, and the computation time at the edge becomes:

$$\mu^c_{T_E} = \frac{T_L}{f^\tau_{p\in e_n}(C, S, M)}$$

The transmission time ($\mu^t_{T_{S,D}}$) for data transfer between the source s and destination d can be calculated as:

$$\mu^t_{T_{S,D}} = \frac{T_V \text{ (task size)}}{N_B \text{ (Network Bandwidth)}}$$

The functions used in the above equations ($f^\tau_{e,p\in E,C}(C, S, M)$ and $f^\tau_{p\in e_n}(d, s, \mu^t_{T_{S,D}})$) should be programmed or modelled based on real systems to determine the available capabilities and utilization of hardware resources through monitoring or predictive modelling.

The distributed resource pooling method described above allows for the optimal utilization of distributed resources at the edge, leading to savings in latency, bandwidth, and cost. It also supports offloading processing to the cloud when required. This can be linked to a scenario described in Figure 6.12 wherein the Device Layer sends data to the Edge Layer, which requests a semantic context to process the received data flow. The Convergence Layer then communicates with the Semantic Layer for related Semantic context and activates the same in the edge layer. Next, the edge layer executes the processing flow based on the received semantic context. In cases where offloading is deemed appropriate, the edge, based on the Processing context, offloads tasks to nearby edges. To determine the availability of nearby edge resources, the edge layer executes semantic discovery context to discover resources. If nearby edge resources are found, they are utilized for service context scheduling. However, if no nearby resources are available, the edge layer utilises the cloud resources. In such cases, the edge layer offloads the tasks to the Cloud Layer, which processes them in the cloud environment. On the other hand, if no offloading (in case the edge has sufficient resources locally and is also allowed by the semantic context) is necessary, the edge layer handles the task processing locally. This system flow enables intelligent decision-making regarding task offloading based on resource availability and utilization at both the edge and cloud layers driven by relevant semantic context.

6.3.9 DLT-Edge Integration Methodology

In the DENOS context, a methodology has been developed to integrate Blockchain with edge systems to support DP and CP operations to provide a secure, transparent, traceable and trusted ecosystem for Dataspace applications. This methodology addresses many of the security challenges of distributed edge systems and the identified limitations of Edge-Blockchain integration.
Issue: As we saw in the literature review, the limitation of DLT (Blockchain) integration lies in its scalability regarding storage and processing. On the one hand, the edge resources are not in abundance, and on the other hand, DLT by design, offers peer-to-peer communication and stores the redundant copy of the ledger in every peer node (as it is distributed) of the DLT network. This poses huge scalability and storage issues, along with delays in the execution of consensus mechanisms among peers in the DLT network.
Methodology: To address the above issue, there is a need for partitioning or segmentation strategy motivated by well-established TCP/IP communication-related VLAN standards [99] and supported by network segmentation-based DLT realization strategy, e.g. using HyperLedgerFabric implementation [100]. In the DENOS context, this segmentation strategy provides five different levels of integration DLT with the edge. The first level includes the selection of a DLT network wherein instead of choosing a public Blockchain, private Blockchain network should always be preferred in the edge ecosystem. After that, the second level comes when a single private network is further segmented into subnetworks (often called channels in Hyperledger Fabric DLT implementation) [101]. The third level comes at the channel level, wherein peers interact with each other to perform different tasks such as consensus, proof of work, signing of transactions, mining/building of blocks etc. [90]. At the third level, each channel or subnetwork can be perceived as a source for a local pool of data within one domain. The fourth level comes at the intercommunication between different channels to build a global pool of data, across domains governed by smart contract-driven policies. The final and fifth level comes wherein applications and services interact to consume or produce data to/from the DLT-Edge-enabled network.

To overcome challenges such as resource constraints and scalability of

Blockchain, this methodology suggests using the concept of Blockchain channels, akin to the concept of a tunnel, inspired by Hyperledger fabric [101]. Channels act as private and permissioned sub-networks of peers, enabling secure data transactions and storage based on access controls and smart contracts. Peers in edge networks can join specific channels (at the local or global level) to exchange or share information, ensuring data's scalability, traceability, and trustworthiness. Therefore, related peer data can be transported over corresponding channels securely and safely between authenticated and authorized peers based on smart contract-driven data classification policies. This integration offers scalability benefits both horizontally (between edge-edge peers) and vertically (between edge-cloud, edge-Metaedge, or edge-Dataspace).

AI integration in the Edge-DLT integration context enables the application of semantic patterns (which can be supported by the Convergence layer through AI-driven, e.g. Random Forest or K-means, semantic adaptation) governed by smart contracts-driven policies to automatically classify incoming data streams, determining whether the data should be processed within or outside the Blockchain channels. For instance, sensitive Control plane data (sensitive to the network) related to device management can be mandated to be processed within the DLT channel, while the processing of device measurement data, i.e. Data plane data can be decided to be processed with or without Blockchain channels based on specific needs such as interoperability with legacy systems where DLT support is not there.

This methodology suggests using a hybrid edge-to-edge or edge-cloud (in case of offloading) enabled Blockchain approach, shown in Figure 6.13. A centralized global view of the distributed edge is maintained, where DLT services are instantiated as an extension to other IoT platform services. The DLT services can also act as digital twins of the meta (reflecting characteristics of semantic and convergence layers of DNEOS) and edge layer peers, storing a centralized (at Dataspace level) and local (at edge layer) copy of the data coming from distributed edge peers and based on smart contracts driven policies.

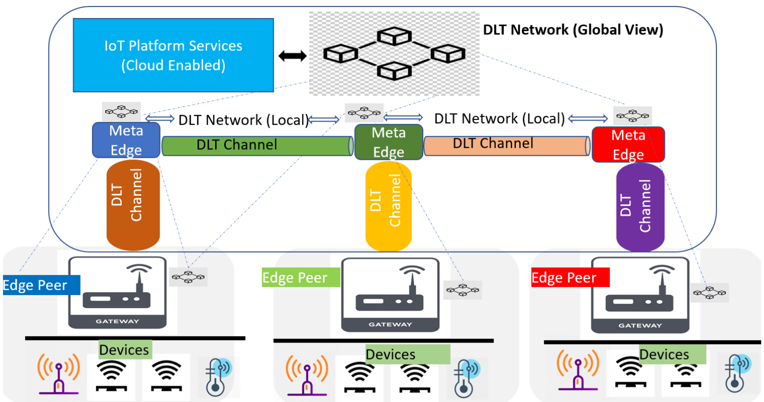


Figure 6.13: Blockchain-Edge Integration Methodology

By implementing this methodology, DENOS model can achieve secure storage, mitigate scalability issues to a large extent, and maintains a global view of the distributed edge for backup and historical purposes. It enables efficient communication, governed by smart contracts, among meta and edge instances through edge channels and meta channels described already. The methodology's realization is backed and supported by different research [58], [59], [75], [101], [102] in the Edge-DLT integration field, demonstrating its potential to address the security challenges combining with traditional security methodologies (e.g. OAUTH2, OpenID, RBAC, ABAC, etc.) and limitations associated with distributed edge systems. By leveraging this approach, DENOS addresses the challenges associated with data control, storage, and scalability requirements within Blockchain-based networks.

Advantages

- Tackles large storage capacity challenges for Blockchain usage at the design level as it reduces storage overhead due to channel segmentation which allows peers to have channel-specific storage only.
- Scalable Blockchain design as the number of channels can grow with the number of gateways grows in the field but each edge and meta edge and related devices can partition or segment their communication in multiple channels which are independent of each other.
- The flow of information in Blockchain networks among peers is controlled via channels binding to different roles.
- Data entering into channels can be classified or labelled further in the control plane (related to the management of devices) and data plane (related to sensing of the environment) using smart contracts and AI based models. This classification will enable performance-efficient data processing flows such as with or without Blockchain.
- Control plane sensitive operations like device management, update or upgrade, discovery or monitoring can be supported through these secure and encrypted channels.
- Enablement of cross-domain (e.g., at the organization level) Dataspace integration and information sharing through granular control, security, trust, transparency, traceability and immutability.

6.3.10 Validation and Use Case - Smart Wind Turbine Monitoring and Optimization

Scenario: This use case has focused on a wind farm equipped with IoT-enabled functions on wind turbines. The objective is to efficiently support digitized operations of bolts on the wind turbine during different phases of the turbine life cycle, predict maintenance needs, and ensure transparent, traceable and trusted information flow in a multi-stakeholder environment. This use case is already demonstrated in [79] as part of the Danish Unwind Project [103] funded by the Danish Industries Foundation. However, the same case is reused and extended in the DENOS model validation context. The relevant test setup is shown in Figure 6.14.

Hardware Components: Wind Turbine, Digital Wrench, QR code Scanner,

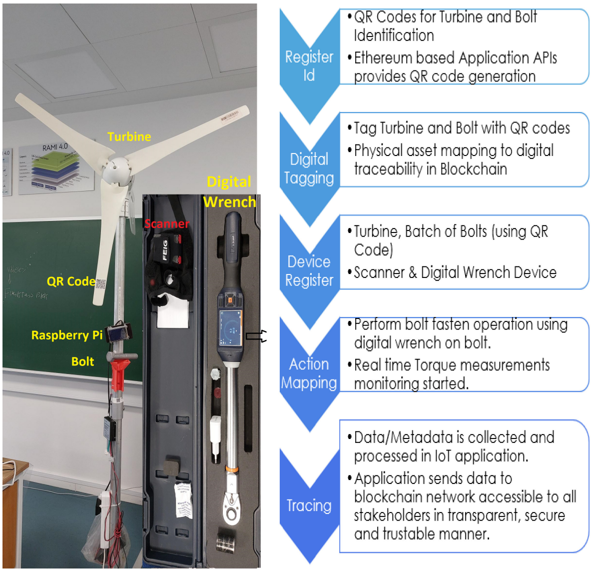


Figure 6.14: Wind Turbine Use Case Test Setup

Bolt, QR codes, Edge device (Raspberry Pi4.0), and Edge/Cloud Server (Ubuntu20.04, 8 GB RAM, 2 VCPU, 60 GB HDD).
Software Components:DLT service (HyperLedgerFabric/Ethereum test net for private DLT networks), Odoo-based ERP to replicate traditional cloud-enabled centralized systems, Node-Red, NodeJS, JavaScript, Python.
Stakeholders Involved: Bolt Vendor, Turbine Operator, Service Engineer (SvcEng.), Insurance official
Requirement: End-to-end digitised bolt operation (latency sensitive to apply precise torque on the bolt) on a wind turbine to ensure quality, error avoidance, and proactive maintenance. In addition, the data during the assembly and maintenance phase should be able to be used during insurance claims of bolt damages within the life cycle of the turbine.
Challenges Involved: Multivendor devices, data and interfaces heterogeneity, integration and interoperability, multiple stakeholders, distributed resources and real-time operations, data and operations quality, security, trust, transparency and traceability.

Implementation Context: To implement the use case, a lab-controlled setup and relevant processing and service context flow are shown in Figure 6.14 and Figure 6.15 respectively. The turbine and Bolt are tagged with the QR codes to represent their physical identity in virtual space. The scanner is the device that scans the QR code for the turbine and bolt. The wrencher device reads the digital torque reading for bolt-tightening operation. The scanner and wrencher device communicates to a nearby edge device (Raspberry Pi) over Bluetooth interface. The data received by the edge device is treated as data

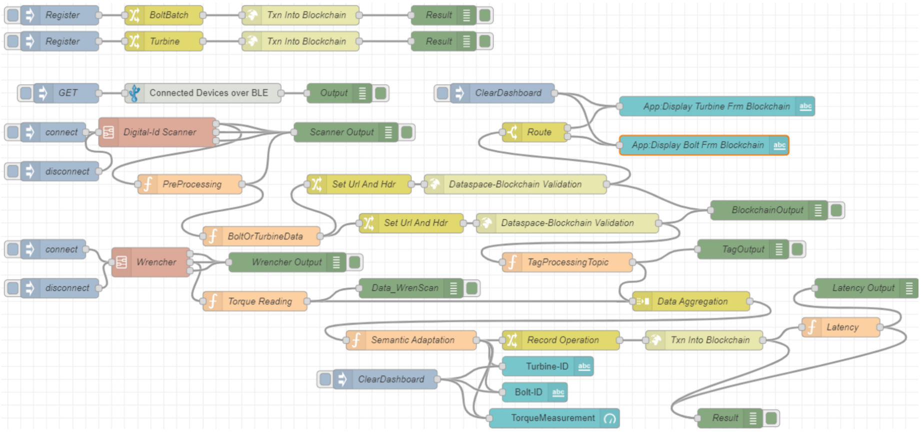


Figure 6.15: Use Case: Processing and Service Context Flow

flow and aggregates at the edge for different events (Turbine/Bolt QR codes and Torque reading) to apply a semantic relationship between Turbine, Bolt and related torque reading with timestamps using semantic adaption service running at the edge. In addition, the incoming data events are also validated against their identity relevance based on the DLT service, where the stakeholder registers the turbine and bolt related initial metadata. Finally, the validated, semantically adapted and aggregated data is sent to the DLT service, where different stakeholders read this data in a secure, transparent, traceable manner for different purposes. The operators can use this data for predictive maintenance and for other ERP purposes. The vendors/insurance entity can check this data to assess the quality of operations level, for example, to support the damage or insurance claim when some unforeseen event occurs at the turbine site.

Operational Flow: The operation flow is presented in the Figure 6.16. In this scenario, a stakeholder, represented by the vendor of the bolt batch or operator of the turbine, uses the Dataspace interface to register bolts and turbines by generating QR codes for the relevant physical entity's digital identification in the virtual world. A Distributed Ledger Technology (DLT) Service (running at the edge instantiated by the predefined processing context) handles blockchain transactions over a permissioned DLT network (i.e. Ethereum/Hyperledger fabric), receiving transactional data from the Dataspace App. This registration data is communicated to an Edge Device, facilitating the registration validation of scanners and wrenches over Bluetooth. A loop then ensures for each QR code related to bolt and turbine identification, involving the scanning of codes, fastening operations, and the real-time transmission of torque readings sent to the Edge Device. The Edge Device performs semantic adaptation and validation as per the configured registration (i.e. Data context), sending integrated data (torque readings and semantic metadata) to both the DLT Service and a Traditional System via REST APIs. The Traditional System receives upstream data, while the DLT Service relays blockchain transactional data back to the

Dataspace App. The relevant stakeholder (Operator/vendor/Insurance official) receives upstream data from the Dataspace App, completing the operational flow of the turbine system integration.

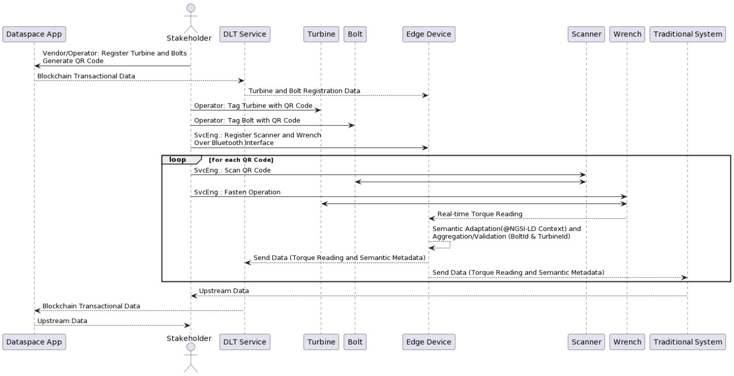


Figure 6.16: Use Case: Digitized and Traceable Edge Enabled Wind Turbine Operations

Architectural Mapping: The exemplary architecture shown in Figure 6.10 has been used to realize the presented use case. The relevant architectural functional, technology, and their usage in the use case context have been shown in Table 7.1.

Result and Discussion: The use case has been validated for capturing the Torque reading operation latencies when the DLT and Dataspace application (to display torque(from service engineer)/turbine(from operator)/Bolt(from vendor) data) services run in the cloud compared to the edge, as shown in Figure 6.17. However, the service context (Node-red that terminates the Bluetooth interface, performs semantic adaption and interfaces with the DLT and Dataspace application) shown in Figure 6.15 always runs in the edge device in both scenarios. We have used an edge server under the Edge-Edge scenario instead of the cloud-based server that is instantiated under the CED scenario to host the related service context for DLT and Dataspace services. In addition, the torque operation was performed for multiple iterations to get maximum latency for four iterations for both scenarios. The average latency for the CED case is found to be around 1 second, and for Edge-Edge, it is around 0.43 seconds. As we can see, the latency is much lower when the complete service context (including Semantic Adaptation, DLT, and Dataspace) runs on the edge, which is an obvious advantage of running applications on the edge layer.

Validation Aspects: In the presented use case, the convergence of IDEAL technologies has been presented in different contexts. IoT is used for bolt, turbine and QR code-related data generation, collection and processing at the edge. Afterwards, AI-based models (Isolation Forest and CNN-based) are proposed to predict anomalies in data quality and maintenance needs. Predefined Processing context across distributed edge layer is realized using Kubernetes and Docker tools. Service contexts have been implemented at the edge using different services,

Table 6.4: Use Case and Architectural Functional Mapping

Functional Block	**Technology Used**	**Use Case Usage Context**
Data Sources	IoT	Wrench, Scanner, QR code (Bolt and Turbine)
Transport Protocols	HTTPS, Bluetooth (BLE)	Torque reading, QR code scanning data stream used BLE between turbine and edge device. HTTPS is used between Edge device and other services interacting with it
Distributed Service Chaining (Service Context)	Node-Red	Realization of service context (BLE, DLT, Dataspace handlers)
Data Ingestion	Node-Red, REST and BLE	Ingest data for onsite turbine operations
Semantic Context	JSON, NGSI-LD (Node-Red based)	For Semantic Adaptation
Custom Programming Interface	Node-Red, NodeJS	For implementing Service Context at the edge
Persistence	HyperledgerFabric/Ethereum, Postgres	Store DLT transactions and traditional system data
Distributed Security	HyperledgerFabric/Ethereum, OAUTH2	For CP operations and User Identity Management
Edge Service Orchestration	Kubernetes/Docker	For implementing predefined Processing context
Data Visualization	Node-Red-Dashboard	For Dataspace App (frontend part), displaying Turbine, Bolt and Torque Data

such as the DLT-Service and Dataspace Application. Also, the role of the cloud in the CED continuum to host DLT/Dataspace services and the possibility of realizing Dataspace directly at the edge using DLT secure networks have been demonstrated. Semantic Adaption (Node-red library) to augment the data with the context (e.g. NGSI-LD core context) after the data aggregation (combining data from different streams of Bolt, turbine and Torque reading) has also been validated. Beyond that, validation of data streams at the edge against the registered data received from the DLT is also achieved.

By employing DENOS exemplary architectural functions in the processing

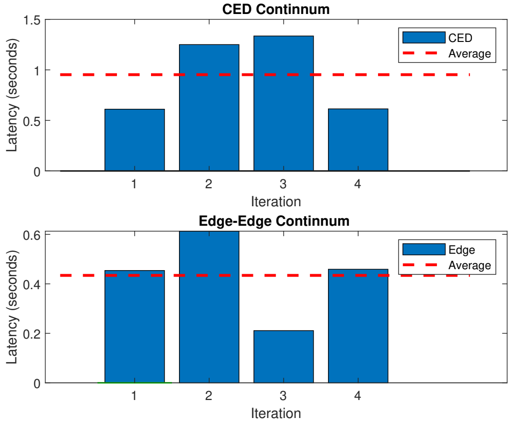


Figure 6.17: CED vs Edge-Edge Continuum Operation Latency

pipeline, this wind turbine use case demonstrates an advanced system for monitoring and optimizing wind farm operations. It ensures efficient energy production, reduces maintenance costs, and maintains transparent, traceable and trusted record-keeping for the bolt-turbine value chain among different stakeholders.

Limitation or Extension Possibilities: In the presented use case, AI context has not been implemented, but this is very relevant to generate predictive modelling using certain anomaly detection algorithms. In the use case, when torque reading is received at the edge, a pre-trained ML model (e.g. Isolation Forest or K-Means Clustering) service can be instantiated to learn about anomalies in the data streams and to generate relevant alarms to stakeholders (e.g. Service Engineer.). In addition, ML-driven CNN models can predict maintenance needs by evaluating the condition of bolts (e.g., rust, crack, etc.) based on drone-enabled turbine inspection video stream inputs. This leads to the generation of alerts for scheduled maintenance activities. Apart from that the use case has used static and pre-defined Processing (Kubernetes/Docker templates), Data and Service context at the Edge-Edge and CED continuum. This can be further made semi or fully automated.

6.3.11 Convergence Significance of Technologies in DENOS

The relevance of the convergence of technologies in DENOS lies in its ability to harness the collective power of IDEAL technologies. This enables converged capabilities, as shown in Figure 6.18, to address many of the challenges and drive innovation in edge networks. Here are the key aspects of the relevance of convergence in DENOS:

Enhanced Capabilities and Responsiveness: The convergence of technologies enables edge networks to operate with greater efficiency and responsiveness. IoT can provide data and related processing in real-time at the edge or at cloud-enabled platform services. AI algorithms can optimize resource allocation, predict future events, and automate tasks. Edge computing brings computing power closer to data sources to reduce latency and enable real-time processing. DLT ensures secure and transparent transactions, thereby enhancing trust and traceability aspects.

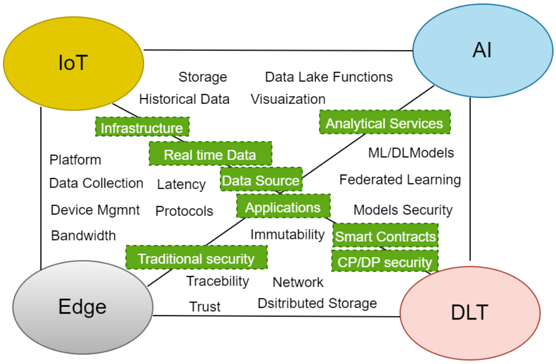


Figure 6.18: Converged Capabilities Of IDEAL Technologies

Scalability and Efficiency: The convergence of technologies supports addressing scalability challenges by distributed computing, storage, and decision-making capabilities through AI-driven predictive modelling at the edge. This enables horizontal scalability in the edge network to handle data volumes and scalable resources wherein the addition or removal of devices and services without disrupting the overall system can be achieved. Thus providing scalability with agility.

Security and Trust: Integrating DLT and edge computing in DENOS enhances security and trust in edge networks, in addition to the traditional security mechanisms such as OAUTH2, OpenID, etc. DLT, at the edge layer, provides a decentralized and tamper-proof network for data-related transaction validations and can be combined with traditional methods to safeguard identity management, identity theft, DDoS attacks, privacy leakages and data provenance [104]. This integration provides an additional layer of security that enables sensitive data protection closer to its source. Thereby reducing the risk of data exposure during transmission.

Therefore, the convergence of IDEAL technologies is highly relevant as it brings together synergies and the strengths of each technology to create innovative and powerful solutions for edge network operations.

6.3.12 How DENOS Model can Address the Challenges?

DENOS provides an architectural framework that offers a convergence of technologies and provides semantics-driven edge processing to address a wide range of relevant challenges. Let's see some of them:

Security and Privacy: By integrating Blockchain technology, DENOS ensures secure data storage, authentication, and access control, enhancing security and privacy at the edge. Through DLT's consensus mechanisms and smart contracts, DENOS establishes trust and reliability, enabling transparent and verifiable transactions and interactions between edge entities. This is applicable for both CP and DP operations. For example, under CP operation, DLT-driven smart contracts can be used to implement certain policies among stakeholders of the Edge infrastructure where discovery and resource pooling of resources can be allowed in a controlled manner. All CP operation-associated control data can be transferred over DLT channels between the peers at the edge to offer secure and tamper-proof communication. Similarly, transparency and traceability, along with immutable security, can be implemented using DLT for DP operations for related data flows. This would not have been possible by using only traditional security methods such as OAUTH2, Open-ID, etc.

Resource Constraints: DENOS aims to optimize resource utilization by facilitating the use of AI (out of IDEAL) in CP operations at Edge, such as intelligent resource scheduling management through predictive modelling and allocating matching (to service deployment requirements) resources among coordinated edge-edge devices. In addition, DENOS allows the static and pre-defined Processing and Service context to be used in a distributed manner. This means SF consisting of multiple services can be deployed over the chain of edges in a coordinated manner to offer SF processing for related DF. The chain of coordinated edges offers sufficient computing capabilities to deploy SF at the edge to suffice the functional processing requirements of the given DF as per the defined Processing context. Thus ensuring efficient performance, by design, in resource-constrained environments through distributed resource-pooling to instantiate the service contexts to alleviate resource constraints at one node.

Interoperability and Integration: DENOS provides a semantics-driven interoperability framework that enables seamless Dataspace integration and interoperation between heterogeneous devices, data sources, systems, platforms, etc. This can be achieved by defining relevant semantic contexts targeting specific use case using common ontologies, semantic mappings, and reasonings compliant with interoperability standards such as NGSI-LD, RDF or OWL etc. Thereby, DENOS ensures compatibility and seamless interoperations at the edge to support dynamic processing, service and data context for relevant data flow integration among heterogeneous entities. DENOS is designed to integrate with existing legacy systems and technologies adaptively. It allows for the gradual migration and integration of legacy systems into the edge ecosystem, ensuring backward compatibility and interoperability.

Scalability: DENOS has the elasticity and distributed nature of edge computing. It allows allocation of resources, distributed loads and dynamic data processing contexts. Thus, enabling scalability while maintaining low latency and efficient performance at the same time.

Management and Orchestration: DENOS incorporates collaborative, static or autonomous orchestration of resources, tasks and services at the edge using corresponding Processing, Data and Service context definitions. All CP operations (discovery, registration, monitoring, etc.) and their related LCM (deploy, start, stop, delete, etc.) events are associated with the orchestration and management process. Thereby minimizing centralized overhead, enhancing fault tolerance, and enabling efficient system management in distributed environments.

Resource Discovery and Pooling: DENOS facilitates efficient semantic resource discovery by enabling edge entities to register and advertise their resources. It incorporates semantic technologies for standardized and machine-understandable descriptions e.g. RDF, NGSI-LD [22]. Thus, enhancing machine-readable resource matching and discovery-related context awareness. This enables the creation of distributed resource pools and aggregating resources from multiple edge entities for efficient utilization across the network.

Optimal Resource Utilization: DENOS optimizes resource utilization by considering proximity to data sources, network latency, and dynamic processing (using Processing and Service context) capabilities. It enables AI-based workload placement, task offloading, intelligent semantic data adaptation, filtering and aggregation techniques at the edge.

6.3.13 How DENOS is Different from Traditional Approaches?

The DENOS model introduces four main novel aspects that differentiate it from traditional approaches. This is given as follows:

Semantic Modelling: The DENOS model incorporates semantic modelling and ontology-based approaches to enable a common understanding of data, processing, services, and related resources. The semantic layer facilitates the same by offering interoperability, resource or service discovery, dynamic deployment and efficient utilization across heterogeneous devices, systems, and platforms.

Convergence of Technologies: The DENOS model brings together the convergence power of IDEAL technologies. It integrates these technologies synergistically to leverage their respective strengths and capabilities. This enables intelligent data processing, secure and transparent transactions, and decentralized decision-making.

Dataspace Integration: DENOS harnesses the power of cross-domain data integration(supported by other layers' functionality) at the edge by enabling the creation of innovative applications which can address complex challenges across different domains and thereby build value chain networks.

Trust and Security: The DENOS model introduces trust and security, incorporating IDEAL capabilities in distributed edge networks. It leverages DLT (in addition to traditional security methods) to provide a decentralized and secure framework for data sharing in Dataspace and control operations beyond verification, validation and identity management. This enables the integrity, privacy, transparency, traceability and authenticity of data and transactions in a distributed environment.

Overall, the novel aspects of the DENOS model lie in its semantic modelling approach, convergence of technologies and resources, building trusted value chain networks using cross-domain Dataspace integrations of assets, distributed resource optimization, and security.

6.3.14 Limitations And Outlook

The study proposes a DENOS model with semantic, convergence, and dataspace capabilities to address challenges of heterogeneity, data integration, distributed security, and resource optimization at the edge layer. However, there is a need to detail further how these capabilities can be used in different contexts to benefit from their relevant strengths. For example, how Processing context can be implemented using different orchestrators or tools at the edge, cloud or a hybrid model, along with relevant experiments and results. This also includes discovery and resource registration functional flow aspects. For semantic contexts, there is a need to improve the information model further (e.g., using already defined ontologies models from different domains) using NGSI-LD and RDF interdependencies. To build a cross-domain Dataspace ecosystem at the edge leveraging DENOS, there is a need to demonstrate more use cases. In this direction, we have identified edge-driven and drone-assisted maintenance operations on wind turbines to extend our current use case. We have already mentioned the high-level AI algorithms (e.g. Random Forest, Isolation, Reinforcement Learning, CNN, etc.) that can be used in different situations. However, there is a need to showcase the capabilities of AI use at different processing levels in CP and DP operations through concrete implementation. We have covered the aspects of security using DLT and related traditional security in combination at a high level. However, having a detailed cyber security perspective on this topic in the DENOS context in the future will be more impactful.

6.4 Conclusion

This study has presented a comprehensive literature review wherein we have explored the convergence of AI, IoT, DLT, and edge technologies to identify relevant trends, challenges and gaps that hinder the effective integration of resources and utilization of these technologies for building Dataspace ecosystem in distributed edge networks. The identified challenges include security, trust, heterogeneity, resource constraints, interoperability, scalability, and resource management complexities. To address these challenges and bridge the architectural gaps, DENOS (Distributed Edge Network Operations oriented Semantic) model has been proposed. DENOS extends the traditional Cloud-Edge-Device architecture to incorporate semantics context awareness, convergence of technologies/resources, and Dataspace integration to build value chain networks to provide a holistic architectural framework for distributed edge networks.

By leveraging distributed layered architecture and semantic context utilization, DENOS enables the integration of diverse data sources, allowing for semantic interoperability and meaningful data exchange among different components in Dataspace applications. It facilitates the integration of AI, IoT, and DLT driven service contexts at the edge by defining dynamic processing and deployment contexts. Beyond that, it also addresses security and privacy concerns of Control and Data plane operations by incorporating distributed security mechanisms leveraging DLT. To demonstrate the validation of the model, a use case pertaining to the digital traceability of operations in the wind energy domain has been presented. As an outlook, we foresee further detailing the different aspects pertaining to DENOS and advancing this work through a complete model prototype implementation. We firmly believe that DENOS has the potential to unleash new business constellations driven by Dataspace integration and foster value chain networks at the edge that can fuel innovation and drive growth in the future for Industry 5.0 applications.

References

[1] Haucap, J., Fritz, D., and Thorwarth, S., "The economic impact of cloud computing in europe," Dice Consult, Tech. Rep., 2022, pp. 1–11.

[2] Research, D.-G. for and Commission), I. (, *Science, research and innovation performance of the EU, 2018: strengthening the foundations for Europe's future*, eng. LU: Publications Office of the European Union, 2017, ISBN: 978-92-79-69744-9.

[3] Sadeeq, M. M., Abdulkareem, N. M., Zeebaree, S. R., Ahmed, D. M., Sami, A. S., and Zebari, R. R., "Iot and cloud computing issues, challenges and opportunities: A review," *Qubahan Academic Journal*, vol. 1, no. 2, pp. 1–7, 2021.

[4] Gill, S. S., Tuli, S., Xu, M., Singh, I., Singh, K. V., Lindsay, D., Tuli, S., Smirnova, D., Singh, M., Jain, U., *et al.*, "Transformative effects of iot, blockchain and artificial intelligence on cloud computing: Evolution, vision, trends and open challenges," *Internet of Things*, vol. 8, p. 100 118, 2019, Times cited: 3 publisher: Elsevier.

[5] Thakur, S. and Chaurasia, A., "Towards green cloud computing: Impact of carbon footprint on environment," in *2016 6th international conference-cloud system and big data engineering (Confluence)*, IEEE, 2016, pp. 209–213.

[6] Iggena, T., Ilyas, E. B., Fischer, M., *et al.*, "Iotcrawler: Challenges and solutions for searching the internet of things," *Sensors*, vol. 21, no. 5, 2021, Times cited: 3, ISSN: 14248220. DOI: 10.3390/s21051559.

[7] *Edge Computing | Trends and Benefits | Lumen*, 2023.

[8] Riemenschneider, R., "Next-generation internet of things and edge computing," en, Euorpean Commission, Tech. Rep., 2022.

[9] Gill, S. S., Tuli, S., Xu, M., Singh, I., Singh, K. V., Lindsay, D., Tuli, S., Smirnova, D., Singh, M., Jain, U., *et al.*, "Transformative effects of iot, blockchain and artificial intelligence on cloud computing: Evolution, vision, trends and open challenges," *Internet of Things*, vol. 8, p. 100 118, 2019.

[10] Barthélemy, J., Verstaevel, N., Forehead, H., and Perez, P., "Edge-computing video analytics for real-time traffic monitoring in a smart city," *Sensors*, vol. 19, no. 9, p. 2048, 2019.

[11] Pathak, T., Patel, V., Kanani, S., Arya, S., Patel, P., and Ali, M. I., "A distributed framework to orchestrate video analytics across edge and cloud: A use case of smart doorbell," in *Proceedings of the 10th International Conference on the Internet of Things*, 2020, pp. 1–8.

[12] Varghese, B. and Buyya, R., "Next generation cloud computing: New trends and research directions," en, *Future Generation Computer Systems*, vol. 79, pp. 849–861, Feb. 2018, ISSN: 0167739X. DOI: 10.1016/j.future.2017.09.020.

[13] Khalaf, K., Nawafleh, O., and Al-Utaibi, G., "The emergence of edge computing technology over cloud computing," *International Journal of P2P Network Trends and Technology*, vol. 10, pp. 1–5, Mar. 25, 2020. DOI: 10.14445/22492615/IJPTT-V10I2P401.

[14] Singh, P. and al., et., "Meta standard requirements for harmonizing dataspace integration at the edge," in *2023 IEEE Conference on Standards for Communications and Networking (CSCN)*, 2023.

[15] Curry, E., *Real-time Linked Dataspaces: Enabling Data Ecosystems for Intelligent Systems*, en. Cham: Springer International Publishing, 2020, ISBN: 978-3-030-29664-3 978-3-030-29665-0. DOI: 10.1007/978-3-030-29665-0.

[16] Hutterer, A. and Krumay, B., "Integrating heterogeneous data in dataspaces-a systematic mapping study," *Pacific Asia Conference on Information Systems*, 2022.

[17] Sipola, T., Alatalo, J., Kokkonen, T., and Rantonen, M., "Artificial intelligence in the iot era: A review of edge ai hardware and software," in *2022 31st Conference of Open Innovations Association (FRUCT)*, IEEE, 2022, pp. 320–331.

[18] Wu, Y., "Cloud-edge orchestration for the internet of things: Architecture and ai-powered data processing," English, *IEEE INTERNET OF THINGS JOURNAL*, vol. 8, no. 16, pp. 12 792–12 805, Aug. 15, 2021, ISSN: 2327-4662. DOI: 10.1109/JIOT.2020.3014845.

[19] Al Ridhawi, I., Aloqaily, M., Boukerche, A., and Jararweh, Y., "Enabling intelligent iocv services at the edge for 5g networks and beyond," English, *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS*, vol. 22, no. 8, pp. 5190–5200, Aug. 2021, ISSN: 1524-9050. DOI: 10.1109/TITS.2021.3053095.

[20] Yazid, Y., Ez-Zazi, I., Guerrero-González, A., El Oualkadi, A., and Arioua, M., "Uav-enabled mobile edge-computing for iot based on ai: A comprehensive review," *Drones*, vol. 5, no. 4, p. 148, 2021, Times cited: 1 publisher: MDPI.

[21] Waher, P., *Mastering IoT: Design and create your own IoT applications using Raspberry Pi 3*, en. s.l.:Packt Publishing Ltd, 2018, ISBN: 978-1-78839-743-8.

[22] CIM, E., "Ngsi-ld cim 009," ETSI, Tech. Rep., 2019, Times cited: 3.

[23] D.T., M. and E, R., "Sme 4.0: The role of small- and medium-sized enterprises in the digital transformation," en, in *978-3-030-25424-7*, ed., Ed., Cham: s.l.:Palgrave Macmillan, 2020, pp. 3–36.

[24] Tu, M., "An exploratory study of internet of things (iot) adoption intention in logistics and supply chain management a mixed research approach," en, *The International Journal of Logistics Management*, 2018.

[25] Tziouvaras, A. and Foukalas, F., "Edge ai for industry 4.0: An internet of things approach," in *24th Pan-Hellenic Conference on Informatics*, 2020, pp. 121–126.

[26] Nguyen Gia, T., Nawaz, A., Peña Querata, J., Tenhunen, H., and Westerlund, T., "Artificial intelligence at the edge in the blockchain of things," in *Wireless Mobile Communication and Healthcare: 8th EAI International Conference, MobiHealth 2019, Dublin, Ireland, November 14-15, 2019, Proceedings 8*, Springer, 2020, pp. 267–280.

[27] Chen, B. and Q. Z, J., "Edge computing in iot-based manufacturing," en, *IEEE Communication Magazine*, 2018.

[28] Pan, J., Wang, J., Hester, A., Alqerm, I., Liu, Y., and Zhao, Y., "Edgechain: An edge-iot framework and prototype based on blockchain and smart contracts," *IEEE Internet of Things Journal*, 2019, Times cited: 1 arXiv: 1806.06185, ISSN: 23274662. DOI: 10.1109/JIOT.2018.2878154.

[29] Wu, Y., Dai, H.-N., and Wang, H., "Convergence of blockchain and edge computing for secure and scalable iiot critical infrastructures in industry 4.0," *IEEE Internet of Things Journal*, vol. 8, no. 4, pp. 2300–2317, 2020, Times cited: 2 publisher: IEEE.

[30] Makhdoom, I., W. N., M., and Jan, "Blockchain's adoption in iot: The challenges, and a way forward," en, *Journal of Network and Computer Applications*, vol. 125, pp. 251–279, 2019.

[31] Kumar, T., Harjula, E., Ejaz, M., Manzoor, A., Porambage, P., Ahmad, I., Liyanage, M., Braeken, A., and Ylianttila, M., "Blockedge: Blockchain-edge framework for industrial iot networks," English, *IEEE ACCESS*, vol. 8, pp. 154 166–154 185, 2020, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2020.3017891.

[32] Du, Y., Wang, Z., and Leung, V. C., "Blockchain-enabled edge intelligence for iot: Background, emerging trends and open issues," *Future Internet*, vol. 13, no. 2, p. 48, 2021, Times cited: 2 publisher: MDPI.

[33] Zhu, C., Zhu, X., Ren, J., and Qin, T., “Blockchain-enabled federated learning for uav edge computing network: Issues and solutions,” English, *IEEE ACCESS*, vol. 10, pp. 56 591–56 610, 2022, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2022.3174865.

[34] Casado-Vara, R., Prieta, F. de la, Prieto, J., and Corchado, J. M., “Blockchain framework for iot data quality via edge computing,” in *Proceedings of the 1st Workshop on Blockchain-enabled Networked Sensor Systems*, 2018, pp. 19–24.

[35] Guo, J., Wu, J., Liu, A., and Xiong, N., “Lightfed: An efficient and secure federated edge learning system on model splitting,” English, *IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS*, vol. 33, no. 11, pp. 2701–2713, Nov. 1, 2022, ISSN: 1045-9219. DOI: 10.1109/TPDS.2021.3127712.

[36] Latif, S. and al., W. et., “Ai-empowered, blockchain and sdn integrated security architecture for iot network of cyber physical systems,” English, *COMPUTER COMMUNICATIONS*, vol. 181, pp. 274–283, Jan. 1, 2022, ISSN: 0140-3664. DOI: 10.1016/j.comcom.2021.09.029.

[37] Wadhwa, S., Rani, S., Kavita, Verma, S., Shafi, J., and Wozniak, M., “Energy efficient consensus approach of blockchain for iot networks with edge computing,” English, *SENSORS*, vol. 22, no. 10, May 2022, ISSN: 1424-8220. DOI: 10.3390/s22103733.

[38] Hu, J., Reed, M., Thomos, N., AI-Naday, M., and Yang, K., “Securing sdn-controlled iot networks through edge blockchain,” English, *IEEE INTERNET OF THINGS JOURNAL*, vol. 8, no. 4, pp. 2102–2115, Feb. 15, 2021, ISSN: 2327-4662. DOI: 10.1109/JIOT.2020.3017354.

[39] Alrubei, S., Ball, E., and Rigelsford, J., “A secure distributed blockchain platform for use in ai-enabled iot applications,” in *2020 IEEE Cloud Summit*, IEEE, 2020, pp. 85–90.

[40] Gharbi, C., Hsairi, L., and Zagrouba, E., “A secure integrated fog cloud-iot architecture based on multi-agents system and blockchain.,” in *ICAART (2)*, 2021, pp. 1184–1191.

[41] Xu, R., Hang, L., Jin, W., and Kim, D., “Distributed secure edge computing architecture based on blockchain for real-time data integrity in iot environments,” English, *ACTUATORS*, vol. 10, no. 8, Aug. 2021, Times cited: 2, ISSN: 2076-0825. DOI: 10.3390/act10080197.

[42] Alsamhi, S. H., Curry, E., Hawbani, A., Kumar, S., Hassan, U. U., and Rajput, N. S., “Dataspace in the sky: A novel decentralized framework to secure drones data sharing in b5g for industry 4.0 toward industry 5.0,” *industrial and manufacturing engineering*, 2023.

[43] Guo, J., Cheng, Y., Wang, D., Tao, F., and Pickl, S., “Industrial dataspace for smart manufacturing: Connotation, key technologies, and framework,” *International Journal of Production Research*, pp. 1–16, 2021.

[44] Ren, L., Li, Y., Wang, X., Cui, J., and Zhang, L., “An abge-aided manufacturing knowledge graph construction approach for heterogeneous iiot data integration,” *International Journal of Production Research*, pp. 1–15, 2022.

[45] Hellmanzik, B. and Sandkuhl, K., “Towards a fair-ready data value chain for dataspaces,” in *Enterprise Engineering Working Conference*, Springer, 2022, pp. 90–105.

[46] Singh, P., Acharya, K. S., Beliatis, M. J., and Presser, M., “Semantic search system for real time occupancy,” in *2021 IEEE International Conference on Internet of Things and Intelligence Systems (IoTaIS)*, IEEE, 2021, pp. 49–55.

[47] Pampus, J., Jahnke, B.-F., and Quensel, R., “Evolving data space technologies: Lessons learned from an ids connector reference implementation,” in *Leveraging Applications of Formal Methods, Verification and Validation. Practice: 11th International Symposium, ISoLA 2022, Rhodes, Greece, October 22–30, 2022, Proceedings, Part IV*, Springer, 2022, pp. 366–381.

[48] Tran, T. T. T., “A systematic literature review on secure iot data sharing,” M.S. thesis, University of Oslo, 2022.

[49] Hedeler, C., Belhajjame, K., Fernandes, A. A., Embury, S. M., and Paton, N. W., “Dimensions of dataspaces,” in *Dataspace: The Final Frontier: 26th British National Conference on Databases, BNCOD 26, Birmingham, UK, July 7-9, 2009. Proceedings 26*, Springer, 2009, pp. 55–66.

[50] Sharma, A., Podoplelova, E., Shapovalov, G., Tselykh, A., and Tselykh, A., “Sustainable smart cities: Convergence of artificial intelligence and blockchain,” *Sustainability*, vol. 13, no. 23, p. 13 076, 2021.

[51] Farahani, B., Firouzi, F., and Luecking, M., “The convergence of iot and distributed ledger technologies (dlt): Opportunities, challenges, and solutions,” *Journal of Network and Computer Applications*, vol. 177, p. 102 936, 2021.

[52] Nguyen, D. C., Ding, M., Pham, Q.-V., Pathirana, P. N., Le, L. B., Seneviratne, A., Li, J., Niyato, D., and Poor, H. V., “Federated learning meets blockchain in edge computing: Opportunities and challenges,” *IEEE Internet of Things Journal*, vol. 8, no. 16, pp. 12 806–12 825, 2021.

[53] Hamdan, S., Ayyash, M., and Almajali, S., “Edge-computing architectures for internet of things applications: A survey,” *Sensors*, vol. 20, no. 22, p. 6441, 2020.

[54] Rahman, H. and Hussain, M. I., “A comprehensive survey on semantic interoperability for internet of things: State-of-the-art and research challenges,” *Transactions on Emerging Telecommunications Technologies*, vol. 31, no. 12, e3902, 2020.

[55] Chen, L. and Lu, Z. e. a., "A resource recommendation model for heterogeneous workloads in fog-based smart factory environment," English, *IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING*, vol. 19, no. 3, pp. 1731–1743, Jul. 2022, ISSN: 1545-5955. DOI: 10.1109/TASE.2022.3141590.

[56] Verbraeken, J., Vos, M. de, and Pouwelse, J., "Bristle: Decentralized federated learning in byzantine, non-iid environments," *arXiv preprint arXiv:2110.11006*, 2021.

[57] Zehir, S. and Zehir, M., "Internet of things in blockchain ecosystem from organizational and business management perspectives," *Digital Business Strategies in Blockchain Ecosystems: Transformational Design and Future of Global Business*, pp. 47–62, 2020.

[58] Bhat, S. A., Sofi, I. B., and Chi, C.-Y., "Edge computing and its convergence with blockchain in 5g and beyond: Security, challenges, and opportunities," *IEEE Access*, vol. 8, pp. 205 340–205 373, 2020, Times cited: 4 publisher: IEEE.

[59] Wu, Y., Dai, H.-N., and Wang, H., "Convergence of blockchain and edge computing for secure and scalable iiot critical infrastructures in industry 4.0," *IEEE Internet of Things Journal*, vol. 8, no. 4, pp. 2300–2317, Feb. 2021, Times cited: 2 event-title: IEEE Internet of Things Journal, ISSN: 2327-4662. DOI: 10.1109/JIOT.2020.3025916.

[60] Jagatheesaperumal, S. K., Pham, Q.-V., Ruby, R., Yang, Z., Xu, C., and Zhang, Z., "Explainable ai over the internet of things (iot): Overview, state-of-the-art and future directions," *IEEE Open Journal of the Communications Society*, 2022.

[61] Li, M., Pei, P., Yu, F. R., Si, P., Li, Y., Sun, E., and Zhang, Y., "Cloud–edge collaborative resource allocation for blockchain-enabled internet of things: A collective reinforcement learning approach," *IEEE Internet of Things Journal*, vol. 9, no. 22, pp. 23 115–23 129, Nov. 2022, Times cited: 1 event-title: IEEE Internet of Things Journal, ISSN: 2327-4662. DOI: 10.1109/JIOT.2022.3185289.

[62] Queralta, J. P., Qingqing, L., Zou, Z., and Westerlund, T., "Enhancing autonomy with blockchain and multi-access edge computing in distributed robotic systems," in *2020 Fifth International Conference on Fog and Mobile Edge Computing (FMEC)*, IEEE, 2020, pp. 180–187.

[63] Dawod, A., Georgakopoulos, D., Jayaraman, P. P., and Nirmalathas, A., "Advancements towards global iot device discovery and integration," in *2019 IEEE International Congress on Internet of Things (ICIOT)*, IEEE, 2019, pp. 147–155.

[64] Rondanini, C., Carminati, B., and Ferrari, E., "Confidential discovery of iot devices through blockchain," in *2019 IEEE International Congress on Internet of Things (ICIOT)*, IEEE, 2019, pp. 1–8.

[65] Seitz, A., Henze, D., Miehle, D., Bruegge, B., Nickles, J., and Sauer, M., "Fog computing as enabler for blockchain-based iiot app marketplaces-a case study," in *2018 Fifth international conference on internet of things: systems, management and security*, IEEE, 2018, pp. 182–188.

[66] Kang, J., Xiong, Z., Niyato, D., Xie, S., and Zhang, J., "Incentive mechanism for reliable federated learning: A joint optimization approach to combining reputation and contract theory," English, *IEEE INTERNET OF THINGS JOURNAL*, vol. 6, no. 6, pp. 10 700–10 714, Dec. 2019, ISSN: 2327-4662. DOI: 10.1109/JIOT.2019.2940820.

[67] Vaño, R., Lacalle, I., Sowiński, P., S-Julián, R., and Palau, C. E., "Cloud-native workload orchestration at the edge: A deployment review and future directions," *Sensors*, vol. 23, no. 4, p. 2215, 2023.

[68] Shahraki, A., Ohlenforst, T., and Kreyß, F., "When machine learning meets network management and orchestration in edge-based networking paradigms," *Journal of Network and Computer Applications*, vol. 212, p. 103 558, 2023.

[69] Karaagac, A., Dalipi, E., Crombez, P., De Poorter, E., and Hoebeke, J., "Light-weight streaming protocol for the internet of multimedia things: Voice streaming over nb-iot," *Pervasive and Mobile Computing*, vol. 59, p. 101 044, Jul. 1, 2019. DOI: 10.1016/j.pmcj.2019.101044.

[70] Williams, P., Dutta, I., Daoud, H., and Bayoumi, M., "A survey on security in internet of things with a focus on the impact of emerging technologies," English, *INTERNET OF THINGS*, vol. 19, Aug. 2022, ISSN: 2543-1536. DOI: 10.1016/j.iot.2022.100564.

[71] Fang, J. and Lei, K., "Blockchain for edge ai computing: A survey," en, *Yingyong Kexue Xuebao/Journal of Applied Sciences*, vol. 38, no. 1, pp. 1–21, 2020.

[72] Vakakis, N., Nikolis, O., Ioannidis, D., Votis, K., and Tzovaras, D., "Cybersecurity in smes: The smart-home/office use case," in *2019 IEEE 24th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, IEEE, 2019, pp. 1–7.

[73] Abdullahi, M., Baashar, Y., Alhussian, H., Alwadain, A. S., Aziz, N., Capretz, L. F., and Abdulkadir, S. J., "Detecting cybersecurity attacks in internet of things using artificial intelligence methods: A systematic literature review," *Electronics*, 2022.

[74] Afrin, M., Jin, J., Rahman, A., Rahman, A., Wan, J., and Hossain, E., "Resource allocation and service provisioning in multi-agent cloud robotics: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp. 842–870, 2021.

[75] Muhati, E., Rawat, D. B., and Sadler, B. M., "A new cyber-alliance of artificial intelligence, internet of things, blockchain, and edge computing," *IEEE Internet of Things Magazine*, vol. 5, no. 1, pp. 104–107, 2022, Times cited: 2 publisher: IEEE.

[76] Lamperti, S., Cavallo, A., and Sassanelli, C., "Digital servitization and business model innovation in smes: A model to escape from market disruption," *IEEE Transactions on Engineering Management*, pp. 1–15, 2023, Times cited: 3 event-title: IEEE Transactions on Engineering Management, ISSN: 1558-0040. DOI: 10.1109/TEM.2022.3233132.

[77] Alsamhi, S. H., Almalki, F. A., Afghah, F., Hawbani, A., Shvetsov, A. V., Lee, B., and Song, H., "Drones' edge intelligence over smart environments in b5g: Blockchain and federated learning synergy," *IEEE Transactions on Green Communications and Networking*, vol. 6, no. 1, pp. 295–312, 2021.

[78] Ming, G., "Exploration of the intelligent control system of autonomous vehicles based on edge computing," en, *PLOS ONE*, vol. 18, no. 2, e0281294, Feb. 2, 2023, Times cited: 1 publisher: Public Library of Science, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0281294.

[79] Singh, P., Holm, K., Beliatis, M. J., Ionita, A., Presser, M., Wolfgang, P., and Goduscheit, R. C., "Blockchain for economy of scale in wind industry: A demo case," in *Internet of Things: 5th The Global IoT Summit, GIoTS 2022, Dublin, Ireland, June 20–23, 2022, Revised Selected Papers*, Springer, 2023, pp. 175–186.

[80] Angles, R., Thakkar, H., and Tomaszuk, D., "Rdf and property graphs interoperability: Status and issues.," *AMW*, vol. 2369, 2019.

[81] Aminu, E. F., Oyefolahan, I. O., Abdullahi, M. B., and Salaudeen, M. T., "A review on ontology development methodologies for developing ontological knowledge representation systems for various domains," *I.J. Information Engineering and Electronic Business*, 2020.

[82] Soleimani Nejad, A. and Soleimani, A., "A systematic review of research in the field of simple knowledge organization (skos) between 2010 and 2020," *Journal of Studies in Library and Information Science*, vol. 13, no. 1, 2021.

[83] Kiselev, B. and Yakutenko, V., "An overview of massive open online course platforms: Personalization and semantic web technologies and standards," *Procedia Computer Science*, vol. 169, pp. 373–379, 2020.

[84] Aburukba, R. O., Landolsi, T., and Omer, D., "A heuristic scheduling approach for fog-cloud computing environment with stationary iot devices," en, *Journal of Network and Computer Applications*, vol. 180, p. 102 994, Apr. 2021, ISSN: 10848045. DOI: 10.1016/j.jnca.2021.102994.

[85] Almalki, F. A., Mutawi, A. M., Turkistani, I. A., Alqurashi, L. K., Fattah, M. T., Almogher, M. T., Aldaher, R. S., Wali, R. A., Almalki, W. M., and Almubayed, Y. M., "Developing CubeSat and AI Framework for Crowd Management Case of Short-Term Large-Scale Events," en, *Advances in Science, Technology and Engineering Systems Journal*, vol. 7, no. 6, pp. 114–125, Nov. 2022, ISSN: 24156698, 24156698. DOI: 10.25046/aj070612.

[86] Beckett, D., Berners-Lee, T., Prud'hommeaux, E., and Carothers, G., "Rdf 1.1 turtle," *World Wide Web Consortium*, pp. 18–31, 2014.

[87] Haghgoo, M., Sychev, I., Monti, A., and Fitzek, F. H., "Entirety — semanntic provisioning and governing iot devices in smart energy domain," en, *SoftwareX*, vol. 18, p. 101 081, Jun. 2022, ISSN: 23527110. DOI: 10.1016/j.softx.2022.101081.

[88] Guindon, C., *The Sparkplug Specification | The Eclipse Foundation*, en, 2023.

[89] Kristiani, E., Yang, C.-T., Wang, Y. T., and Huang, C.-Y., "Implementation of an Edge Computing Architecture Using OpenStack and Kubernetes," en, in *Information Science and Applications 2018*, Kim, K. J. and Baek, N., Eds., ser. Lecture Notes in Electrical Engineering, Singapore: Springer, 2019, pp. 675–685, ISBN: 9789811310560. DOI: $10.1007/978-981-13-1056-0_66$.

[90] Ahmed, I., Zhang, Y., Jeon, G., Lin, W., Khosravi, M. R., and Qi, L., "A blockchain- and artificial intelligence-enabled smart IoT framework for sustainable city," *International Journal of Intelligent Systems*, 2022, ISSN: 1098111X. DOI: 10.1002/int.22852.

[91] Consortium, O. *et al.*, "Ieee standard for adoption of openfog reference architecture for fog computing," *IEEE Std*, vol. 2018, no. 2018, pp. 1–176, 2018.

[92] David, R., Duke, J., Jain, A., Janapa Reddi, V., Jeffries, N., Li, J., Kreeger, N., Nappier, I., Natraj, M., Wang, T., *et al.*, "Tensorflow lite micro: Embedded machine learning for tinyml systems," *Proceedings of Machine Learning and Systems*, vol. 3, pp. 800–811, 2021.

[93] Munoz-Arcentales, A., López-Pernas, S., Conde, J., Alonso, Á., Salvachúa, J., and Hierro, J. J., "Enabling context-aware data analytics in smart environments: An open source reference implementation," *Sensors*, vol. 21, no. 21, p. 7095, 2021.

[94] Gupta, P. *et al.*, "A survey of application layer protocols for internet of things," in *2021 International Conference on Communication information and Computing Technology (ICCICT)*, IEEE, 2021, pp. 1–6.

[95] Hartner, F., Löwen, U., and Franke, J., "Digital industrial b2b platform patterns from a business perspective," *ESSN: 2701-6277*, 2021.

[96] Lee, C., Huo, Y., Zhang, S., and Ng, K., "Design of a smart manufacturing system with the application of multi-access edge computing and blockchain technology," English, *IEEE ACCESS*, vol. 8, pp. 28 659–28 667, 2020, Times cited: 2, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2020.2972284.

[97] Wang, K., Jin, J., Yang, Y., Zhang, T., Nallanathan, A., Tellambura, C., and Jabbari, B., "Task offloading with multi-tier computing resources in next generation wireless networks," *IEEE Journal on Selected Areas in Communications*, 2022.

[98] Ekwe-Ekwe, N. N., “All weather scheduling: Towards effective scheduling across the edge, fog and cloud,” Ph.D. dissertation, The University of St Andrews, 2023.

[99] Sun, X., Sung, Y.-W., Krothapalli, S. D., and Rao, S. G., “A systematic approach for evolving vlan designs,” in *2010 Proceedings IEEE INFOCOM*, IEEE, 2010, pp. 1–9.

[100] Androulaki, E., Barger, A., Bortnikov, V., Cachin, C., Christidis, K., De Caro, A., Enyeart, D., Ferris, C., Laventman, G., Manevich, Y., *et al.*, “Hyperledger fabric: A distributed operating system for permissioned blockchains,” in *Proceedings of the thirteenth EuroSys conference*, 2018, pp. 1–15.

[101] Srivastava, A. and Desai, Y., “Performance analysis of hyperledger fabric based blockchain for traceability in food supply chain,” in *2021 IEEE 2nd International Conference on Technology, Engineering, Management for Societal impact using Marketing, Entrepreneurship and Talent (TEMSMET)*, IEEE, 2021, pp. 1–5.

[102] Sagirlar, G., Carminati, B., Ferrari, E., Sheehan, J. D., and Ragnoli, E., “Hybrid-iot: Hybrid blockchain architecture for internet of things-pow sub-blockchains,” in *2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, IEEE, 2018, pp. 1007–1016.

[103] Name, A., *Unwind - blockchain i vindmolleindustrien*, https://industriensfond.dk/projekt/unwind-blockchain-i-vindmolleindustrien/, Project Report, Accessed: November 14, 2023, 2022.

[104] Gadekallu, T. R., Pham, Q.-V., Nguyen, D. C., Maddikunta, P. K. R., Deepa, N., Prabadevi, B., Pathirana, P. N., Zhao, J., and Hwang, W.-J., “Blockchain for edge of things: Applications, opportunities, and challenges,” *IEEE Internet of Things Journal*, vol. 9, no. 2, pp. 964–988, 2021.

Chapter 7

Paper V - Data-Driven IoT Ecosystem for Cross Business Growth: An Inspiration Future Internet Model with Dataspace at the Edge

Parwinder Singh, Nidhi, Michail J. Beliatis, Mirko Presser

Originally published in *INTERNET 2024:The Sixteenth International Conference on Evolving Internet.*, March 2024, ISSN: 2308-443X, pp. 19-25. DOI:978-1-68558-133-6.

V

Abstract

Data is the bloodline for a business to grow, compete, and sustain in the market. It empowers businesses to build diverse services comprising innovative business models. For this, businesses must adopt an open collaboration approach, making their data and associated services available for sharing and reuse purposes, leading towards a positive and collaborative win-win business model instead of competing with each other. This creates the need for a digital ecosystem that allows data and services to be shared, reused, and exchanged in a governed and secure manner. Dataspace (DS) caters to the same objective that facilitates many data operations for stakeholders, such as search, query, aggregation, federation, integration, analysis, etc., over geo-spatially distributed and diverse resources. Therefore, we propose a novel edge-enabled context-aware Dataspace model, presented for the first time in literature, as a potential solution to integrate cross-domain and cross-organization data and associated services in local or regional contexts. This model aligns with the architectural vision of the future internet model, which can create collaborative innovation and shape the futuristic industry 5.0 and beyond ecosystems. In this context, each participating organization will act as an edge that supports DS computing resource requirements and offers edge-oriented advantages in saving latency, bandwidth, and data operations near or at the data source. The model has also been validated over a local IoT edge-cluster emulated Dataspace testbed and found to fulfill the functional aspects of the proposed model.

7.1 Introduction

Data, in the Internet of Things (IoT) ecosystem, is an asset to active (primary) and passive (secondary) users, i.e., generated data for specific purposes can be useful for other applications based on data sharing and exploitation rights in its raw or processed form. Dataspace (DS) has emerged as a paradigm to facilitate seamless data integration from various heterogeneous data sources, including corporate databases, files, web services, IoT-oriented devices, platforms, gateways, services, etc. It administers a virtual space to pool data from various sources under its owner rights until requested access from another application or service [1].

DS expedites cross-domain data management operations and creates a unified data catalog, acting as a regulated data marketplace adhering to relevant policies for fair data usage [2]. It enables a user-friendly semantic representation of data context with built-in security and privacy measures, leading to numerous opportunities and innovative business models for different stakeholders engaged in the data life cycle and connected over the Dataspace value chain network [3]. For example, DS can enable the Pay-as-You-Go business model [4] and

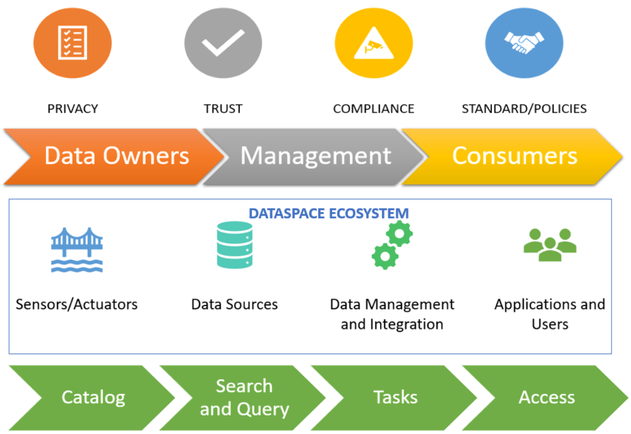


Figure 7.1: Dataspace Ecosystem and Associated Players

generate revenue from available data through its pooling, sharing, reusability, and access capabilities [5]. Figure 7.1 illustrates the interaction between different stakeholders in the DS ecosystem. However, data integration for DS faces many challenges in developing cross-domain data and service value chains. Therefore, this raises an important Research Question (RQ):

How to build DS in the local context for developing data-driven cross-domain service value-chain enablement?

The surge in connected IoT devices demands a resilient and robust future internet infrastructure to facilitate efficient data management and associated operations [6]. Therefore, the role of distributed edge computing becomes more significant in supporting edge-enabled DS ecosystem [7] to address and optimize the arising challenges of security, privacy, standardized integration practices, and transforming the digital landscape towards sustainability [8]. The concept of DS revolutionizes the way we perceive and utilize data across the entire value chain, facilitating diverse services enablement and monetization opportunities that drive growth and create lasting impact. Aligning with the RQ, we have broadened the understanding of the DS concept with a focus on *how such an ecosystem can be realized at the edge* or on-premises environment, contrary to a centralized cloud facility to avail optimized latency, bandwidth, and data operations.

DS at the edge can allow data and associated services to be shared, reused, exchanged, and integrated across domains in local or regional contexts. However, realizing such a cross-domain integration ecosystem is often bundled with challenges like linked computing resources and data pool, heterogeneity, dynamic deployment context, interoperability, trust, governance, participatory motivation, etc. [3]. Therefore, it becomes critical to enable the semantic capabilities of the data to build a context-aware edge-enabled DS model. Data context awareness enhances understanding, aiding discovery, quality assurance, and integration. It establishes a semantic layer for linked data within the DS ecosystem, ensuring higher data quality and reliability [7], [9]. The context-aware linked DS can enable semantic integration and harmonize the relationships within data, unlocking new insights and possibilities [3], [7]. Additionally, it will bring synergy with constantly evolving user requirements by facilitating data and technology convergence [10]. In the context of cross-domain edge (representing organization, domain, system, or service) integration empowers DS with required computing resources, availability, and convergence of technologies that enable diverse stakeholders to build a unified ecosystem for innovative business models and dynamic data-driven applications [3], [7], [11].

Therefore, this study has contributed to the semantics enablement and smart governance of the data management and associated services in future internet hyper-connected applications, particularly considering 6G and beyond [12] network ecosystems. This is achieved by identifying relevant stakeholders' common requirements, proposing and designing a Dataspace model with context-aware data processing, smart governance, and semantic adaption capabilities. In addition, a novel service artifact methodology, consisting of a service catalog and relevant toolchain, is also introduced to realize such a DS model efficiently over a distributed edge network.

The rest of the paper is given as follows: Section II will summarise relevant literature on DS and highlight key takeaways, and Section III will explain the overall methodology of this study. Further, Section IV will provide the system model, and deployment architecture framework to realize the proposed DS platform. Finally, Section V will conclude the paper.

7.2 Literature Review Outcomes

This section summarises the relevant literature on DS and related enabling techniques and technologies. The DS ecosystem offers a promising solution by breaking down data silos and promoting cross-domain data sharing with contextual semantics [1]. Initiatives like the International Data Space (IDS) and GAIA-X in the EU have outlined architectural frameworks and guidelines to strengthen the data economy by developing DS ecosystems [13] to facilitate seamless data integration in a larger context. StreamPipes Connect, a distributed edge-driven semantic adaptation toolbox, allows harmonizing data in Industrial IoT analytics by enabling data ingestion, sharing, and data model automation [14]. In realizing DS, addressing heterogeneity [15] is critical and can be resolved by leveraging semantics wherein ontologies represent machine-readable conceptualization of knowledge understanding at the domain level, and metadata represents a data structure at the business and technical level [5]. Thus, it is evident that metadata and ontology are essential for developing semantic information by mapping the business-level domain information to relevant technical-level information, consisting of data encapsulated entities, objects, and their inter-relationships that represent associated operations.

To build a DS ecosystem, multiple participants or entities are required. Here each entity consists of data sources and associated services with a specific or cross-domain that are geo-spatially distributed [3] and supports diverse data types or formats to represent the relevant domain-level information [3], [5]. DS essentially provides data co-existence, sharing, and reusability while promoting pay-as-you-go methods or services over the integrated data [5]. DS, in general, does not control or own the data sources, thereby, the data maintenance and administration falls under the individuals or their relevant organizational management systems [16]. Therefore, the European GAIA-X project [17] has focused on a cross-ecosystem data exchange with data sovereignty based on linking data principles. It facilitates the "common data space" concept for implementing a future "space data economy" in a cooperative business space through a common GAIA-X standard [18] supporting interoperability, portability, and data sovereignty as guided in the European data strategy [19]. Semantic modeling development tools such as Plasma are really helpful for non-technical users in providing a visual editing interface to build semantic models for DS operations[20]. These tools allow the creation, extension, and export of the semantic models and related ontologies along with relevant maintenance of knowledge graphs to annotate the datasets with semantic descriptions and convert them into unified and Resource Description Framework (RDF) standard format [21]. In IoT landscape, an edge-driven DS incorporating 'virtual sensors' allows for abstracting and mapping high-level user-driven application behaviors [22]. The user actions (in the form of HTTP verbs PUT/GET/POST, etc.) are to be reflected at the edge device, which is linked to the virtual sensor, through the application and leveraging Next Generation Service Interface - Linked Data (NGSI-LD) semantic standards information model [23].

There are also some DS-related architectural studies found in the literature.

For example, [24] presents a DS testbed for maritime domain-driven data management operations which is based on a Service-Oriented Architecture (SOA) and layer-based structure emphasizing data protection and sovereignty to cater to diverse needs and support activities among multiple stakeholders. This model, however, does not address heterogeneity among various data sources. Similarly, [7] presents a Dataspace integration enablement framework based on the convergence of technologies and extending the (Cloud-Edge-Device) CED model with semantics capabilities that offer dynamic data, processing, and service context. This study has been used as the basis to define our current proposed model with a focus on context-aware DS development at the application and data management level.

Subject to limited literature about building edge-enabled DS platforms in the local context, this study contributes at the design level by proposing a distributed edge-enabled DS model with context-aware linked data and semantic adaptation capabilities.

7.3 Methodology

To address the RQ, we have identified the requirements based on DS stakeholders analysis [25], established methods for utilizing shared services [26], data reusability, embedding semantics in data, and creating values through context-aware linked data [27] within our local context at the Department of Business Development and Technology, Aarhus University. Our stakeholders include students, teachers, researchers, and industrial partners, where we find that *data and related operations* are the common entities among different projects. Therefore, we set a vision to extract useful information from the data semantically and collaboratively while the actors still have sovereign control over data with a readily available toolchain to perform certain semantic operations over the fusion of data in a context-aware and cross-domain manner. In this context, Figure 7.2 shows the value chain interaction (color-coded lines) among different stakeholders

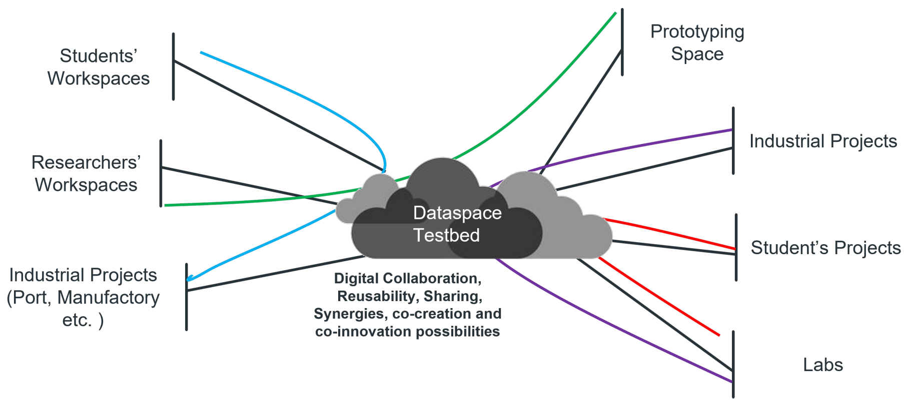


Figure 7.2: Local Context Dataspace - Stakeholders and Value Chains.

for cross-project (representing cross-domain) data-driven events and operations. This emerges as a requirement to develop a DS ecosystem in the local context to cater to diverse data management requirements. The functionality for identified requirements has been fulfilled by building a context-aware DS solution (i.e. testbed) following the proposed system model based on Onion architectural [28] methodology, deployment architecture [29], and selected use of toolchain as per target use case defined by the A-La-Carte (ALC) approach [30]. The solution is further validated for functional compliance against a cross-domain wind turbine supply-chain use case.

The main objective of this local context-driven DS platform is to empower hyper-connected applications and use cases in future internet-based distributed edge computing models where multiple stakeholders (dealing with different use cases, e.g., cross-lab collaboration activities, prototyping and training initiatives, external industrial projects, student education, etc.) and their data interactions will develop relevant value chains in their contextual space, as shown in Figure 7.2. Therefore, we proposed a semantics-driven DS model with context-aware data lake functional capabilities and realized it in our local lab environment. The next section covers the relevant details.

7.4 System Model and Framework

This section proposes a reference semantic DS model implemented with context-aware and semantic adaptation capabilities to ensure that the context associated with the data under diverse DS operations enables data value in a given context and empowers data usefulness.

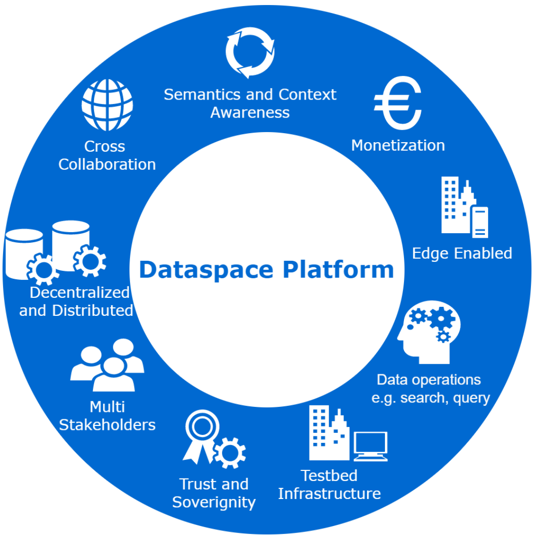


Figure 7.3: High-Level Requirements for Edge Enabled Dataspace.

7.4.1 Requirements Analysis

Figure 7.3 illustrates high-level requirements to realize the DS ecosystem based on our stakeholder discussions, which are explained as follows:

- *Multistakeholder and Cross-Collaboration* - This indicates that the DS should support multi-tenancy operations across domains to promote collaboration while securing ownership, isolation and segregation aspects. This will enable the development of cross-domain service value chains over the data integrated in DS.
- *Monetization* - One of the main objectives for DS development is to generate monetary values from the DS integrated ecosystem by building innovative business models based on each other's data strengths. This can be the basis for a data marketplace where data and associated services generate real value and motivation.
- *Data Operations* - The system should allow data management i.e., CRUD (creation, updation, deletion, and read), operations along with federation, analytical, and visualization contextually.
- *Decentralization* - The DS platform should be decentralized and distributed regarding its resources, i.e., computing, storage, and networking for data management. This makes it scalable and near to real-time prototyping in nature. In addition, this platform will be geo-spatially distributed to extend its functionality to target use cases, where this platform serves as a toolchain for data management operations.
- *Semantically Context Awareness* - The DS platform is perceived to be context-aware based on semantics-driven data linkage. This is important to generate knowledge graphs and cross-domain linked information required to build data-driven value chains among stakeholders.
- *Trust and Sovereignty* - This is an important feature in any DS platform that ensures the stakeholder who owns data shall have complete control over their data. This is also needed for General Data Protection Regulation (GDPR) compliance within the EU.
- *Edge-enabled Infrastructure* - DS platform shall be able to realize on-premises near the data sources and with all required relative toolchains available to cater to specific needs for the target use case and related stakeholders. Anyway, in the DS context, the data mostly lies with the generator, and it only expects the data to be searched, indexed, and accumulated on a temporary need basis. Hence, it eliminates the need for expensive cloud-enabled recurring costs and centralized facilities. Therefore, such platforms can be realized with relatively smaller costs.

7.4.2 Context Aware Dataspace Model

The system model for our context-aware DS is shown in Figure 7.4. It is based on the identified requirements and our previous work on the Distributed Edge Network Operations oriented Semantic (i.e. DENOS) model, presented in [7]. It is motivated by the "Onion Architecture" design [28], wherein the key idea is to map the dependencies of the outer layers towards the inner layers and the core, providing a clear separation and segregation of concerns, thus simultaneously improving functional and non-functional concerns. Our proposed architecture has five layers and one main core, explained below from outer to inner direction.

- **Data Source Layer:** This layer represents the source of data that needs to be searched, indexed, queried, etc., by different applications for specific purposes or needs. It stores and manages data to a specific domain or organization and has specific metadata or structure. Different data sources represent different metadata or data models, though they may be semantically identical. Thus, it induces the challenge of heterogeneity and interoperability during the data integration operations.

- **Semantic Adaptation Layer:** To harmonize heterogeneity, this layer provides tools and methods to annotate the incoming data (from the Data Source layer) semantically as per ontology and metadata models. This layer also provides tools to define/reuse relevant ontology and metadata models. Semantic modeling standards like NGSI-LD, RDF, Web Ontology Language (OWL), JSON for Linking Data (JSON-LD), etc. can be used here.

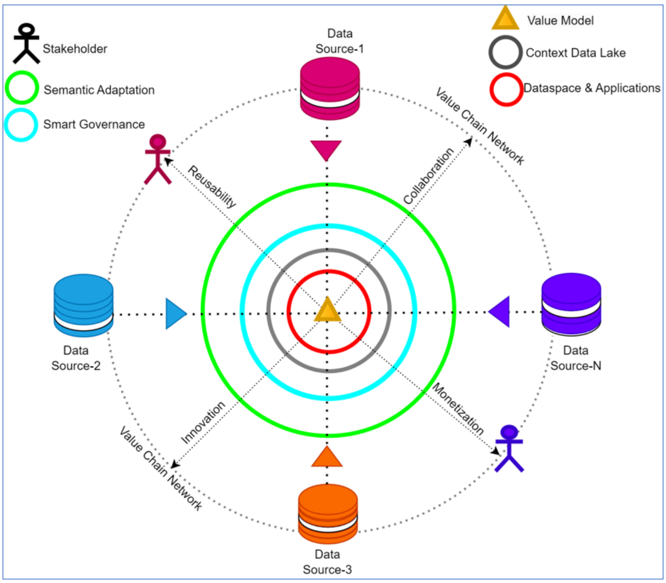


Figure 7.4: Context-Aware Dataspace Model

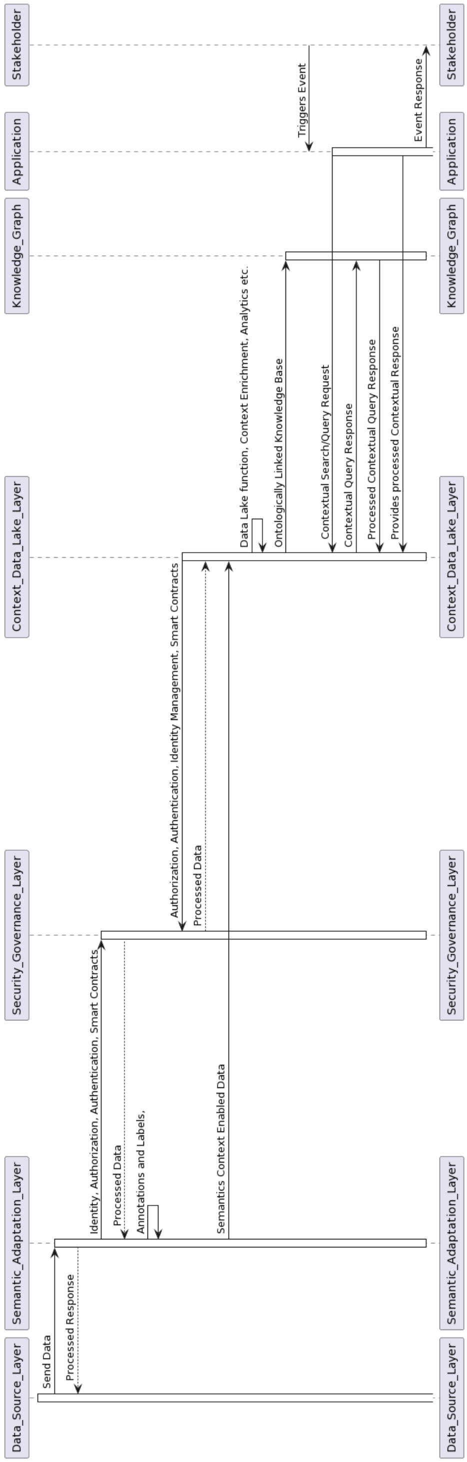


Figure 7.5: Sequence Diagram for the Context-Aware Dataspace Operations

- **Smart Governance Layer:** This layer provides mechanisms to offer identity and access management to maintain trust and sovereignty of the data being operated. This can be achieved using Identity and Role and Attribute access management in a traditional way leveraging standards such as Security Assertion Markup Language (SAML), OpenID Connect (OIDC), OAuth 2.0, System for Cross-domain Identity Management (SCIM), etc., implemented or integrated through DLT/Blockchain-driven smart contracts to have fine-grained granular control[31]. It ensures identity, role, and attribute-based access in a decentralized, transparent, and tamper-resistant manner.
- **Context Data Lake Layer:** This layer represents a specialized data lake offering temporary storage and contextual data management using relevant toolchains. Here, contextual data includes semantically annotated data presenting information at the ontology, domain, and metadata level, providing additional context like metadata, lineage, quality, relationships, origin, etc., for the data to be linked with other domain-level information in different contexts to help machines understand and interpret the data as per the contexts. Further, it facilitates data governance, tracking, discovery, and cataloging efforts, enabling stakeholders to find and utilize the right data for their analytical or operational needs.
- **Application Layer:** This layer provides the DS operations enablement, as per the target use case-driven value context (extraction) needs, over the contextual data in different contexts offered by the contextual data lake.
- **Value Model:** This is the framework's core that triggers different events, such as *Collaboration* for data *Reusability* to *Innovate* new values that can be *Monetized* through building of a *Value Chain Network* among collaborating *Stakeholders* who inspires to derive value out of their *Data Sources*. This drives the value extraction out of the diverse data sources for the given business value context of the use case, leveraging all the upper layers. The business value context can be defined using the relevant business value model, such as St. Gallens Magic Triangle [32] for the given use case.

Functional Flow - Figure 7.5 illustrates the sequence diagram for the context data lake-centered DS operations. Data comes from the Data Source Layer and enters the Semantic Adaptation Layer, where context annotations and labeling occur using semantic models defined by the domain's ontology. Moreover, before performing adaptation, it requests authorization, authentication, and identity management from the Security Governance Layer based on agreed-upon smart contract-driven policies. Then, the data is ingested inside the Context Data Lake Layer, which holds the data in relevant semantic service context [7] after the Data Lake's pre-configured pipeline operations, such as data/context enrichment, storage, analytics, etc. Thus, the Context Data Lake Layer holds data from multiple sources with multiple semantic contexts and builds a converged knowledge graph for the entire DS model.

7.4.3 Deployment Architecture

Multiple reasons motivated us to build DS at the edge. First, the DS is perceived to utilize edge network infrastructure in a coordinated manner, as shown in Figure 7.6, wherein each edge acts as the organizational entity holding the data with the ownership and providing the relevant semantics context and infrastructure for processing the data at the edge. This offers many advantages, such as the availability of infrastructure by resource pooling across edge networks, which will be a cost-efficient method and allow control of data processing at the edge, thereby raising trust and participatory stake in a multi-stakeholder DS environment. In this context, we intend to emphasize that all participating stakeholders interested in building the DS for mutual benefits can provide the necessary edge network infrastructure required to deploy the proposed DS model. Anyway, saving and optimum utilization of resources at the edge is always the

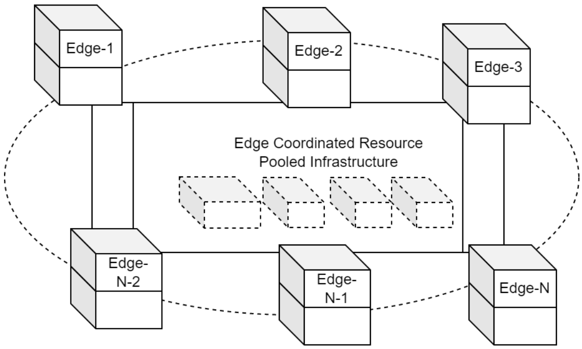


Figure 7.6: Edge Coordinated Resource Pooling

objective of edge computing and the future internet paradigm. Therefore, we have extended an ALC approach [30] to be used in the DS implementation context. ALC provides the flexibility to choose and pick different services from the service catalog and relevant open-source tools, as shown in Figure 7.7, to develop the pre-configured processing pipeline artifacts to implement the DS layer operations. This way, it helps to choose, select, and deploy only the required services to certain stakeholder or use-case contexts. Thus, saves a lot of computing resources, energy, and cost while addressing the challenges of heterogeneity, integration, and interoperability along with pre-defined processing pipelines and resource requirements. Under the ALC approach, the user selects the packages from the service catalog and generates the relevant artifacts, which can be deployed easily over the edge infrastructure in a distributed manner.

The deployment architecture for the DS platform/testbed is shown in Figure 7.8. The testbed is developed utilizing on-premises infrastructure and is incrementally scalable. The testbed's infrastructure, system, services, or applications can be scaled without disturbing the existing setup to accommodate the elasticity in the computing and processing demands.

The testbed's infrastructure is provisioned by Kubernetes which is a

Dataspace Testbed

Layer	Service Catalogue	Realization Tools	Aritifacts for Targeted Use Case: Industrial Projects	R&D Projects	Training & Education
Application	Search & Query Interfaces	SQL/REST/GraphQL	X	X	
	Data Visualization	Grafana		X	X
	Resources Management	Horizon (Openstack)		X	
	Stream Visualization	StreamPipes Connect	X	X	X
Platform	Analytics tool chain	Anaconda/Jupyter	X		X
	Programming modules	Node-Red/Python/NodeJS		X	X
	Semantic Broker	Scorpio NGSI-LD		X	
	Simple Broker	Mosquitto-MQTT Broker	X		X
	Stream Processor	StreamPipes Connect	X		X
	Onotology Modeller	Onotology Modeller		X	X
Virtualization	Virtual Storage	Cinder/Ceph Volumes	X	X	
	Virtual Networks	OpenVswitch/Neutron	X	X	X
	Virtual Routers and Ports	Neturon (Openstack)	X	X	
	Virtual Machines	Libvirt	X	X	X
	Virtual Containers	Docker	X	X	X
Hardware	Connectivity Media	RJ-45, Wifi Adapters	X	X	X
	Router and Switch	Any router and Switch	X	X	X
	Raspberry Pis	Raspberry Pis 4.0		X	X
	Servers/PCs	16GB, 8 cores, 200G HDD, 2 NICs, 1 wifi Adapter	X	X	X
	Internet Connectivity	5G/LTE/Wifi/WLan	X	X	X
Gateway Protocols	MQTT/AMQP	Mosquitto	X		X
	NGSI-LD	Scorpio		X	X
	Modbus/Bluetooth/LoraWan	Modbus/Bluetooth/LoraWan	X		X
	OPC-UA	OPC-UA			X
	HTTP REST	HTTP REST			X
Data Sources	Platforms	AWS, Azure, Private Cloud		X	
	Systems	Data Management Systems	X		X
	Open Data Services/APIs	CKAN, Orion		X	X
	Devices	IoT sensors, Actuators, UAVs	X	X	X
	Gateways	IoT or industrial gateways	X	X	X

Figure 7.7: A-La-Carte Approach for Dataspace Model Implementation

distributed microservices orchestrator [33]. We have used K3s which is a lightweight distribution of Kubernetes [34]. It supports Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), and Software-as-a-Service (Saas) models, catering to the diverse needs of stakeholder's use-case in the DS ecosystem. The testbed leveraged Infrastructure-as-a-Code, based on Ansible [33] for bootstrapping of infrastructure. Following this, the PaaS and SaaS are provisioned using the ALC approach, incorporating relevant toolchains like helm charts or Kubernetes templates [33].

The testbed's Platform Layer contains the core implementation of the DS model, encompassing the context data lake functions like data ingestion, authentication, storage, metadata management, and cataloging. The architecture is organized into distinct operational namespaces for resource isolation like (i) *Admin namespace* to manage the infrastructure and resource provisioning, (ii) *Stakeholder-specific namespace* to emulate cross-domain organizational projects for DS with limited access based on predefined roles, along with virtual resource allocation tailored to project needs, (iii) *Common-services namespace* to host shared services like broker, database, NodeRed, and Jupyter, accessible via agreed-upon APIs and permission. The DS testbed is provisioned with various artifacts utilizing Kubernetes/helm-based templates tailored to ALC package selection. These artifacts empower a broad spectrum of services for semantic adaptation and context-aware data lake processing. This encompasses Integration-as-a-Service (e.g., IDS connector) for semantic context-aware data operations, AI-as-a-Service (e.g., StyleGAN) with GPU-enabled edge-instances for machine learning, Database-as-a-Service (e.g., Postgres, and MySQL) for managing diverse types of data, and Programming-as-a-Service (e.g., Node-Red, and WordPress) for custom data processing flow development.

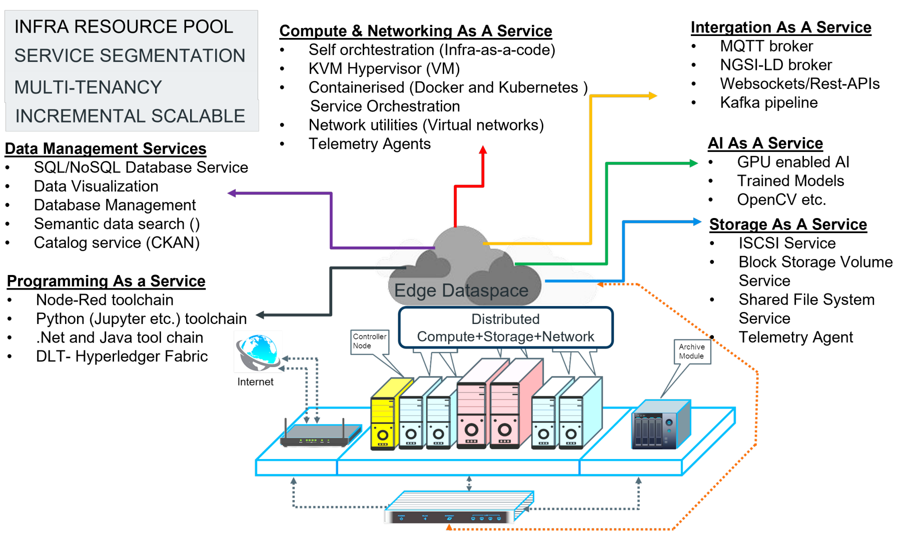


Figure 7.8: Deployment Architecture for Dataspace Model

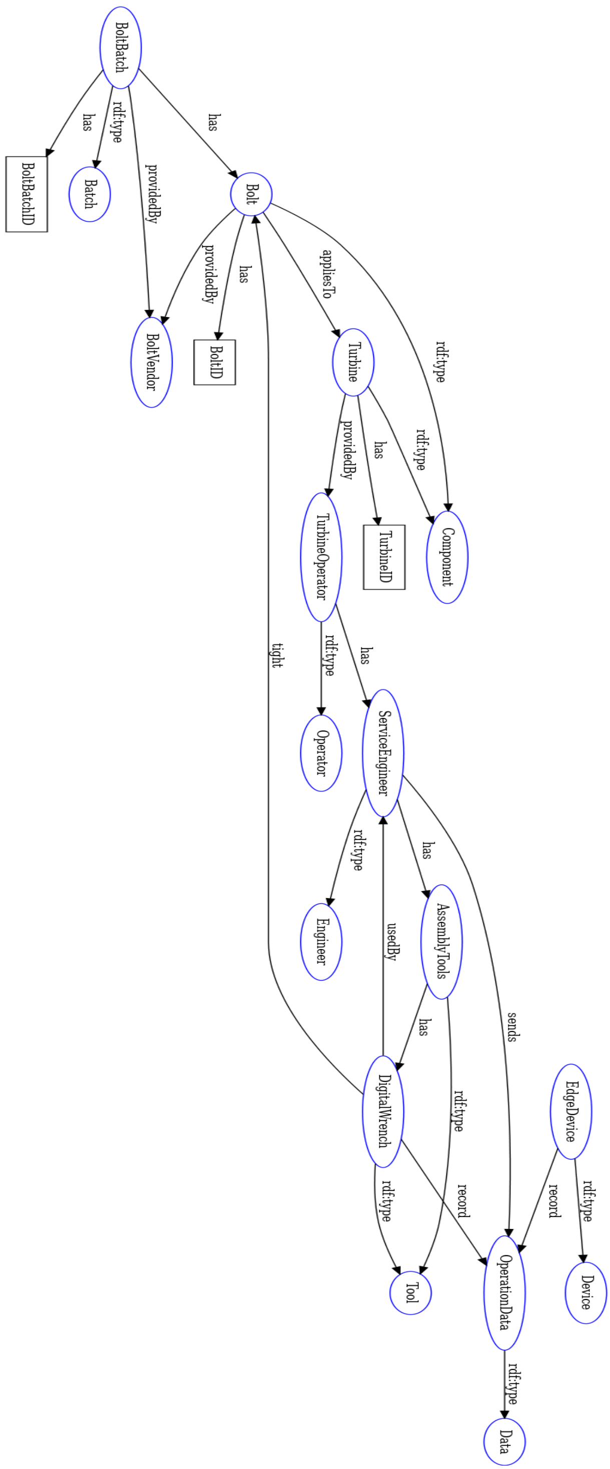


Figure 7.9: Digitized Wind Turbine Bolt Semantic Operations Context Model

7.4.4 Validation

The proposed architecture is validated against the wind turbine use case, presented already in [31]. However, the operations of this use-case have been represented semantically, for the first time, in Figure 7.9 using RDF standard. This bolt-specific operations semantic model serves as the basis and shows the path to define harmonized cross-domain data models among diverse stakeholders collaborating in wind turbine supply chains in the energy sector. This use case demonstrates the cross-domain digital traceability requirement for bolt, turbine, and related stakeholders that need to deal with diverse events being managed through our local Dataspace testbed. This use-case has been expressed semantically as - *A Service engineer with Name/Employee-ID (Domain-1) performing bolt, coming with Batch-No./ID (Domain-2) coming from supplier with ID, tightening operation at the turbine of turbine operator/manufacturer with turbine ID (Domain-3) at a certain location and time with timestamp.* So, the use-case deals with data from three different domains namely service engineer, turbine operator, and bolt supplier. The complete functional flow consists of nearby edge to the installed turbine capturing the relevant events (e.g., Service engineer registering for the device at the edge, Bolt batch registration by the turbine manufacturer, turbine/bolt identification via QR code scanning, bolt-supplier mapping registration, etc.) over radio interface (e.g. Bluetooth in our case) in the turbine assembly area or on the field. At the nearby edge, the semantic adaptation (static) functionality is provisioned using the ALC service artifact approach. In this case, the node-red based flow service artifact is provisioned on the edge. This adaptation service receives data over a Bluetooth radio interface on one hand, and it converges data from different events to create a semantically linked message using the NGSI-LD standard, on the other hand. Afterward, the transformed semantic data is pushed into the permissioned and private Blockchain, implemented using the HyperledgerFabric service artifact, and running at the neighboring edge. This provides us with the smart governance layer based on smart contracts-driven policies validating the pre-registered data model in our case. This can also be used to validate identification, authorization, and authentication through relevant smart contracts in combination with traditional security methods such as identity management or OAUTH2 standards. Finally, the data is processed further for context data lake layer functionality (e.g., StreamPipes, NGSI-LD broker) that allows the building of a knowledge base (based on semantically adapted contexts) and semantic CRUD operations (e.g., SPARQL/NGSI-LD) over data. Different stakeholders can now read this data over semantic contextual interfaces based on their role and permission level agreed upon in smart contracts. To validate this, various cross-domain semantic queries were executed by the stakeholder application, such as - *Fetch bolts from a specific batch ID that impact certain turbines to predict their maintenance requirements or inspection of operational events (e.g., torque value recorded during bolt tightening) for insurance claims under unseen events.*

The average response time results for different operations and their

Table 7.1: Data Operation and Their Response Times

Operation Type	**Response Time (ms)**	**Functional Context and Dataspace Model Relevance**
- Registration of Device- turbine or Bolt	800	Stakeholder application registers for turbine or Bolt attributes. - Application, Smart Governance, and Context Data Lake layers are involved.
Bolt or turbine ID Validation	1200	Service engineer scans the QR code for Turbine and Bolt ID and the relevant event at the edge creates a query to fetch Dataspace from the registered knowledge base. - Data source, Semantic Adaptation, Smart Governance, and Context Data lake layers are involved.
Torque Recording	500	Digital wrench is used to tight the bolt, and the relevant torque value is recorded by the nearby edge over Bluetooth and this is then recorded in Blockchain and Application backend both. - Data source, Semantic Adaptation, Smart Governance, and Context Data lake and application layers are involved.
Read Turbine, Bolt, or Log entry	600	Application interface reading the Dataspace backend for relevant event data. - Application, Smart Governance, and Context Data Lake are involved.

explanation are given in Table 7.1. In addition, this demonstrates the functional validation of the proposed DS model in local and cross-domain contexts. This shows the possibility of a collaborative data-driven value chain development among multiple stakeholders through the proposed model.

The artifacts for this use case consist of frontend (Node-Red, Bluetooth libraries, Web3.js) and backend (REST API, Blockchain Ganache/Hyperledger Fabric) components packaged and provisioned using the ALC approach and Kubernetes orchestrated distributed infrastructure, respectively. The frontend and backend components deployed in different namespaces (representing stakeholder system) over the edge (using two x86 servers-8 core, 16 GB RAM, 80 GB HDD) enabled-DS testbed.

7.5 Conclusions and Outlook

This paper has introduced the motivation for developing a DS platform at Edge and its realization being presented for the first time in literature, along with the background and relevant work in this area. This study identifies the requirements for edge-enabled DS based on discussions with stakeholders dealing with different data integration, reusability needs, and desire for integrated value-chain development. As a result, a novel context-aware DS model with semantic capabilities is proposed and prototyped in a lab environment. In addition, the deployment is supported by the edge-oriented resources pool infrastructure and orchestrated following the extended ALC approach based on predefined service catalog artifacts. The proposed DS model is also validated against identified requirements following a wind turbine use case. As an outlook, we would like to detail this model further for each layer with concrete implementation for diverse cross-domain use cases. Finally, this study contributes knowledge on how context-aware DS ecosystems for data integration can be realized in local or regional contexts at a small scale by exploiting relevant resources at the edge in real-world scenarios. In addition, this study advances the knowledge on the use of semantic adaptation, smart governance, and context-aware data lake for enhancing the efficiency of cross-domain data management operations and value chain development. Thus, adding value in the context of evolving industry5.0 ecosystems and upcoming technologies, such as 6G, in the future internet landscape.

References

[1] Solmaz, G., Cirillo, F., Fürst, J., Jacobs, T., Bauer, M., Kovacs, E., Santana, J. R., and Sánchez, L., "Enabling data spaces: Existing developments and challenges," in *Proceedings of the 1st International Workshop on Data Economy*, 2022, pp. 42–48.

[2] Baloup, J., Bayamlıoğlu, E., Benmayor, A., Ducuing, C., Dutkiewicz, L., Lalova-Spinks, T., Miadzvetskaya, Y., and Peeters, B., "White paper on the data governance act," 2021.

[3] Singh, P. and al., et., "Meta standard requirements for harmonizing dataspace integration at the edge," in *2023 IEEE Conference on Standards for Communications and Networking (CSCN)*, 2023.

[4] Jeffery, S. R., Franklin, M. J., and Halevy, A. Y., "Pay-as-you-go user feedback for dataspace systems," in *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, 2008, pp. 847–860.

[5] Ning, H. and Wang, T., "Constructing a dataspace based on metadata and ontology for complicated scientific data management," in *2007 2nd International Conference on Pervasive Computing and Applications*, IEEE, 2007, 512–514a.

[6] Curry, E., Derguech, W., Hasan, S., Kouroupetroglou, C., and Hassan, U. ul, "A real-time linked dataspace for the internet of things: Enabling "pay-as-you-go" data management in smart environments," *Future Generation Computer Systems*, vol. 90, pp. 405–422, 2019.

[7] Singh, P., Beliatis, M. J., and Presser, M., "Enabling edge-driven dataspace integration through convergence of distributed technologies," *Internet of Things*, vol. 25, p. 101 087, 2024, ISSN: 2542-6605. DOI: https://doi.org/10.1016/j.iot.2024.101087.

[8] Curry, E., Scerri, S., and Tuikka, T., *Data Spaces: Design, Deployment and Future Directions*. Springer Nature, 2022.

[9] Heath, T. and Bizer, C., *Linked data: Evolving the web into a global data space*. Springer Nature, 2022.

[10] Duan, Q., Wang, S., and Ansari, N., "Convergence of networking and cloud/edge computing: Status, challenges, and opportunities," *IEEE Network*, vol. 34, no. 6, pp. 148–155, 2020.

[11] Farahani, B., Firouzi, F., and Luecking, M., "The convergence of iot and distributed ledger technologies (dlt): Opportunities, challenges, and solutions," *Journal of Network and Computer Applications*, vol. 177, p. 102 936, 2021.

[12] Lee, H., Lee, B., Yang, H., Kim, J., Kim, S., Shin, W., Shim, B., and Poor, H. V., "Towards 6g hyper-connectivity: Vision, challenges, and key enabling technologies," *Journal of Communications and Networks*, 2023.

[13] Association, I. D. S., *Gaia-x and ids. position paper, version 1.0 01*, 2021.

[14] Zehnder, P., Wiener, P., Straub, T., and Riemer, D., "Streampipes connect: Semantics-based edge adapters for the iiot," in *The Semantic Web: 17th International Conference, ESWC 2020, Heraklion, Crete, Greece, May 31–June 4, 2020, Proceedings 17*, Springer, 2020, pp. 665–680.

[15] Hutterer, A. and Krumay, B., "Integrating heterogeneous data in dataspaces-a systematic mapping study," *Pacific Asia Conference on Information Systems*, 2022.

[16] Hernandez, J., McKenna, L., and Brennan, R., "Tikd: A trusted integrated knowledge dataspace for sensitive healthcare data sharing," in *2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC)*, IEEE, 2021, pp. 1855–1860.

[17] Seidel, A., Wenzel, K., Hänel, A., Teicher, U., Weiß, A., Schäfer, U., Ihlenfeldt, S., Eisenmann, H., and Ernst, H., "Towards a seamless data cycle for space components: Considerations from the growing european future digital ecosystem gaia-x," *CEAS Space Journal*, pp. 1–15, 2023.

[18] Otto, B., "A federated infrastructure for european data spaces," *Communications of the ACM*, vol. 65, no. 4, pp. 44–45, 2022.

[19] Commission, E., "Data act," 2022.

[20] Paulus, A., Pomp, A., and Meisen, T., "The plasma framework: Laying the path to domain-specific semantics in dataspaces," in *Companion Proceedings of the ACM Web Conference 2023*, 2023, pp. 1474–1479.

[21] Angles, R., Thakkar, H., and Tomaszuk, D., "Rdf and property graphs interoperability: Status and issues.," *AMW*, vol. 2369, 2019.

[22] Martella, F., Lukaj, V., Fazio, M., Celesti, A., and Villari, M., "On-demand and automatic deployment of microservice at the edge based on ngsi-ld," in *2023 31st Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP)*, IEEE, 2023, pp. 314–320.

[23] CIM, E., "Ngsi-ld cim 009," ETSI, Tech. Rep., 2019, Times cited: 3.

[24] Möller, J., Jankowski, D., Lamm, A., and Hahn, A., "Data management architecture for service-oriented maritime testbeds," *IEEE Open Journal of Intelligent Transportation Systems*, vol. 3, pp. 631–649, 2022.

[25] Bryson, J. M., "What to do when stakeholders matter: Stakeholder identification and analysis techniques," *Public management review*, vol. 6, no. 1, pp. 21–53, 2004.

[26] Niranjan, V., Anand, S., and Kunti, K., "Shared data services: An architectural approach," in *IEEE International Conference on Web Services (ICWS'05)*, IEEE, 2005.

[27] Bader, S. *et al.*, "The international data spaces information model–an ontology for sovereign exchange of digital content," in *International Semantic Web Conference*, Springer, 2020, pp. 176–192.

[28] Khalil, M. E., Ghani, K., and Khalil, W., "Onion architecture: A new approach for xaas (every-thing-as-a service) based virtual collaborations," in *2016 13th Learning and Technology Conference (L&T)*, IEEE, 2016, pp. 1–7.

[29] Aksakalli, I. K., C, T., C, A. B., and T, B., "Deployment and communication patterns in microservice architectures: A systematic literature review," *Journal of Systems and Software*, vol. 180, p. 111 014, 2021.

[30] Sibold, J., “Learning" a la carte": A theory-based tool for maximizing student engagement.,” *Journal of College Teaching & Learning*, vol. 13, no. 2, pp. 79–84, 2016.

[31] Singh, P., Holm, K., Beliatis, M. J., Ionita, A., Presser, M., Wolfgang, P., and Goduscheit, R. C., “Blockchain for economy of scale in wind industry: A demo case,” in *Global IoT Summit*, Springer, 2022, pp. 175–186.

[32] Gassmann, O., Frankenberger, K., and Csik, M., “The st. gallen business model navigator,” *Int. J. Prod. Dev*, vol. 18, pp. 249–273, 2013.

[33] Berton, L., “Ansible for k8s management,” in *Ansible for Kubernetes by Example*, Springer, 2023, pp. 201–237.

[34] Hao, Y., “Edge computing on low availability devices with k3s in a smart home iot system,” Ph.D. dissertation, The Cooper Union for the Advancement of Science and Art, 2022.

Chapter 8

Paper VI - Navigating the International Data Space To Build Edge-Driven Cross-Domain Dataspace Ecosystem

Parwinder Singh, Nirvana Meratnia, Michail J. Beliatis, Mirko Presser

Originally published in *Global Internet of Things and Edge Computing Summit. GIECS 2024. Communications in Computer and Information Science, Springer, Cham*, September 2024, volume 2328, pp. 151–168. DOI: 10.1007/978-3-031-78572-6_10.

VI

Abstract

Data is an asset for the modern industrial era that plays a crucial role in driving intelligence through machine learning techniques and methods for business, technical, economic, and social decision making. Generally, data generated in a specific domain is often used in other domains, which requires cross-domain data integration (CDDI) and sharing across edges. Here, an edge represents an organization, system, or entity with computing data processing capabilities. This CDDI and sharing across edges need data sovereignty to govern the relevant usage context, wherein the sharing edge domain's data context must be respected by the shared edge domain. This edge domain's data context can be seen as a medium to define rules for associated data usage, access, and identity management. This can be achieved by the common dataspace vision of the European Union for CDDI to turn the European market into a unified data-driven European market. International Data Space (IDS) is such an effort backed up by the European Commission to develop relevant standards and specifications for CDDI. However, even though a vast amount of IDS ecosystem information is available online, it is scattered and hard to navigate. This obstructs the use and adoption of the IDS at a desired pace and in a simplified manner. This study contributes to this need by converging the information in one place and leveraging theoretical and pragmatic insights on building IDS-based edge-driven dataspace in real-world scenarios. We dive into a wind industry supply-chain-specific use case realized through a locally developed IDS platform to showcase and validate the use of CDDI.

8.1 Introduction

Data has recently become a critical asset for modern industries worldwide. This can also be seen in the fourth industrial revolution, i.e., Industry 4.0, which revolves around Big data and associated technologies such as Cloud Computing, Artificial Intelligence (AI), Machine Learning (ML), and Internet of Things (IoT) [1] for the processing of data with 4V (Volumes Veracity, Velocity, and Variety) characteristics [2]. Data is crucial in deducing intelligence by applying ML-based models, including regression, classification, clustering, association, and control methods [3]. Applying these methods to data generates informational insights that help business, technical, economic, and social decision making [4]. Therefore, data has an economic value that can play a significant role in the sustainability and growth of any given industry or domain [5]. In addition, with the advancement of 5G/6G networks and Internet of Things (IoT) technologies, the generation and collection of data has become more viable [6].

Generally, data generated in a specific organization is often associated with its domain knowledge or context [7], [8]. In the current study context, an edge (following edge computing paradigm) represents an organization, system, device, or any entity that has significant infrastructural and system capabilities for different data processing requirements under Dataspace (DS) ecosystem [9],[10]. Therefore, each edge will have an associated domain context, which can be represented by the ontological methods that define how different entities within domain w.r.t data are linked to each other [7], [8]. This domain-level informational context can also include the terms or conditions under which the relevant data of this domain can be used in other domains [11]. We call this contextual use of data, which is a fundamental block in achieving data sovereignty [12]. Most of the time, data generated within a domain remains in the domain boundaries, making data silos and failures to achieve the mutual gain (e.g., generating revenue streams or services) that can arise from its sharing with others [9]. This happens for multiple reasons, including domain-specific problems, competitiveness feeling, and lack of data-sharing methods that support data owners' interests, trust, security, transparency, privacy, complexity, interoperability and integration, and lack of knowledge [10]. All these result in data not being used to its full potential, which otherwise could have created numerous opportunities [9]. This problem can be addressed through the concept of Dataspace. Dataspace is not a new concept, but its definition and usage have evolved. However, one thing that remains persistent w.r.t its definition is that *data owners always perceived to have complete control over their data* [13], which also reflects through data sovereignty objective [14]. Authors of [14] defined Dataspace as a *"federated, open infrastructure for sovereign data sharing based on common policies, rules, and standards."* From EU perspective [14], [15], the Dataspace is supposed to have the following functional characteristics:

- **Security and Privacy Preserving Infrastructure**: The Dataspace should have a robust IT infrastructure (generally distributed in nature) that supports the security and privacy of data. This infrastructure should enable data pooling, accessing, processing, and sharing while safeguarding individual's or sharing domain's privacy rights.

- **Data Governance Methods:** Dataspace comprises of administrative and contractual rules or policies that will govern different rights to different data operations such as access, process, use, share, etc. It is important for exchanging clear information in the form of guidelines and responsibilities for data management while data is crossing its organizational boundaries.

- **Protecting Data Owner Rights:** Data holders always have control over their data life cycle regarding its usage conditions, access, and permissions related to data management to define data usage contexts and intentions while data is being shared in cross domains [9]. From the data owner's perspective, it is important to determine how the data is being used or accessed to protect the data owner's rights over data.

- **Monetary Value:** Data made available in a Dataspace, based on a volunteer basis, can be used against remuneration or without cost, depending on the decision of the data owner.

- **Open Participation:** The Dataspace should support open participation for all, including any organization or individual, adhering to participation rules, ensuring a fair and level playing field by avoiding any discrimination for any participant.

To support cross-domain data integration (CDDI), the EU has taken many initiatives to promote data sharing, making it open and accessible to all with fair usage [16]. The initiative includes developing the European Data Strategy [17], Common European Data Spaces [13], for which a working document was released in 2022, a cross-sectoral legislative framework comprised of the Data Governance Act [18], Data Act [19], Implementing Act on High-Value Datasets also known as Open Data Directive [20]. EU also sets DSSC (Data Spaces Support Center) to coordinate and govern the actions of these initiatives. In addition, the EU has also set up a European Data Innovation Board, which is the consultative and advisory body that provides guidelines for the interoperability of Common European Data Spaces [14].

The objective of all aforementioned initiatives is to enable the opening up and sharing of data in different domains through CDDI, so that the true potential of converged data can be exploited in different dimensions. Figure 8.1 illustrates the CDDI concept among different edge domains, which can generate cross-organizational value chain streams [10]. However, this requires standardized methods for harmonizing data across disparate edge domains [9]. In addition, equally important is safeguarding the interests of data owners through implementation of data sovereignty. This entails respecting the contextual and regulatory frameworks governing the data life cycle within each domain during the CDDI or sharing process [18]. Essentially, CDDI necessitates technical integration and adherence to legal and regulatory standards to ensure integrity and security of data while facilitating effective cross-domain collaboration. In CDDI, the domain-specific data context can be seen as the boundary at the edge where the data owner can define usage, access, and identity control or management of associated data operations. European Union foresaw building such data spaces for CDDI, which is essential to support the vision of turning the European market into a unified market based on data-driven value streams [13].

Aligning to the EU vision, an effort for International/Industrial Data Space (IDS) was initiated by the Fraunhofer Institute in 2015 [21] and backed up by the European Commission to develop relevant standards and specifications. To this end, IDS Association was formed in 2017 and developed IDS Reference Architecture Model, Dataspace protocol specification, and related IDS prototype, which can uptake the CDDI activity aligning EU vision to converge data among European industries. However, understanding the IDS ecosystem and setting it up at the local or regional level is a huge challenge due to the vast amount of scattered information available online, which is complex and hard to navigate. This obstructs the use and adoption of IDS in a simplified and understandable manner. This raises an important research question (RQ), namely,

How can IDS-based standards, specifications, and reference implementation be used to build a local or regional Dataspace for a target use case focusing on edge-driven cross-domain data integration?

This paper contributes to addressing the above RQ by converging the information around the EU's vision of building Dataspaces, related work, and IDS specifically in one place and in a simplified manner. We leverage theoretical and pragmatic insights on building IDS-based Dataspaces in real-world scenarios through a wind turbine industry supply chain-specific use case to achieve CDDI. In addition, we also provide knowledge on how a specific use case can be semantically modeled through the IDS information model, followed by its implementation through a locally developed IDS platform prototype, which supports the requirement for CDDI with functional compliance and validation.

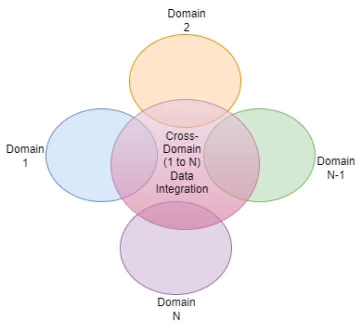


Figure 8.1: Cross-Domain Data Driven Dataspace

8.2 Literature Review

Applications of Dataspace in diverse use cases can bring various advantages for CDDI, such as the wind industry supply chain or the role of digital transformation in roll-to-roll printing in Industry 4.0 settings [22],[23]. However, CDDI must address interoperability and heterogeneity challenges (through meta-standards) induced by diverse standards and protocols used in different organizational boundaries for data sharing, reuse, and distributed operations [9]. [24] presents the role of Blockchain in the IDS-based data exchange ecosystem for data sovereignty and implementation of a clear housing data trading platform. [25] has presented in-depth knowledge on security aspects for usage and access control for achieving data sovereignty, identity, and integrity management of components to embed trust through Public Key Infrastructure (PKI) methods among different participants in the Dataspace ecosystem. [26] presents an architectural framework to develop a cross-domain distributed infrastructural ecosystem based on Processing, Service, and Data context by extending the traditional Cloud-Edge-Device (CED) continuum by introducing three new layers, namely, Semantic, Convergence, and Dataspace integration.

[27] introduced a prototype for IDS Security Components tailored to the Textile and Clothing Industry, ensuring data sovereignty and facilitating secure interactions among supply chain participants. [28] suggests leveraging Dataspaces and digital twins to implement the Industry 4.0 vision, exploring key components such as integrating digital twins within Dataspaces according to the latest Industry 4.0 Asset Administration Shell specifications. [29] targets bridging the gap between IDS and Plattform Industrie 4.0 (PI4.0) by analyzing their concepts and tools and recommending a generic approach for combining different technologies, irrespective of specific IDS or Industry 4.0 implementations. [30] proposed the establishment of an International Testbed for Dataspace Technology, i.e., a testbed for developing and testing data platform technologies' interoperability, portability, and customizability. [31] presents the reference implementation of IDS specifications, named *True Connector*, developed by Engineering. Authors of [32] present the learning from implementing IDS-based Dataspace Connector (DSC) and their interactions at a high level. This is quite similar to our work, but it only covers the architectural perspectives. In a way, we are extending their study with the latest findings and insights for IDS implementation.

8.3 Methodology

The constructive research approach [33] has been followed in our study to guide the design, development, and deployment of our IDS-based Dataspace prototype, ensuring its effective and sustainable adoption. We studied many EU Commission policies and working documents to determine the role of Dataspace in the futuristic industrial landscape in terms of CDDI. Finally, we recorded theoretical information, such as relevant background and semantic modeling of IDS from a target use case perspective, as the basis for our study to conduct practical experiments and to build an IDS-based Dataspace ecosystem. In this context, we have investigated the IDS Connector, IDS Testbed [30], and True Connector (from Engineering) [31] as IDS reference implementation [32]. After that, based on this open-source implementation [31], [34], we developed our prototype with default settings for IDS-based data integration in different domains and participants acting as data producers and consumers. After completing the prototype with default implementation, we applied our wind turbine supply chain use case to an IDS information model. We then prototyped and evaluated it, focusing on IDS connector implementation, standards compliance, and protocol message exchange sequence. The relevant services of IDS implementation have been deployed following the Microservices design architecture. For this, Docker-based toolchain was used over virtual infrastructure to instantiate and validate the prototype. An online Resource Description Framework (RDF) graphic visual tool was used for semantic modeling (ontology) visualization. During prototype development, many challenges, such as the deployment of services failure, semantic modeling adaptation, API validation failures, inter-component connectivity, and service discovery-related issues, were experienced and solved, and the relevant lessons learned have been documented.

As a contribution, we have advanced the relevant knowledge for building IDS-based Dataspace at edge, focusing on the target use case from theoretical and practical perspectives. We believe this will also contribute to accelerating the adoption of IDS-based standardized Dataspace implementation across Europe, specifically focusing on the industrial region of Denmark in our study.

8.4 Use Case Specific IDS Development

8.4.1 IDS Background

IDS offers a framework to exchange data in a cross-domain environment offering data sovereignty, including access and usage control [35]. The Reference Architecture Model [35] of IDS consists of the following layers:

- *Business Layer:* provides different roles (data providers, consumers, brokers, etc.) and their interactions, encompassing contracts and relevant data usage policies.
- *Functional Layer:* provides requirements for trust, security, data sovereignty, interoperability, value-added applications, and data marketization.
- *Process Layer:* specifies the processes and interactions between components for data exchange, including policy enforcement.
- *Information Model:* provides a standardized way to describe and represent data entities using the RDF/JSON-LD contextual format and IDS ontology, enabling interoperability.

- *System Layer:* defines and realizes the technical components, unified namespace, and their interactions for implementing the IDS infrastructure.

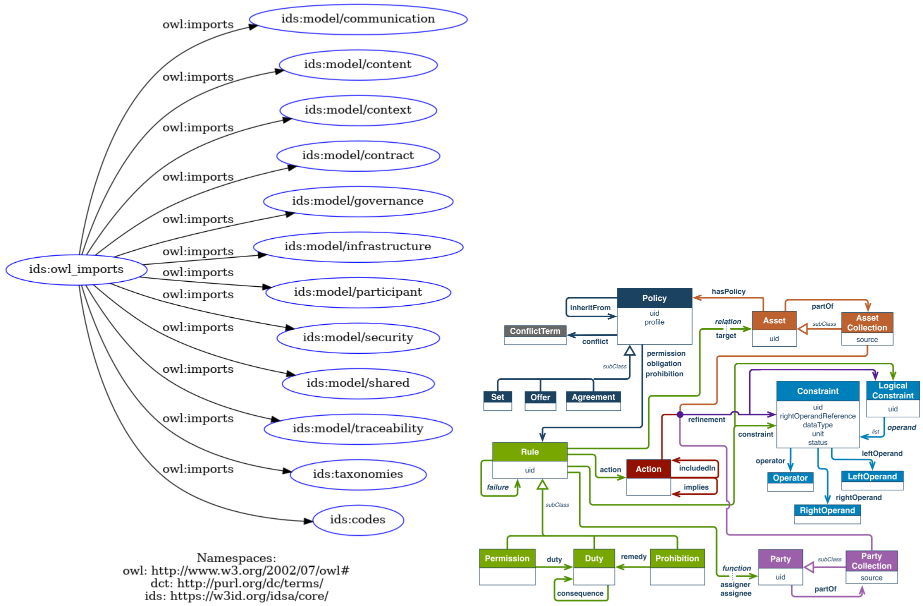
IDS Connector is a key component in realizing IDS architecture that facilitates secure, traceable, and trusted data exchange among stakeholders or participants in the cross-domain data ecosystem [21]. The connector may have multiple components, such as data adapters, security/identity modules, policy/rule engines, metadata repositories, and cross-domain communication interfaces, to support IDS functionality effectively [32]. This connector is a main gateway for data exchange between participating parties, including data providers, consumers, and intermediaries. It enables the organizational edges to share, reuse, and integrate data within the IDS-distributed infrastructure, including deployment over cloud, edge, on-premises, etc., depending on the specific requirements of the target use case. This connector enforces data governance to achieve data sovereignty through policies and access controls, ensuring data is accessed and shared according to predefined rules, regulations, and exchange protocols, providing interoperability among participating entities. This includes security mechanisms such as authorization, authentication, encryption, and access control to protect CIA (confidentiality, integrity, and availability) Triad. It also offers the publication of catalogs through a metadata broker that provides the availability of metadata to the requesting participant for the data to be exchanged [31]. This metadata context provides semantics and provenance information to facilitate data discovery, understanding, and utilization. IDS Connector can be applied to various use cases across industries such as healthcare, smart cities, logistics, energy, etc.

8.4.2 IDS information Model

IDS information model is the key core element for the participating entity to participate in an IDS-based Dataspace [11]. The IDS information model defines metadata or data elements using a common vocabulary, technically called ontology. It must be used for data modeling while sharing the data in cross-domain integration. The relevant information model is shown in Figure 8.2a. IDS semantic information model bridges the IDS ecosystem's expression, infrastructure, and enforcement components for delivering seamless data integration and interoperability in cross-domain integration [11].

IDS info model uses the Open Digital Rights Language (ODRL) model, shown in Figure 8.2b, [36] to define the relevant usage policies and permissions to achieve data sovereignty. ODRL, as shown in Figure 8.2b, provides a standardized vocabulary and information model for representing digital rights expressions, such as permissions, prohibitions, obligations, and constraints associated with digital content or services. It allows data holders to specify how their data can be shared, distributed, or consumed by other participants of the IDS ecosystem.

Figure 8.2a provides a high-level view of IDS information model/ontology generated based on IDS source code [34]. It is observed that IDS ontology (OWL-based) imports from different models related to Communication, Content,



(a) IDS Information Model (Code View). (b) ODRL Model. Sourced from [36].

Figure 8.2: IDS and ODRL Information Model

Context, Contract, Governance, Infrastructure, Participant, Security, Shared, Traceability, Taxonomies, and Codes. Each importing model/ontology encapsulates specific functional entities relevant to its domain. For instance, Communication encompasses components like AppRoute, Endpoint, Message, and Proxy, reflecting elements involved in data transmission. Context incorporates SpatialEntity and TemporalEntity, which are indicative of spatial and temporal considerations within the system's environment. Content includes Artifact, Asset, DataApp, DigitalContent, Language, MediaType, PaymentModalities, Representation, Resource, UsageControlObject related ontologies suggesting diverse content types managed within the IDS system. Contract includes ODRL-based Action, Constraints, Rules, BinaryOperator, UsageControl, Contracts, LeftOperand, and UsagePolicyClasses, which are elements pertinent to contractual agreements. Similarly, Infrastructure encompasses infrastructure components such as AppStore, Broker, Catalog, ClearingHouse, ConfigurationModel, Connector, DAPS, IdentityProvider, ParIS, and PublicKey, hinting at the foundational elements supporting IDS soft infrastructure [15] system operations. In a nutshell, this info model represents the IDS system's architecture and organization at the implementation level, facilitating a better understanding and communication of its intricacies.

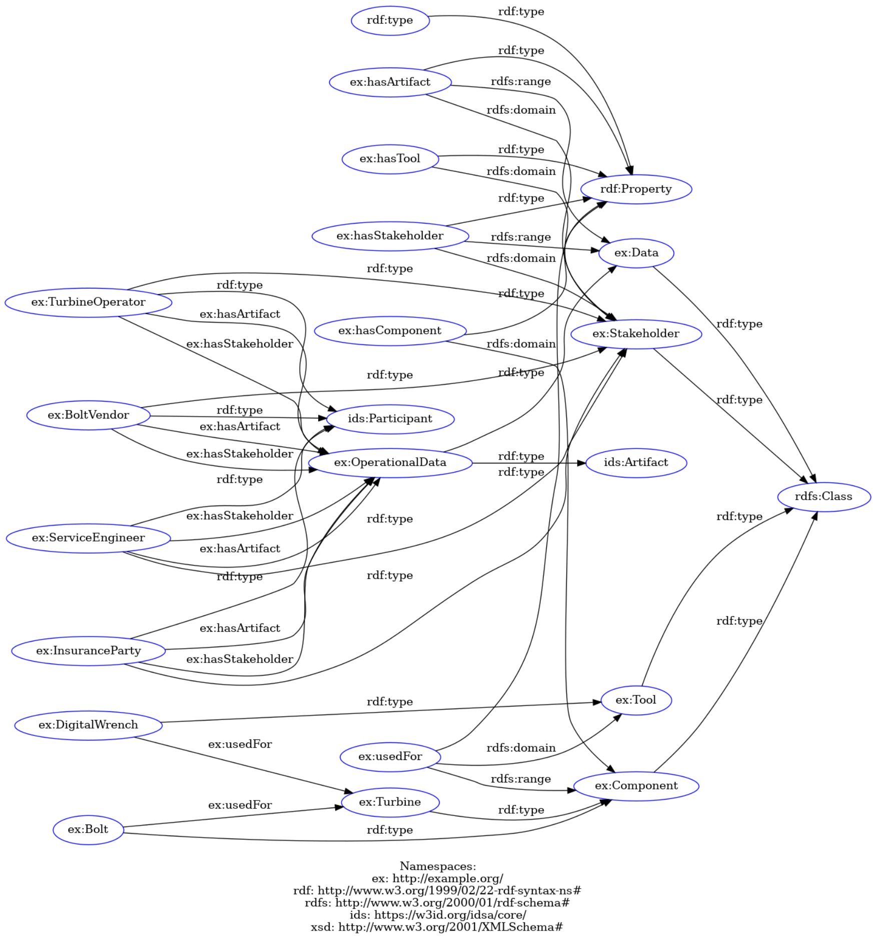


Figure 8.3: IDS Based Wind Turbine Bolt Operations Info Model

8.4.3 Use Case

Our use case, presented in [22], and [10], revolves around the supply chain events associated with wind turbine operations, with a particular focus on full product life cycle and bolt-related activities. This illustrates the collaboration among various stakeholders across different domains. The bolt vendors or suppliers are at the heart of this supply chain, responsible for manufacturing bolts in their production units and delivering them in batches to the turbine operators. The turbine operators, who own and manage the turbines, rely on third-party service engineers for maintenance and operational support. Additionally, an insurance partner plays a crucial role in mitigating risks by offering insurance services to protect the interests of the turbine operators against unforeseen events. Wind turbines have multiple components, including blades, towers, bolts, motors, electrical cables, rotors, electronic sensors, etc. However, our focus is on bolt-related operations management within the turbine context. Bolts play a critical role in turbine operations, and ensuring their timely maintenance is paramount. The main objective of this use case is to facilitate the exchange of bolt-specific operational information among various stakeholders, including turbine operators, bolt suppliers, service engineers, and insurance partners across different domains. This relationship is illustrated in the semantic modeling diagram of Figure 8.3. This shows the semantic relationship among stakeholders operating in different domains, tools, and components like bolt-batch and turbine. Operational data represents the event data the Service Engineer recorded for the bolt operation over the turbine. Then, this data is exchanged for different purposes or actions and provided to different stakeholders, such as bolt suppliers and insurance parties. For example, the insurance party can validate the quality of the bolt operation during the assembly of the turbine to settle any claim in the future. Figure 8.3 shows that the stakeholders will be treated as Participants as per the IDS information model. This information model is developed using the ontology provided by the IDS GitHub repository [34] and the use case-specific data model [10], [22] mapped to this IDS model to get it processed through the IDS connector running over the IDS prototype explained in the next section.

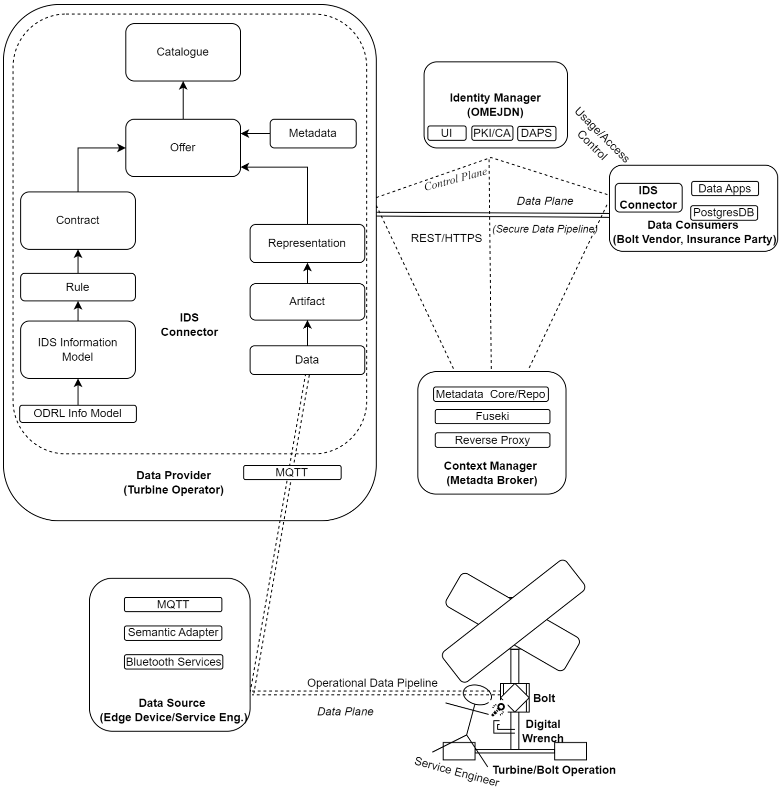


Figure 8.4: IDS Provisioned Use Case Specific System Architecture

8.4.4 System Context

Figure 8.4 provides the system architecture for our use case presented earlier in section 8.4.3. There are multiple architectural blocks whose description is given below:

- **Data Source**: is the turbine site, where the physical operation on the bolt is performed. The Service Engineer uses the digital wrench to perform the bolt-tightening operation and scan the turbine and bolt-related bar codes as their digital identity. The relevant operational events are transmitted as data streams through a data pipeline from these physical objects to a nearby edge computing device. This data source, consisting of multiple services such as Bluetooth (BLE) protocol stack termination, MQTT agent, and Semantic Adapter (service to transform the BLE data into JSON format) are running at the edge, technically a Raspberry Pi device in our prototype setup, and sending the data towards the Data Producer.
- **Data Producer**: is an organizational edge [9], [10], which has on-premises setup, consumes data over the MQTT interface from Data Source and forwards the data towards the IDS connector that is also running as a service on the same system. This is called Data Producer from the IDS ecosystem view, as the data stream is offered/produced as a catalog in the Dataspace.
- **Data Consumers**: are other organizational edges [9], [10], which have on-premises setup and consume data over the IDS interface (HTTPS/REST in our case) from the catalog made available to Dataspace by the Data Producer. Multiple applications can run on the consumer side that needs the consumed data to offer innovative services. In our case, consumers such as bolt vendors and insurance partners are using this data to reconcile their entries and monitor the quality of their supplied products (bolt batches) to settle any issues based on received data in the future.
- **Identity Management(IDM)**: is the security module of the IDS ecosystem, which is responsible for establishing trust between the data-exchanging parties or participants. The IDM also consists of a functional component called Dynamic Attribute Provisioning Service (DAPS) within the IDS [32]. This service manages the digital identity of participants/systems in IDS that rely on different attributes linked to that identity [27]. The DAPS service provides dynamic and updated attribute information about participants and connectors. IDM consists of services and mechanisms designed to ensure the CIA-Triad of data exchanged between participants in the IDS ecosystem. It provides the requested parties a DAPS token as per OAUTH2 [37] token workflow and standards. It also uses the Public Key Infrastructure (PKI) to generate X.509 certificates [38]. IDM architectural block also consists of an open-source implementation of OAUTH2, which Fraunhofer provided and called *Omejdn* in the IDS implementation context.

- **Metadata Broker**: is the cataloging system where the offers, resources, or catalogs are registered, discovered, and made available to different participants in the IDS ecosystem [32]. In our case, the Data Producer registers its catalog offerings (with resources data/metadata) with the broker. The data consumer can query the availability of desired/interested resources (data/metadata) from the metadata brokers to fetch the endpoints of the data producer and kick off the data exchange negotiation process. This metadata broker is also responsible for managing metadata-related contexts. In our case, there is a service called *Fuseki* running as part of the IDS context management architectural block, which is an abstract name for Apache Jena Fuseki, i.e., a SPARQL Protocol and RDF Query Language server [39]. This provides services for semantic queries based on RDF standards and ontologies-driven data models. It allows requesting parties within the IDS ecosystem to query certain catalogs or resources based on RDF query strings, which typically follow the Subject-Verb-Predicate (aka triple) structure [39]. This semantic capability enables data producers and consumers to access and retrieve relevant data within the IDS ecosystem efficiently. In addition, there is also a proxy service running, which is at the forefront and exposes the interface for all broker services to the external world through a unified URL interface.

The architectural message flow consists of two communication planes: Control Plane (CP) and Data Plane (DP). CP is responsible for establishing contracts, registering catalogs or artifacts, authentication, authorization, and negotiation between data producers and consumers. On the other hand, DP manages the actual forwarding and processing of data between producer and consumer.

8.4.5 System Validation

The system context explained earlier is deployed independently in two different modes, i.e., (i) in the Azure cloud environment and (ii) in the on-premises environment (on a local laptop). In both cases, the virtual infrastructure was instantiated using virtualization services with the following specifications: VCPU: 2cores; RAM: 8GB; Hard-Disk: 20GB; OS: Ubuntu20.04. A VM was created using these specifications, followed by the necessary Docker and Docker-Compose packages installation for microservices instantiations [40]. After completion of the Docker platform, the Gitlab repository was cloned from [34] to build our use case-specific IDS environment. [34] provides detailed steps to clone, install, and provision the IDS environment and start validation for different IDS standard API requests (based on the Postman tool).

Figure 8.5 depicts the deployment view of all the microservices running successfully and ready to be validated for handling API requests between the Data Producer and Data Consumer. All API requests and their end-to-end response time are shown in Figure 8.6. One can see that it is less time-consuming when the requests are landing directly at specific components such as Data Producer or Consumer. However, the response time is significantly

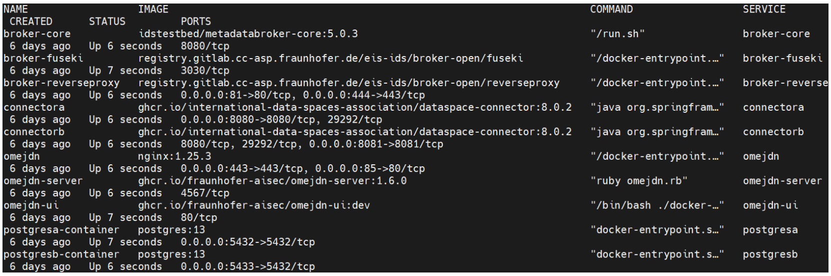


Figure 8.5: IDS Microservices Based Deployment View

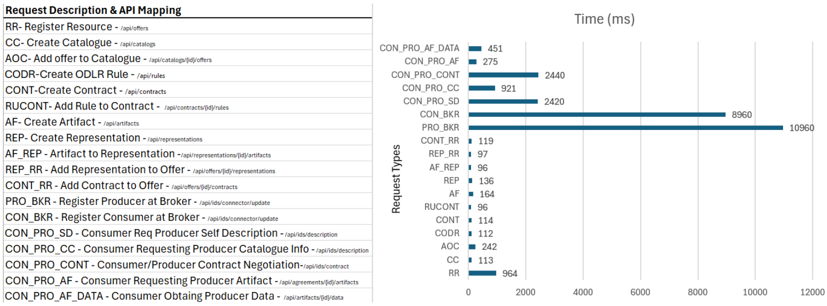
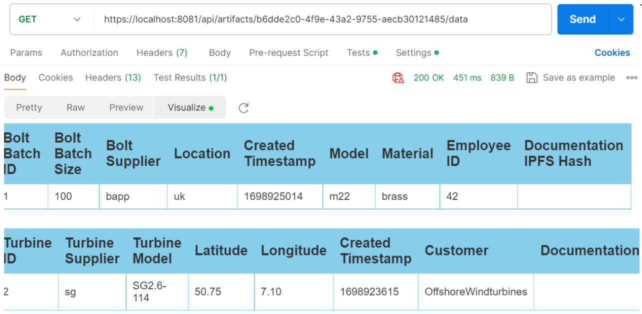


Figure 8.6: IDS Validation Results - Request Vs Response Time

increased when the Metadata Broker comes into play for the registration of an entity (producer/consumer) at the metadata broker or the initial discovery of information by an entity in the IDS ecosystem when the broker needs to find the availability of a catalog. This happens because the IDM intercepts the flow when one entity wants to interact with the other party. In the entity registration case, the entity needs to be validated by IDM through the OAUTH2 token workflow before allowing access to the Metadata Broker. Similarly, when a Data Consumer interacts with the Producer directly, it has to be authenticated, authorized by the IDM, as a trusted third party, in order to start negotiation with the producer. This also demonstrates the usage of access control and data sovereignty-related functional validation of IDS specifications.

The request message types that are shown in Figure 8.6 and Figure 8.4 are executed by Dataspace applications (in our case, Postman tool) towards the Data Producer and Data Consumer, and these requests interact with the IDM and Metadata Broker services in their backends (via relevant front-ends, e.g. proxy URL). For example, a register connector request `{PoC_EP}/api/ids/connector/update?recipient={BROKER_EP}` at the data producer or consumer endpoints registers them with the provided metadata broker endpoint after it gets validated through IDM based on OAuth2 standard (using Omejdn service) workflow to enable them for cross-domain digital negotiations for data exchange. The following list presents the high-level flow of functional requests illustrated in the system architecture of Figure 8.4:

- **Offer:** creates an offer/resource for the catalog at Data Producer.



Bolt Batch ID	Bolt Batch Size	Bolt Supplier	Location	Created Timestamp	Model	Material	Employee ID	Documentation IPFS Hash
1	100	bapp	uk	1698925014	m22	brass	42	

Turbine ID	Turbine Supplier	Turbine Model	Latitude	Longitude	Created Timestamp	Customer	Documentation
2	sg	SG2.6-114	50.75	7.10	1698923615	OffshoreWindturbines	

Figure 8.7: IDS Based Cross-Domain (Consumer/Producer) Data Request

- **Catalog:** creates a catalog at Data Producer.
- **Contract:** creates a contract template for the data to be used from the catalog by external parties.
- **Rule:** creates a rule or Policies (as per IDS information and ODRL model) to be enforced for data usage during cross-domain data exchange. This rule is then added to the Contract metadata.
- **IDS Information model:** is the IDS ecosystem's default ontology or domain model to stay compliant with IDS standards and specifications to achieve interoperability and integration. Data models of all requests must comply with this model.
- **Representation:** specifies the format and language in which the artifact will be presented. For example, it could be in JSON/CSV format using English.
- **Artifact:** holds the actual data to be shared under the "*value*" field. Here, the JSON data payload needs to be encoded in string format.

During validation, entities such as offers, catalogs, contracts, rules, artifacts, and representations were generated through HTTPS POST requests at the Data Producer's end (see Figure 8.6). After validation, the Data Producer and Consumer register these entities with the Metadata Broker, enabling data exchange through IDM validation. Once registered, the Data Consumer acquires metadata about the catalog from the Metadata broker and initiates direct negotiations with the Data Producer. These negotiations (trusted through IDM) involve agreeing upon and accepting the contract terms and rules stipulated by the producer. Upon successful negotiation, the consumer gains access to the artifact to retrieve the desired data. In our case, one sample of the received data, at the consumer end application, after the negotiation is shown in Figure 8.7.

During deployment, we faced several challenges, which we fixed. This includes `Docker` level changes where the `version` attribute was missing in *docker-compose.yml*. Also, during the `Registration` request at MetaDatabroker, the `Omejdn` service was found to be unreachable due to using `localhost` in the service discovery URL. This was resolved by switching to a container-level URL, `broker-reverse-proxy`, representing the container-level IP, ensuring proper internal microservice discovery. Apart from that, at the environment level, we found that the `.env` file, while instantiating the docker-compose environment, using quotes, caused issues. It's recommended to omit them or use `absolute paths` for clarity and to avoid potential errors. Finally, during test validation, the `Postman` collection script encountered difficulties extracting parameter values, particularly the `Consumer Agreement Id`, from the current request to populate subsequent requests. This issue has been resolved by putting a generic code to extract parameter values, and the fix has already been pushed to the IDS main repository [41].

8.5 Conclusion

Data has become an economic asset for modern era. It is the key to the future for any industry to sustain and grow through driving data intelligence based on modern machine learning techniques and methods. Industries that operate in silos cannot exploit the true potential of their data. Therefore, the European Commission (EC) has envisioned developing cross-domain data integration (CDDI) platforms called Common Data Spaces to thrive in data-driven value chains and benefit the social, industrial, and economic factors. International Data Space (IDS) was such an initiative backed by the EC to uplift CDDI activities at regional and national levels across Europe. IDS association has developed IDS specifications, data exchange standards, and reference implementations to be used by the target industrial use case. However, the edge-driven CDDI through IDS was found to be complex due to the vast amount of scattered and hard-to-navigate information. Therefore, as a contribution, this study has developed the knowledge on the main research question of how to build IDS-based Dataspace for the target use case focusing on CDDI. As a result, we have modeled and evaluated a wind turbine industry use case leveraging the IDS information model. The platform is deployed following microservices architecture and evaluated over Docker-based virtualized infrastructure in the cloud and on-premises environment. The platform comprises identity management, data producer and consumer, metadata broker, and data source-based services that run and enable the edge-driven cross-domain Datsapce ecosystem together. As a contribution, we have advanced the relevant knowledge base on building IDS-based Dataspace at edge by applying target use case from theoretical and practical perspectives. We believe this will contribute to pace up the adoption of IDS-based standardized Dataspace implementations across industries.

References

[1] Meindl, B., Ayala, N. F., Mendonça, J., and Frank, A. G., "The four smarts of industry 4.0: Evolution of ten years of research and future perspectives," *Technological Forecasting and Social Change*, vol. 168, p. 120 784, 2021.

[2] Sun, Z., “Big data 4.0= meta4 (big data)= the era of big intelligence,” in *Proceedings of the 2024 7th International Conference on Software Engineering and Information Management*, 2024, pp. 14–22.

[3] Dol, S. M. and Jawandhiya, P. M., “Classification technique and its combination with clustering and association rule mining in educational data mining—a survey,” *Engineering Applications of Artificial Intelligence*, vol. 122, p. 106 071, 2023.

[4] Allam, Z. and Dhunny, Z. A., “On big data, artificial intelligence and smart cities,” *Cities*, vol. 89, pp. 80–91, 2019.

[5] Bachmann, N., Tripathi, S., Brunner, M., and Jodlbauer, H., “The contribution of data-driven technologies in achieving the sustainable development goals,” *Sustainability*, vol. 14, no. 5, p. 2497, 2022.

[6] Qiao, L., Li, Y., Chen, D., Serikawa, S., Guizani, M., and Lv, Z., “A survey on 5g/6g, ai, and robotics,” *Computers and Electrical Engineering*, vol. 95, p. 107 372, 2021.

[7] Pundt, H. and Bishr, Y., “Domain ontologies for data sharing–an example from environmental monitoring using field gis,” *Computers & Geosciences*, vol. 28, no. 1, pp. 95–102, 2002.

[8] Sun, J. and Fang, Y., “Cross-domain data sharing in distributed electronic health record systems,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 21, no. 6, pp. 754–764, 2009.

[9] Singh, P., Haq, A. U., Beliatis, M., *et al.*, “Meta standard requirements for harmonizing dataspace integration at the edge,” in *2023 IEEE Conference on Standards for Communications and Networking*, IEEE, 2023, pp. 130–135.

[10] Singh, P., Beliatis, M., Presser, M., *et al.*, “Data-driven iot ecosystem for cross business growth: An inspiration future internet model with dataspace at the edge,” in *INTERNET 2024: The Sixteenth International Conference on Evolving Internet, ISBN: 978-1-68558-133-6*, 2024.

[11] Lange, C., Langkau, J., and Bader, S. R., *The ids information model: A semantic vocabulary for sovereign data exchange.* 2022.

[12] Jarke, M., “Data sovereignty and the internet of production,” in *International Conference on Advanced Information Systems Engineering*, Springer, 2020, pp. 549–558.

[13] Scerri, S., Tuikka, T., Vallejo, I. L. de, and Curry, E., “Common european data spaces: Challenges and opportunities,” *Data Spaces: Design, Deployment and Future Directions*, pp. 337–357, 2022.

[14] Reiberg, A., Niebel, C., and Kraemer, P., “What is a data space,” *White Paper 1*, 2022.

[15] Nagel, L., Hierro, J. J., Perea, E., Lycklama, D., Mertens, C., Taillandier, A.-S., Marques, M., Gelhaar, J., Marguglio, A., Ahle, U., *et al.*, “Design principles for data spaces: Position paper,” E. ON Energy Research Center, Tech. Rep., 2021.

[16] European Commission, “A European strategy for data,” European Commission, Brussels, Communication COM(2020) 66 final, Feb. 2020.

[17] Lutz, S. U., “The european digital single market strategy: Local indicators of spatial association 2011–2016,” *Telecommunications Policy*, vol. 43, no. 5, pp. 393–410, 2019.

[18] Baloup, J., Bayamlıoğlu, E., Benmayor, A., Ducuing, C., Dutkiewicz, L., Lalova-Spinks, T., Miadzvetskaya, Y., and Peeters, B., “White paper on the data governance act,” 2021.

[19] Pathak, M., “Data governance redefined: The evolution of eu data regulations from the gdpr to the dma, dsa, dga, data act and ai act.,” *DSA, DGA, Data Act and AI Act.(February 6, 2024)*, 2024.

[20] European Commission, *Implementing Regulation for High Value Datasets*, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32023R0138, Accessed: Insert Date Here, 2023.

[21] Otto, B., Hompel, M. ten, and Wrobel, S., *Designing Data Spaces: The Ecosystem Approach to Competitive Advantage*. Springer Nature, 2022.

[22] Singh, P., Holm, K., Beliatis, M. J., Ionita, A., Presser, M., Wolfgang, P., and Goduscheit, R. C., “Blockchain for economy of scale in wind industry: A demo case,” in *Global IoT Summit*, Springer, 2022, pp. 175–186.

[23] Singh, P., Karpavice, J., Beliatis, M., Presser, M., *et al.*, “Digital dataspace and business ecosystem growth for industrial roll-to-roll label printing manufacturing: A case study,” in *SENSORCOMM 2023: The Seventeenth International Conference on Sensor Technologies and Applications*, IARIA, 2023.

[24] Prinz, W., Rose, T., and Urbach, N., *Blockchain technology and international data spaces*. 2022.

[25] Huber, M., Wessel, S., Brost, G. S., and Menz, N., *Building trust in data spaces*. 2022.

[26] Singh, P., Beliatis, M. J., and Presser, M., “Enabling edge-driven dataspace integration through convergence of distributed technologies,” *Internet of Things*, p. 101 087, 2024.

[27] Torres, N., Chaves, A., Toscano, C., and Pinto, P., “Prototyping the ids security components in the context of industry 4.0-a textile and clothing industry case study,” in *International Conference on Ubiquitous Security*, Springer, 2022, pp. 193–206.

[28] Volz, F., Sutschet, G., Stojanovic, L., and Usländer, T., “On the role of digital twins in data spaces,” *Sensors*, vol. 23, no. 17, p. 7601, 2023.

[29] Larrinaga, F., "Bridging the gap between ids and industry 4.0–lessons learned and recommendations for the future," 2024.

[30] Matsunaga, I., Michikata, T., and Koshizuka, N., "Itdt: International testbed for dataspace technology," in *2023 IEEE International Conference on Big Data (BigData)*, IEEE, 2023, pp. 4740–4747.

[31] Engineering Ingegneria Informatica S.p.A. "TRUE Connector." (Accessed 15 May 2024).

[32] Pampus, J., Jahnke, B.-F., and Quensel, R., "Evolving data space technologies: Lessons learned from an ids connector reference implementation," in *International Symposium on Leveraging Applications of Formal Methods*, Springer, 2022, pp. 366–381.

[33] Lehtiranta, L., Junnonen, J.-M., Kärnä, S., and Pekuri, L., "The constructive research approach: Problem solving for complex projects," *Designs, methods and practices for research of project management*, pp. 95–106, 2015.

[34] Parwinder. "Wind Turbine Supply Chain Dataspace Git Repository." (Accessed 15 May 2024).

[35] *International data spaces reference architecture model (ids-ram)*, Conceptual Framework, Version 4.2, International Data Spaces Association, 2024.

[36] Iannella, R. and Villata, S., "Odrl information model 2.2," W3C, Recommendation REC-odrl-model-20180215, Feb. 2018, W3C Recommendation.

[37] Hardt, D., *The OAuth 2.0 Authorization Framework*, RFC 6749, [Online]. Available: https://www.rfc-editor.org/rfc/rfc6749.txt, Oct. 2012.

[38] Cooper, D., Santesson, S., Farrell, S., Boeyen, S., Housley, R., and Polk, W., *Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile*, RFC 5280, [Online]. Available: https://www.rfc-editor.org/rfc/rfc5280.txt, May 2008.

[39] Chokshi, H. J. and Panchal, R., "Using apache jena fuseki server for execution of sparql queries in job search ontology using semantic technology," *International Journal of Innovative Research in CST*, vol. 10, no. 2, pp. 497–504, 2022.

[40] Ibrahim, M. H., Sayagh, M., and Hassan, A. E., "A study of how docker compose is used to compose multi-component systems," *Empirical Software Engineering*, vol. 26, pp. 1–27, 2021.

[41] Singh, P., *Fixes in ids testbed*, GitHub Pull Request #143, International Data Spaces Association, https://github.com/International-Data-Spaces-Association/IDS-Testbed/pull/143, Jun. 2023.

Chapter 9

Discussion & Conclusion

9.1 Discussion

This PhD research explores how distributed edge architectures and the convergence of (IDEAL) technologies such as AI, IoT, DLT, Edge Computing, and related frameworks—can address challenges in building an Edge-driven value-chain ecosystem for Industry 4.0 and beyond. Consequently, this dissertation is composed of six research papers to answer the following main research question: RQ: *How can distributed edge architectures and convergence/integration of IDEAL technologies be applied to build an edge-driven value-chain ecosystem in Industry 4.0 and beyond contexts?*

The overall research process combines theoretical exploration, system modeling, and empirical validation to address the main RQ and SRQs to fulfill the research objectives that have been mentioned in Section 1.3. Design Science Research (DSR) has been chosen as the overarching methodology for this PhD research and focused on designing and evaluating artifacts to solve complex problems and generate actionable knowledge. This dissertation leverages six research papers and different phases of the DSR process to address the different aspects of the main RQ in the following manner:

1. **Problem Identification and Motivation:** This PhD work addresses certain challenges in Industry 4.0 and emerging Industry 5.0 contexts. These challenges include addressing inefficiencies of wind industry supply chain events through data-driven digital traceability functions, digital interventions for optimized production operations in SMEs where upgrade of old machinery is not feasible, addressing heterogeneity due to standards and as well as its existence at data, service, or system level, dealing with resource constraints at the Edge during the convergence of technologies, need for dynamic implementation contexts to fulfill the diverse use case requirements, stakeholder participatory motivation in developing cross-domain value chains and distributed data operations emphasizing security and data sovereignty. These challenges are identified and explored through industrial stakeholder engagement, systematic and narrative literature reviews, prototyping, and case studies, providing the foundation for theoretical extensions and designing target solutions. These challenges are addressed cumulatively through the research work covered in the six papers.

2. **Objective Definition for Solutions:** Specific research objectives were articulated in Section 1.3 to map the SRQs to address the main RQ. These

objectives guided the research and development of artifacts to address the identified gaps and challenges mentioned earlier and in Section 1.2.

3. **Design and Development of Artifacts/Prototypes:** This PhD research has culminated various artifacts and prototypes in multiple papers. Paper I has prototyped the Edge, IoT, and Blockchain-enabled digital traceability solution for the wind turbine supply chain events related to bolt components to enhance transparency and trust in a multi-stakeholder environment. A digital transformation-related design artifact in traditional label printing processes to enable data-driven decision-making and error-monitoring ecosystems aligning with the Industry 4.0 vision has been presented in Paper II. This is done by mapping business and technical processes that include data point augmentation and system architecture development. Focusing on heterogeneity and Dataspace development challenges, a Participatory Motivation system model, Extended Data Life Cycle design for data sharing in cross-organization, and Integration Harmonization design have been presented in Paper III. Similarly, Paper IV presents **DENOS** (Distributed Edge Network Operations-oriented Semantic) model, which extends traditional Cloud-Edge-Device continuum architecture, supporting semantic processing model to realize dynamic implementation context. In addition, it also presents the conceptual IDEAL convergence capabilities model and Edge-Blockchain integration methodology.

   Finally, context-aware Dataspace and IDS-based Edge-enabled Dataspace architectural designs and prototypes to develop the Datspace ecosystem in regional or local contexts have been demonstrated in Paper V and Paper VI, respectively.

4. **Evaluation:** Proposed artifacts representing different models and architecture in this PhD research have been evaluated through relevant real-world use cases and prototypes. This includes the Blockchain demonstrator for wind turbine supply chain traceability of events in Paper I, and the DENOS model presented in Paper IV, which was iteratively refined based on insights gained during earlier studies in Paper I till Paper III and based on relevant prototype testing and stakeholder feedback. Similarly, local context-aware Dataspace in Paper V developed based on the knowledge of all earlier papers but mainly considering the requirements that emerged in Paper II and its evaluation followed by building microservices (Kubernetes-based) enabled Dataspace demonstrator based on ALC approach and validating it for wind turbine use-case presented in Paper I. Similarly, IDS-based Dataspace in Paper VI has been developed based on the objectives of Paper V and validated in controlled lab environments to measure feasibility success.

5. **Communication of Results:** This PhD research addressed multiple dimensions: security, trust, digital transformation, theoretical extensions, cross-domain resource management, and heterogeneity harmonization. The

overall research context shows that applying IoT, AI, DLT, semantic methods, and distributed architectural frameworks enables a holistic approach to solving complex industry challenges. Apart from that, each paper systematically communicates the problem, the designed solution, its implementation, and its evaluation, contributing to the broader knowledge base.

To summarize, this PhD research explored and investigated the role of distributed architectures and the convergence of technologies at Edge in different use case scenarios in diverse industrial settings. This leads to the advancement of theoretical knowledge around CED continuum with data life cycle extensions and the development of edge-driven Dataspace, reflecting the method of developing a value chain for CDRI requirements. Thereby designing and delivering actionable solutions for Industry 4.0 and beyond and laying a strong foundation for future innovation in distributed edge-driven value-chain ecosystems. This has been done by fulfilling the five defined research objectives (RO1–RO5), each corresponding to its SRQ defined in section 1.3, in the following manner:

- **RO1** focused on establishing the role of IDEAL technologies and distributed architectures through an exhaustive literature review, prototyped solutions, and case studies. All papers, i.e., from Paper I to Paper VI fulfill this objective. It is observed that different technologies have their own strengths and weaknesses. However, if they combine, they can fill the gaps or limitations of each other. For example, IoT brings real-time data sensing and processing capabilities; Edge provides the required distributed infrastructure to process data close to the data source and save on latency and bandwidth fronts; AI brings in the analytical abilities, and DLT/Blockchain brings the distributed security capabilities that enable digital traceability, transparency, and trusted ecosystem in a multistakeholder environment. Therefore, IDEAL technologies are particularly valuable for processing data across distributed architectures, which is crucial for optimizing business value chains in Industry 4.0 and beyond ecosystems.

- RO2 explored distributed, interoperable, integrated, traceable, and immutable secure Edge operations for heterogeneous devices or data sources, cross-domain services, and infrastructural resources in a multi-stakeholder environment. Papers Paper I to Paper VI contribute to this objective. It is observed that events or functional flows in Industry 4.0 are often associated with distributed architectural design. For example, the customer or production units are distributed, as seen in Paper II, or in the wind industry use case of Paper I, where distributed data events in the supply chain are common. Therefore, when distributed events, functional flows, or architectural systems come into play, interoperability, digital traceability, trust, and transparency become essential requirements for multi-stakeholder environments. The major blocker observed is the issue of heterogeneity and dealing with it through the Integrated Harmonization and CDRI-driven Dataspace approach is discussed in Paper III to Paper VI.

- RO3 investigates and extends the theoretical knowledge of the CED computing paradigm through a systematic literature review and insights from earlier ROs. Consequently DENOS model emerged that offers a semantic processing model that provides dynamic implementation contexts tailored to the diverse use case requirements in Industry 4.0. It is also observed that a deadlock occurs when Edge is perceived as a device with limited computing capabilities. However, this perception is outdated due to significant advancements at the technology level. Even if Edge is traditionally viewed as a device, why not consider on-premises computing as Edge rather than just a device? This perspective emerged during the research exploration of Paper II, where label printing manufacturers were observed to have machines and systems operating in the same building, with many peers and customers working nearby. This scenario stimulates the need for Edge-to-Edge direct communication or data exchange, where Edge is perceived as an organizational entity. This approach foresees the possibility of sharing resources across organizational boundaries but with governance control in place, constituting the cross-domain concept and relevant value-chain development in these CDRI scenarios. Papers Paper II to Paper VI primarily contribute to this objective.

- RO4 focused on designing decentralized frameworks to achieve CDDI and data sovereignty in cross-domain contexts. It is well established through this objective that the Dataspace concept is the method by which CDDI and data sovereignty can be achieved. However, in this PhD research, Dataspace over Edge has been perceived to fulfill the CDRI requirements and is backed by semantic processing or modeling and IDEAL technological capabilities. Paper V to Paper VI primarily demonstrates the functional background of this objective.

- RO5: Validating the overall research context set by the main RQ and its impact on real-world industrial use cases. Paper I and Paper II provide real-world industrial use cases and relevant practical insights for validation of the relevant artifacts themselves and in other papers too. Here, the industrial use case provided by Paper I is extended with different perspectives and prototypes in multiple papers, starting from Paper IV to Paper VI.

9.2 Key Contributions

Each of the six key papers contributes unique insights into technological challenges and opportunities for achieving the research objectives for the main RQ. The overall contributions for this research work have been shown in Figure 9.1. However, some significant contributions have been highlighted below :

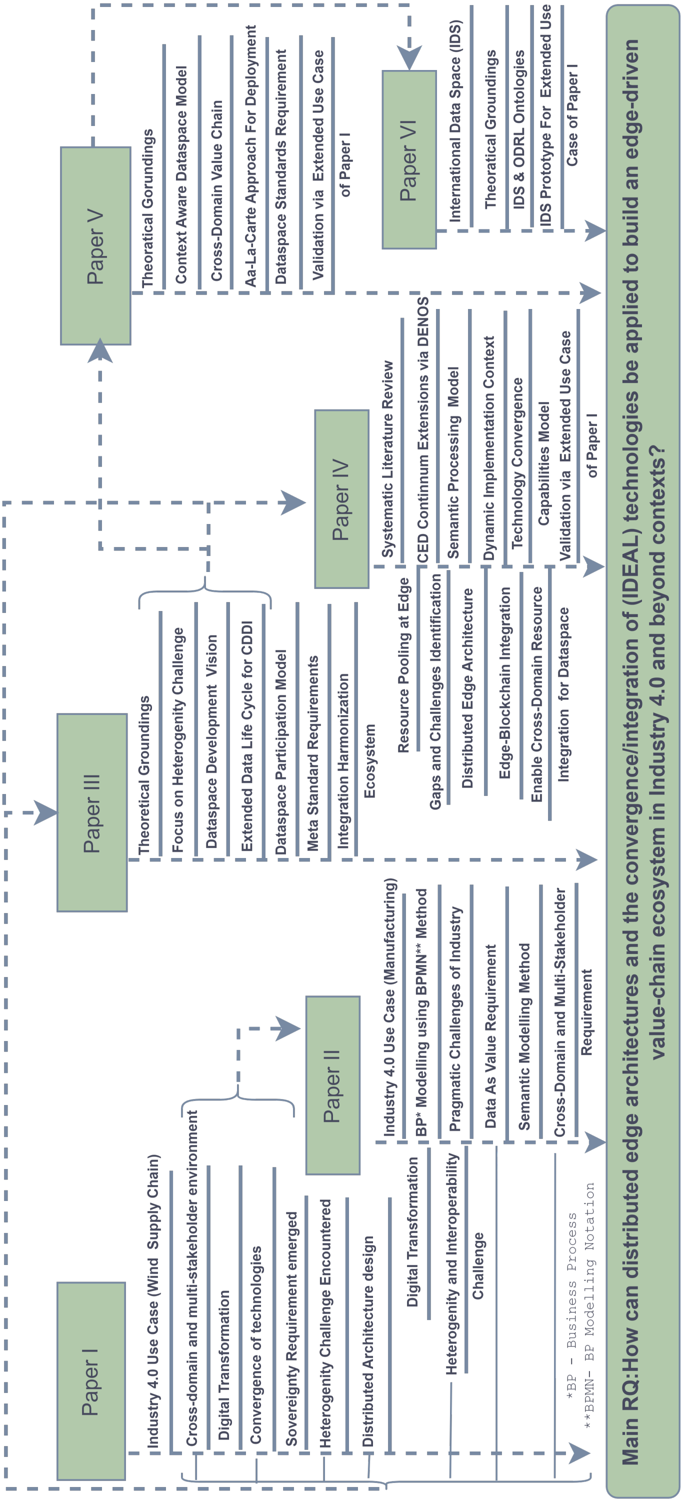


Figure 9.1: PhD Key Contributions

- **Contributions to Industry 4.0 & Technological Convergence Impact (Paper I)** The first study explored the application of blockchain technology in enhancing the economy of scale within the wind energy supply chain, focusing on commodity components such as bolts and fasteners. The study revealed significant challenges in achieving digital traceability and maintaining trust and transparency across the supply chain. By integrating blockchain technology, this research demonstrated that the visibility and traceability of these components could be improved throughout the entire lifecycle of wind turbines. The study developed a working demo to validate the feasibility of blockchain in a controlled lab environment using IoT and edge computing to host the relevant services and process the necessary actions near the data source, i.e., the turbine site. Thereby addressing issues of digital traceability and supporting sustainability, and proactive maintenance for operational efficiency. The findings underscored that Blockchain, IoT, and Edge-enabled solutions could transform traditional supply chains into more dynamic, transparent, and sustainable systems, offering a foundational model for applying similar approaches across other industrial domains.

  These contributions demonstrate how Blockchain (DLT), IoT, and Edge, as part of the IDEAL technology convergence, enable not only operational efficiencies through digital transformation but also foster trust among stakeholders, which is critical in wind turbine-related commodity supply-chain events in Industry 4.0 ecosystems.

- **Industry 4.0 - Enabling Digital Transformation in SMEs (Paper II)** The second study addressed the challenges faced by a small and medium enterprise (SME) in Denmark that specializes in industrial label printing. Operating with substantial printing operations and outdated installed machinery, the SME struggled to integrate advanced technologies due to constraints such as capital investment, resource limitations, and affordability. Therefore, an in-depth analysis of the SME's operational processes was conducted wherein bottlenecks and data point augmentation locations were identified that can support productivity, error management, communication, waste reduction, and compliance-related issues. A semantic process and system model were proposed to address these challenges, mainly focusing on error management in printing operations. The proposed solution incorporated open standards, protocols, and a low-cost open-source tools ecosystem to ensure affordability and practicality. The study demonstrated that even resource-constrained businesses could achieve digital transformation through targeted interventions enabled by distributed edge technologies.

  This work highlights the potential of distributed architectures and the need for cross-domain data integration requirements to empower SMEs to innovate their business models by adopting Industry 4.0 technologies-driven digital transformation and improving their value chains.

- **Highlighting Need For Meta Standards and Participatory Motivation For Dataspace Ecosystems (Paper III)**

  The third study focused on the complexities of cross-organizational data sharing in Industry 5.0 contexts. As organizations open their resources for collaboration, they encounter diverse standards and protocols that hinder seamless integration. This study proposed a meta-standard approach to harmonizing dataspace integration through an Edge-driven ecosystem, addressing the "Why," "What," and "How" aspects of enabling data sharing. Following the "Why" part, it presents a system model (i.e., participatory motivation model) that guides the interested organization or entities in determining mutual gain if they participate in the Datsapce-driven value chain development. Afterward, how they can join in building such an Edge-driven Dataspace ecosystem through a harmonized integration ecosystem was presented to mitigate the challenges posed by standards-induced heterogeneity, paving the way for building integrated value chains among industries. The findings emphasize that achieving such harmonization requires collaborative efforts in standardization and the development of clear guidelines for cross-business data sharing.

  This work contributes to a broader understanding of on designing and implementing cross-business value-chain systems that can leverage distributed architectures while addressing the intricacies of cross-domain integration and heterogeneity.

- **Advancing Edge Focusing Cross-Domain Value Chains (Paper IV)**
  The fourth study explored the theoretical groundings of CED continuum, traditionally encompassing Cloud, Edge, and Device layers, that can be reimagined as an ecosystem of interconnected domains with multiple stakeholders at each layer. This perspective highlights the classical challenges of heterogeneity, interoperability, distributed security, trust, and resource management, compounded by modern-day requirements for cross-domain integration. Addressing these complexities demands an evolution of the CED continuum, augmented by the combined strengths of IDEAL technologies. Therefore, a systematic literature review is conducted on these aspects, and many challenges and research gaps are identified, which are already mentioned in Paper IV under section 6.2.4. To address these challenges and considering cross-domain value chain requirements (i.e., Dataspace), the DENOS (Distributed Edge Network Operations oriented Semantic) architectural framework was designed, which also considered insights from earlier studies. DENOS extends the CED continuum by introducing three additional layers (namely Semantics, Convergence, and Dataspace Integration) to address both technical and business challenges in cross-domain or multi-stakeholder environments. It achieves this through semantic modeling and technological convergence, functioning on five core principles: *Discovery* of resources, and defining the relevant semantic contexts (Processing, Service, and Data), Execution of defined contexts leveraging technological services, Negotiation of resources at technical

and business levels (i.e., Dataspace), Optimized operations with resource pooling and proximity and Security for cross-domain operations covering distributed trust, traceability, identity access management, integrity and confidentiality. DENOS can dynamically fulfill use-case-specific requirements by modeling Processing Context, Service Context, and Data Context, facilitating value chain creation at technical (Infrastructure, Data, and Services) and business (i.e., Dataspace) levels. In addition, it presents the methodology to integrate a private Blockchain network at the Edge by addressing the known blockchain limitations.

By converging digital technologies and considering industrial ecosystems, DENOS can transform multi-stakeholder operations, address the challenges of legacy systems (distributed security, resource pooling, dynamic implementation, cross-domain integration, trust, etc.), and foster cross-domain collaboration. Its federation capabilities and dynamic adaptability make it particularly valuable for industries like wind energy and manufacturing, enabling secure, transparent, and efficient workflows aligned with Industry 4.0 goals. Therefore, this framework contributes towards building scalable, cost-effective solutions for enhancing operational productivity, cross-domain data sharing, and trust in diverse industrial ecosystems.

- **Context-Aware Dataspace Model for Industrial Ecosystems (Paper V)**

  The fifth study expanded on the concept of cross-domain value-chain integration by developing a context-aware Dataspace model with semantic capabilities. This study identifies the requirements for Dataspace, and afterward, a Datspace model is designed and prototyped in a lab environment to address data integration, reusability, and governance challenges. The proposed model leverages Edge resources (microservices-based) and an A-La-Carte approach (for developing artifact packages of services) to create a context-aware Datspace ecosystem to support the efficiency of cross-domain data management operations. In addition, it also demonstrates how localized or regional dataspace ecosystems could be realized through minimal investments, providing a scalable solution for small-scale deployments that will be particularly useful for SMEs.

  This work contributes to the growing knowledge base on how Edge-enabled Dataspace ecosystems in regional contexts can support value chain development in Industry 5.0, highlighting the importance of semantic adaptation and smart governance. The findings are particularly relevant for advancing upcoming technologies such as 6G in the future internet landscape.

- **Standards Compliant Dataspaces for Cross-Domain Data Integration (Paper VI)**

  The sixth study focused on the European Commission's initiative to develop Common Data Spaces (CDS) for cross-domain data integration. This research addressed the complexities of Edge-driven data sharing in the wind turbine industry by leveraging the IDS (International Data Spaces) standards, ontologies, and connector framework. The study proposed and evaluated a microservices-based architecture for deploying IDS-based dataspace ecosystems. The platform integrates components such as identity management, metadata brokers, and data source services, enabling secure and standardized data sharing across domains while managing related data sovereignty.

This research provides a practical pathway for adopting IDS specifications across industries in the local context, showcasing how IDS initiatives and standards can be leveraged to build distributed, scalable, sovereign, secure, and interoperable industrial ecosystems.

9.3 Key Take Aways

This section presents novel research contributions that have appeared for the first time in the literature and can be useful for industrial practitioners. These are given as follows:

1. **Customized Edge-Driven Dataspace:** Paper V presents a context-aware Dataspace model, which is perceived to be developed following the Edge layer. Here, the Edge has been perceived as an organizational Edge, providing the required infrastructure, services, and data sources to build an Edge-driven ecosystem. This will open the pathways for industries to share data with other industries or partners in their value chain (as shown by the use case in Paper II) using such a customized Dataspace ecosystem.

2. **IDS Compliant Local Dataspace Development:** Paper VI presents the method to develop IDS standards compliant Edge-driven Dataspace ecosystem in regional or local context. This will enable industries to develop value chains based on cross-domain data integration in a standard, secure, and sovereign manner.

3. **Standards Induced Heterogeneity**: Throughout different studies, we identified that heterogeneity is a significant challenge in the traditional CED continuum and the IoT Data ecosystem. Heteregoenity can exist at multiple levels, i.e., at systems, service, data, interface, or organization levels. Traditionally, using standards is the way to resolve this challenge. However, it is observed in Paper III that it is still a major challenge despite the use of standards. This is due to the development and use of diverse standards to perform the same task in different domains of operations. This develops a chicken-egg problem wherein heterogeneity is induced due to using diverse standards to perform the same task in different domains. Therefore, this research highlights the need for Metastandards and presents a related harmonization framework and initial guidelines to address the identified problem.

4. **Extended Data Life Cycle (DLC) for CDDI**: In Paper III, while investigating the cross-domain data integration, it is found that there is a research gap in lacking the data life cycle coverage over cross-domain data operations, especially during the data is being shared. Therefore, DLC has been extended to include the sharing and reusability phases and emphasize the need for further research and investigation in this direction.

5. **Edge-enabled CDDI through Dataspace area has been opened up for research:** It is observed during different studies of this research that much of the existing research is conducted in silos—for instance, focusing exclusively on either edge computing or Dataspace due to their origins in distinct domains. This fragmented approach limits the realization of their full potential, which lies in their convergence and integration. Therefore, this PhD research contributes by opening CDDI areas at the Edge and leveraging Dataspace concepts. This has been shown collectively in Paper IV, Paper V and Paper VI.

6. **Bridging Business and Technical Perspectives Through Semantic Modelling**: Semantic modeling is not a new concept, but in this research this has been leveraged as a tool in the architectural design to present or reflect both business and technical perspectives simultaneously, as shown in Paper II, Paper IV , Paper V and Paper VI. This approach is crucial for understanding diverse viewpoints and their interconnections. Ultimately, it serves industrial requirements by enabling the seamless mapping of business processes to technical or digital specifications. This accelerates digital transformation without compromising the balance between business and technical priorities.

7. **Participatory Motivation Model for Building a CDDI Ecosystem:** Participatory motivation is essential for developing a CDDI ecosystem using Dataspace methods, which are crucial for the sustainability of industries in Industry 5.0 and beyond. This is presented in Paper III wherein a system model for participatory motivation has been explained. This also emphasizes the need for a shift in mindset in industrial stakeholders that they should focus more on collaborative and win-win business models rather than competition.

8. **Dynamic Implementation Context:** During the research, it is observed that industries comprise diverse domains and use cases where the same set of technologies can be applied in varied contexts. This diversity necessitates a model capable of offering a dynamic implementation framework. *DENOS* is one such model, as presented in Paper IV that holds the potential to make a significant impact through dynamic implementation contextual capabilities to accommodate industries with varying technological and use-case requirements.

9. **Redefining the Edge with Expanding Possibilities:** Traditionally, the Edge is viewed as a device with constrained resources, raising the

question: how can it host resource-intensive technologies? This perspective limits the evolution of edge computing. While much research found during the literature review focuses on optimizing Edge capabilities, this PhD work explores a broader perspective—extending the horizon of edge computing. Thereby uncovering opportunities like cross-domain integration and Edge-driven value chains by perceiving the Edge more abstractly, such as an organizational or factory Edge. This approach enables the development of CDRI, unlocking pathways for data, services, and system interoperability across domains. The refined Edge concept has been presented and positioned in Paper II to Paper VI.

9.4 Broader Implications and Reflections

The contributions from this PhD research demonstrate the immense potential of distributed edge architectures and IDEAL technologies in addressing the multifaceted challenges of Industry 4.0 and beyond. The findings reveal that:

- **Distributed Architectures** can provide efficient infrastructure, services, systems, and data integration management, fostering collaboration and innovation across industries to develop win-win business models by sharing and reusing cross-domain resources through value chain development, wherein participatory motivation by stakeholders plays a significant role.
- **Convergence of Technologies** like AI, IoT, and DLT at the Edge can offer transformative capabilities in addressing intelligence of operations for optimizations, low latency, resources efficiency, real-time events monitoring, digital trust, traceability, security, and scalability challenges in modern industrial value chains which are often found to be distributed in nature.
- **Semantic Models and Context-Aware Systems** can be the game changer in addressing interoperability due to heterogeneity in data sources, systems, services, and their interfaces for creating new cross-domain operations and collaboration opportunities.
- **Cross-Domain Edge-Enabled Dataspace Ecosystems**: Leveraging a relaxed Edge perspective and the convergence of IDEAL technologies offers practical solutions to traditional business challenges. These include addressing struggles with digitalization due to constrained computing resources and reluctance to embrace collaborative approaches. Such ecosystems facilitate cross-domain resource integration by enabling distributed infrastructural resource pooling, digital trust, transparency, traceability, and sustainability. Within this framework, data, services, and systems can function as valuable assets across related supply chain life cycles and industrial processes.

This PhD dissertation emphasizes the critical role of distributed edge architectures and the convergence of technologies in bridging the gap between technological innovation and practical business challenges observed at the SME level. By enabling cross-domain data integration methods, these architectures pave the way for more resilient, sustainable, and dynamic ecosystems, aligning with the broader objectives of Industry 5.0 and the evolving data-driven digital economy.

9.5 Limitations and Future Directions

While this research has laid a foundation for building cross-domain Edge-driven value-chain ecosystems, several areas remain for further exploration:

- **Prototyping and Real-World Deployments:** Though multiple models have been prototyped, such as DENOS, context-aware Dataspace, and IDS-specific Datsapce models in the wind turbine use case. However, expanding the proposed models and frameworks to include more use cases and large-scale, real-world deployments across multiple industries will be of significant interest.

- **AI-Driven Support:** Due to the time limitations, the full potential of AI-based capabilities in CDDI has not been fully explored yet. Therefore, incorporating advanced machine learning techniques will progress this research work further to automate diverse tasks in CDDI, such as supporting the automated data ingestion phase by harmonizing data schemas or ontologies to make them useful for SMEs to integrate quickly and cost-effectively with Dataspace ecosystems.

- **Standardization Efforts:** It is observed during the research that there is a need for Meta Standards in general and as well as for Edge-enabled CDRI operations. Therefore, efforts in this direction are needed to develop Meta Standards while considering aspects of dynamic implementation context for CDRI operations to offer seamless Dataspace integration and address the relevant challenges.

- **Sustainability and Circular Economy:** It will be interesting to explore how distributed architectures and the proposed IDS-based Dataspace model in this research can support circular economy principles, enhancing sustainability and coordination among industrial entities and their processes.

9.6 Conclusion

To conclude this PhD research, there is an urgent need to address the limitations and challenges of Edge-driven architectures to enable cross-domain resource integration and support the development of industry-wide value chains. These limitations include constrained Edge capabilities for technology convergence, while challenges encompass distributed architectures, multiple levels of heterogeneity, diverse use-case requirements, distributed security, as well as a lack of trust and transparency in multi-stakeholder environments and value-chain silos. To address these issues, there is a critical need for distributed architectural extensions that integrate robust security mechanisms, harmonize data heterogeneity, and establish sovereign, governed, and trusted frameworks for cross-domain data integration (CDDI) operations at the Edge. In alignment with this vision, this research investigates how extending traditional distributed edge paradigms and leveraging the convergence capabilities of IDEAL technologies can provide practical solutions to real-world industrial challenges.

Furthermore, this PhD work introduces several key contributions, including innovative methods for cross-domain resource integration, ensuring data sovereignty, enabling digital traceability, seamless data sharing, and resource chaining through Dataspace enablement at the Edge. These advancements pave the way for creating integrated cross-domain ecosystems across industries, empowering actors to engage at the Edge directly. This transformative approach has the potential to revolutionize business operations and foster the emergence of novel value-chain constellations, thereby addressing the demands of Industry 4.0 and beyond.

In a nutshell, this PhD research through six papers contributes significantly to the knowledge base on distributed edge architectures and convergence of IDEAL technologies, providing actionable insights for building Edge-driven value-chain ecosystems. These advancements hold the potential to transform industries, foster innovation, and drive sustainable growth in an increasingly data-driven world in the future.

Appendices

Appendix A

Awards and Recognitions

AIOTI Award For Developing Solution in SME Category For Paper I, Brussels Belgium, Sep. 2023

Best Research Paper Award For Paper V in Internet 2024 Conference, Athens, Greece, March 2024

IARIA

International Academy, Research, and Industry Association

Best Paper Award

Data-Driven IoT Ecosystem for Cross Business Growth: An Inspiration Future Internet Model with Dataspace at the Edge

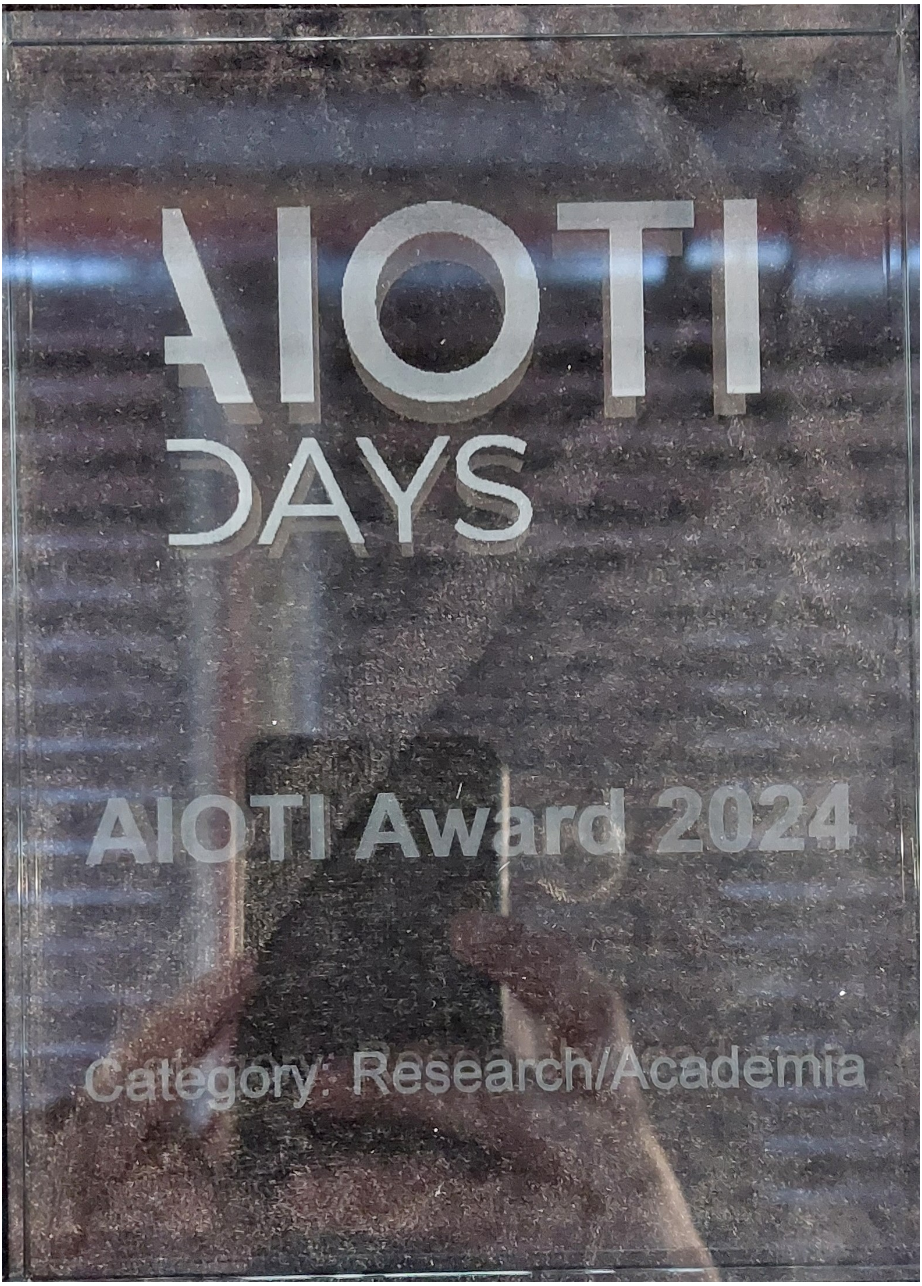
By

Parwinder Singh, Nidhi Nidhi, Michail J. Beliatis, Mirko Presser

Presented during INTERNET 2024, The Sixteenth International Conference on Evolving Internet, held in Athens, Greece during March 10 - 14, 2024

IARIA Board

AIOTI Award Winner in Research Category For Paper IV, Brussels Belgium, Sep. 2024



Appendix B

Co-Author Statements/Declaration

B.1 Information Interpretation

This section presents the coauthor statements, outlining their contributions across various aspects of the research. The numerical values represent the level of involvement in different areas. If a cell in the table is marked as N/A, it should be interpreted as "Not Applicable" for the corresponding author in that specific category. Tables B.1 and B.2 provide the detailed contribution descriptions and the scale used to quantify involvement, respectively.

Table B.1: Coauthor Contributions

Label	**Contribution Description**
A	Research idea: Identifying, developing, specifying, and formulating overarching research questions and aim.
B	Theory: Organizing theoretical perspectives, developing arguments and hypotheses, specifying theoretical model.
C	Research design: Developing and planning design for test or exploration of the research question.
D	Data collection: Preparing and organizing data collection, data collection, preparing data for analysis and storage.
E	Data analysis: Application of empirical techniques to analyze or synthesize study data including providing support for interpretations such as visualizations, etc.
F	Writing: Drafting and revising manuscript presenting the research idea and results.

Table B.2: Contribution Scale

Scale	**Meaning**
4	Has essentially delivered this part.
3	Major contribution.
2	Equal contribution.
1	Minor contribution.
0	No contribution.
N/A	Not Applicable.

The table B.3 presents my contributions to various papers, highlighting my role as the main author and my level of involvement in different aspects of each study.

Table B.3: My Contributions by Paper Title

Paper Title	**A**	**B**	**C**	**D**	**E**	**F**
Blockchain for Economy of Scale in Wind Industry: A Demo Case	4	4	4	2	4	4
Digital Dataspace and Business Ecosystem Growth for Industrial Roll-to-Roll Label Printing Manufacturing: A Case Study	4	4	4	4	4	4
Meta Standard Requirements for Harmonizing Dataspace Integration at the Edge	4	4	4	4	4	4
Enabling edge-driven Dataspace integration through convergence of distributed technologies	4	4	4	4	4	4
Data-Driven IoT Ecosystem for Cross Business Growth: An Inspiration Future Internet Model with Dataspace at the Edge	4	4	4	4	4	4
Navigating the International Data Space to Build Edge-Driven Cross-Domain Dataspace Ecosystem	4	4	4	4	4	4

Co-Author Declaration For Paper I

SCHOOL OF BUSINESS AND SOCIAL SCIENCES
AARHUS UNIVERSITY

Declaration of co-authorship[1] Date:

This declaration concerns the following article/manuscript:

Title:	**Blockchain for Economy of Scale in Wind Industry: A Demo Case**
Authors:	Parwinder Singh, Kristoffer Holm, Michail J. Beliatis, Andrei Ionita, Mirko Presser, Prinz Wolfgang & René C. Goduscheit

The article/manuscript is:

☒ Published, state full reference: Singh, P. et al. (2022). Blockchain for Economy of Scale in Wind Industry: A Demo Case. In: González-Vidal, A., Mohamed Abdelgawad, A., Sabir, E., Ziegler, S., Ladid, L. (eds) Internet of Things. GIoTS 2022. Lecture Notes in Computer Science, vol 13533. Springer, Cham. https://doi.org/10.1007/978-3-031-20936-9_14
☐ Accepted, state journal:
☐ Invited for revision, state journal:
☐ Submitted
☐ In preparation

Date of the current version of the manuscript, if not published or accepted:

Please fill out Table 1 regarding contribution to the manuscript for all authors. The respective author has contributed to the elements:

A. Research idea: Identifying, developing, specifying, and formulating the overarching research question and aim.
B. Theory: Organizing theoretical perspectives, developing arguments and hypotheses, specifying theoretical model.
C. Research design: Developing and planning design for test or exploration of the research question.
D. Data collection: Preparing and organizing data collection, data collection, preparing data for analysis and storage.
E. Data analysis: Application of empirical techniques to analyze or synthesize study data including providing support for interpretations such as visualizations etc.
F. Writing: Drafting and revising manuscript presenting the research idea and results

of this article/manuscript as follows:

4 Has essentially delivered this part.
3 Major contribution
2 Equal contribution
1 Minor contribution
0 Did not contribute to this part.

[1] Attribution of authorship should be based on criteria a-d adopted from the Vancouver guidelines (see also rules and guidelines from Aarhus University) and all individuals who meet these criteria should be recognized as authors. The co-author has contributed:

a) to the conception or design of the work, or the acquisition, analysis, or interpretation of data for the work, *and*
b) drafting the work or revising it critically for important intellectual content, *and*
c) to the final approval of the version to be published, *and*
d) agrees to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Co-Author Declaration For Paper I (continued)

AARHUS BSS SCHOOL OF BUSINESS AND SOCIAL SCIENCES
AARHUS UNIVERSITY

N/A Not relevant or not applicable

Table 1. Individual contributions and signature of each co-author[1]

Author	Extent of contribution (4-0) per element (A.-F.)						Signature of the author[2]
	A. Research Idea	B. Theory	C. Research Design	D. Data Collection	E. Data Analysis	F. Writing	
Parwinder Singh	4	4	4	2	4	4	Parwinder Singh
Kristoffer Holm	1	1	1	2		1	Kristoffer Holm
Michail J. Beliatis	1	1	1	1	1	1	
Andrei Ionita	1	1	1	1	1	1	
Mirko Presser	1	1	1	N/A	1	1	
Prinz Wolfgang	1	N/A	1	N/A	1		
René C. Goduscheit	1	N/A	1	N/A	1	1	

[1]More rows can be added for additional authors.
[2]All authors must confirm the declaration either by signature or email.

If relevant, you may add more information on the work and collaboration such as open science practices or more detailed specifications of authors' contributions here:

Co-Author Declaration For Paper II

SCHOOL OF BUSINESS AND SOCIAL SCIENCES
AARHUS UNIVERSITY

Declaration of co-authorship[1]

Date: 17-10-2024

This declaration concerns the following article/manuscript:

Title:	**Digital Dataspace and Business Ecosystem Growth for Industrial Roll-to-Roll Label Printing Manufacturing: A Case Study**
Authors:	Parwinder Singh, Nidhi, Justina Karpavice, Michail Beliatis

The article/manuscript is:

☒ Published, state full reference: Singh, Parwinder, et al. "Digital dataspace and business ecosystem growth for industrial roll-to-roll label printing manufacturing: A case study." *SENSORCOMM 2023: The Seventeenth International Conference on Sensor Technologies and Applications*. IARIA, 2023.
☐ Accepted, state journal:
☐ Invited for revision, state journal:
☐ Submitted
☐ In preparation

Date of the current version of the manuscript, if not published or accepted:

Please fill out Table 1 regarding contribution to the manuscript for all authors. The respective author has contributed to the elements:

A. Research idea: Identifying, developing, specifying, and formulating the overarching research question and aim.
B. Theory: Organizing theoretical perspectives, developing arguments and hypotheses, specifying theoretical model.
C. Research design: Developing and planning design for test or exploration of the research question.
D. Data collection: Preparing and organizing data collection, data collection, preparing data for analysis and storage.
E. Data analysis: Application of empirical techniques to analyze or synthesize study data including providing support for interpretations such as visualizations etc.
F. Writing: Drafting and revising manuscript presenting the research idea and results

of this article/manuscript as follows:

4 Has essentially delivered this part.
3 Major contribution
2 Equal contribution
1 Minor contribution
0 Did not contribute to this part.
N/A Not relevant or not applicable

[1] Attribution of authorship should be based on criteria a-d adopted from the Vancouver guidelines (see also rules and guidelines from Aarhus University) and all individuals who meet these criteria should be recognized as authors. The co-author has contributed:

a) to the conception or design of the work, or the acquisition, analysis, or interpretation of data for the work, *and*
b) drafting the work or revising it critically for important intellectual content, *and*
c) to the final approval of the version to be published, *and*
d) agrees to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Co-Author Declaration For Paper II (continued)

AARHUS BSS SCHOOL OF BUSINESS AND SOCIAL SCIENCES AARHUS UNIVERSITY

Table 1. Individual contributions and signature of each co-author[1]

Author	Extent of contribution (4-0) per element (A.-F.)						Signature of the author[2]
	A. Research Idea	B. Theory	C. Research Design	D. Data Collection	E. Data Analysis	F. Writing	
Parwinder Singh	4	4	4	4	4	4	Parwinder Singh
Nidhi	N/A	1	1	1	1	1	Nidhi
Justina Karpavice	N/A	1	1	1	1	1	
Michail Beliatis	1	1	1	1	1	1	

[1]More rows can be added for additional authors.
[2]All authors must confirm the declaration either by signature or email.

If relevant, you may add more information on the work and collaboration such as open science practices or more detailed specifications of authors' contributions here:

Co-Author Declaration For Paper III

SCHOOL OF BUSINESS AND SOCIAL SCIENCES
AARHUS UNIVERSITY

Declaration of co-authorship[1] Date: 17-10-2024

This declaration concerns the following article/manuscript:

Title:	**Meta Standard Requirements for Harmonizing Dataspace Integration at the Edge**
Authors:	Parwinder Singh, Nidhi, Asim Ul Haq, Michail Beliatis

The article/manuscript is:

☒ Published, state full reference: Singh, Parwinder, Asim Ul Haq, and Michail Beliatis. "Meta Standard Requirements for Harmonizing Dataspace Integration at the Edge." *2023 IEEE Conference on Standards for Communications and Networking (CSCN)*. IEEE, 2023.
☐ Accepted, state journal:
☐ Invited for revision, state journal:
☐ Submitted
☐ In preparation

Date of the current version of the manuscript, if not published or accepted:

Please fill out Table 1 regarding contribution to the manuscript for all authors. The respective author has contributed to the elements:

A. Research idea: Identifying, developing, specifying, and formulating the overarching research question and aim.
B. Theory: Organizing theoretical perspectives, developing arguments and hypotheses, specifying theoretical model.
C. Research design: Developing and planning design for test or exploration of the research question.
D. Data collection: Preparing and organizing data collection, data collection, preparing data for analysis and storage.
E. Data analysis: Application of empirical techniques to analyze or synthesize study data including providing support for interpretations such as visualizations etc.
F. Writing: Drafting and revising manuscript presenting the research idea and results

of this article/manuscript as follows:

4 Has essentially delivered this part.
3 Major contribution
2 Equal contribution
1 Minor contribution
0 Did not contribute to this part.
N/A Not relevant or not applicable

[1] Attribution of authorship should be based on criteria a-d adopted from the Vancouver guidelines (see also rules and guidelines from Aarhus University) and all individuals who meet these criteria should be recognized as authors. The co-author has contributed:

a) to the conception or design of the work, or the acquisition, analysis, or interpretation of data for the work, *and*
b) drafting the work or revising it critically for important intellectual content, *and*
c) to the final approval of the version to be published, *and*
d) agrees to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Co-Author Declaration For Paper III (continued)

AARHUS BSS

SCHOOL OF BUSINESS AND SOCIAL SCIENCES
AARHUS UNIVERSITY

Table 1. Individual contributions and signature of each co-author[1]

Author	Extent of contribution (4-0) per element (A.-F.)						Signature of the author[2]
	A. Research Idea	B. Theory	C. Research Design	D. Data Collection	E. Data Analysis	F. Writing	
Parwinder Singh	4	4	4	4	4	4	Parwinder Singh
Nidhi	N/A	1	1	1	1	1	Nidhi
Asim Ul Haq	N/A	1	N/A	1	1	1	
Michail Beliatis	N/A	N/A	1	N/A	1	1	

[1]More rows can be added for additional authors.
[2]All authors must confirm the declaration either by signature or email.

If relevant, you may add more information on the work and collaboration such as open science practices or more detailed specifications of authors' contributions here:

Co-Author Declaration For Paper IV

SCHOOL OF BUSINESS AND SOCIAL SCIENCES
AARHUS UNIVERSITY

Declaration of co-authorship[1] Date: 17-10-2024

This declaration concerns the following article/manuscript:

Title:	**Enabling edge-driven Dataspace integration through convergence of distributed technologies**
Authors:	Parwinder Singh, Michail J. Beliatis, Mirko Presser

The article/manuscript is:

☒ Published, state full reference: Parwinder Singh, Michail J. Beliatis, Mirko Presser, Enabling edge-driven Dataspace integration through convergence of distributed technologies, Internet of Things, Volume 25, 2024, 101087, ISSN 2542-6605, doi: https://doi.org/10.1016/j.iot.2024.101087.

☐ Accepted, state journal:
☐ Invited for revision, state journal:
☐ Submitted
☐ In preparation

Date of the current version of the manuscript, if not published or accepted:

Please fill out Table 1 regarding contribution to the manuscript for all authors. The respective author has contributed to the elements:

A. Research idea: Identifying, developing, specifying, and formulating the overarching research question and aim.
B. Theory: Organizing theoretical perspectives, developing arguments and hypotheses, specifying theoretical model.
C. Research design: Developing and planning design for test or exploration of the research question.
D. Data collection: Preparing and organizing data collection, data collection, preparing data for analysis and storage.
E. Data analysis: Application of empirical techniques to analyze or synthesize study data including providing support for interpretations such as visualizations etc.
F. Writing: Drafting and revising manuscript presenting the research idea and results

of this article/manuscript as follows:

4 Has essentially delivered this part.
3 Major contribution
2 Equal contribution
1 Minor contribution
0 Did not contribute to this part.
N/A Not relevant or not applicable

[1] Attribution of authorship should be based on criteria a-d adopted from the Vancouver guidelines (see also rules and guidelines from Aarhus University) and all individuals who meet these criteria should be recognized as authors. The co-author has contributed:

a) to the conception or design of the work, or the acquisition, analysis, or interpretation of data for the work, *and*
b) drafting the work or revising it critically for important intellectual content, *and*
c) to the final approval of the version to be published, *and*
d) agrees to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Co-Author Declaration For Paper IV (continued)

AARHUS BSS
SCHOOL OF BUSINESS AND SOCIAL SCIENCES
AARHUS UNIVERSITY

Table 1. Individual contributions and signature of each co-author[1]

Author	Extent of contribution (4-0) per element (A.-F.)						Signature of the author[2]
	A. Research Idea	B. Theory	C. Research Design	D. Data Collection	E. Data Analysis	F. Writing	
Parwinder Singh	4	4	4	4	4	4	Parwinder Singh
Michail Beliatis	1	N/A	1	N/A	1	1	[illegible]
Mirko Presser	N/A	N/A	1	N/A	1	1	[illegible]

[1]More rows can be added for additional authors.
[2]All authors must confirm the declaration either by signature or email.

If relevant, you may add more information on the work and collaboration such as open science practices or more detailed specifications of authors' contributions here:

Co-Author Declaration For Paper V

SCHOOL OF BUSINESS AND SOCIAL SCIENCES
AARHUS UNIVERSITY

Declaration of co-authorship[1] Date: 17-10-2024

This declaration concerns the following article/manuscript:

Title:	**Data-Driven IoT Ecosystem for Cross Business Growth: An Inspiration Future Internet Model with Dataspace at the Edge**
Authors:	Parwinder Singh, Nidhi, Michail Beliatis, Mirko Presser

The article/manuscript is:

☒ Published, state full reference: Singh, Parwinder, Nidhi, Michail Beliatis, and Mirko Presser. "Data-Driven IoT Ecosystem for Cross Business Growth: An Inspiration Future Internet Model with Dataspace at the Edge." *INTERNET 2024:: The Sixteenth International Conference on Evolving Internet*. 2024.
☐ Accepted, state journal:
☐ Invited for revision, state journal:
☐ Submitted
☐ In preparation

Date of the current version of the manuscript, if not published or accepted:

Please fill out Table 1 regarding contribution to the manuscript for all authors. The respective author has contributed to the elements:

A. Research idea: Identifying, developing, specifying, and formulating the overarching research question and aim.
B. Theory: Organizing theoretical perspectives, developing arguments and hypotheses, specifying theoretical model.
C. Research design: Developing and planning design for test or exploration of the research question.
D. Data collection: Preparing and organizing data collection, data collection, preparing data for analysis and storage.
E. Data analysis: Application of empirical techniques to analyze or synthesize study data including providing support for interpretations such as visualizations etc.
F. Writing: Drafting and revising manuscript presenting the research idea and results

of this article/manuscript as follows:

4 Has essentially delivered this part.
3 Major contribution
2 Equal contribution
1 Minor contribution
0 Did not contribute to this part.
N/A Not relevant or not applicable

[1] Attribution of authorship should be based on criteria a-d adopted from the Vancouver guidelines (see also rules and guidelines from Aarhus University) and all individuals who meet these criteria should be recognized as authors. The co-author has contributed:

a) to the conception or design of the work, or the acquisition, analysis, or interpretation of data for the work, *and*
b) drafting the work or revising it critically for important intellectual content, *and*
c) to the final approval of the version to be published, *and*
d) agrees to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Co-Author Declaration For Paper V (continued)

SCHOOL OF BUSINESS AND SOCIAL SCIENCES
AARHUS UNIVERSITY

Table 1. Individual contributions and signature of each co-author[1]

Author	Extent of contribution (4-0) per element (A.-F.)						Signature of the author[2]
	A. Research Idea	B. Theory	C. Research Design	D. Data Collection	E. Data Analysis	F. Writing	
Parwinder Singh	4	4	4	4	4	4	Parwinder Singh
Nidhi	N/A	1	1	N/A	1	1	Nidhi
Michail Beliatis	N/A	1	1	N/A	1	1	MJB
Mirko Presser	N/A	N/A	1	N/A	1	1	

[1]More rows can be added for additional authors.
[2]All authors must confirm the declaration either by signature or email.

If relevant, you may add more information on the work and collaboration such as open science practices or more detailed specifications of authors' contributions here:

Co-Author Declaration For Paper VI

AARHUS BSS
SCHOOL OF BUSINESS AND SOCIAL SCIENCES
AARHUS UNIVERSITY

Declaration of co-authorship[1] Date: 17-10-2024

This declaration concerns the following article/manuscript:

Title:	**Navigating the International Data Space To Build Edge-Driven Cross-Domain Dataspace Ecosystem**
Authors:	Parwinder Singh, Nirvana Meratnia, Michail Beliatis, Mirko Presser

The article/manuscript is:

☒ Published, state full reference: Singh, Parwinder, et al. "Navigating the International Data Space to Build Edge-Driven Cross-Domain Dataspace Ecosystem." GIECS 2024, Belgium, Brussels, Springer Nature Switzerland AG, 2024. Lecture Notes in Computer Science (LNCS).
☐ Accepted, state journal:
☐ Invited for revision, state journal:
☐ Submitted
☐ In preparation

Date of the current version of the manuscript, if not published or accepted:

Please fill out Table 1 regarding contribution to the manuscript for all authors. The respective author has contributed to the elements:

A. Research idea: Identifying, developing, specifying, and formulating the overarching research question and aim.
B. Theory: Organizing theoretical perspectives, developing arguments and hypotheses, specifying theoretical model.
C. Research design: Developing and planning design for test or exploration of the research question.
D. Data collection: Preparing and organizing data collection, data collection, preparing data for analysis and storage.
E. Data analysis: Application of empirical techniques to analyze or synthesize study data including providing support for interpretations such as visualizations etc.
F. Writing: Drafting and revising manuscript presenting the research idea and results

of this article/manuscript as follows:

4 Has essentially delivered this part.
3 Major contribution
2 Equal contribution
1 Minor contribution
0 Did not contribute to this part.
N/A Not relevant or not applicable

[1] Attribution of authorship should be based on criteria a-d adopted from the Vancouver guidelines (see also rules and guidelines from Aarhus University) and all individuals who meet these criteria should be recognized as authors. The co-author has contributed:

a) to the conception or design of the work, or the acquisition, analysis, or interpretation of data for the work, *and*
b) drafting the work or revising it critically for important intellectual content, *and*
c) to the final approval of the version to be published, *and*
d) agrees to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Co-Author Declaration For Paper VI (continued)

AARHUS BSS
SCHOOL OF BUSINESS AND SOCIAL SCIENCES
AARHUS UNIVERSITY

Table 1. Individual contributions and signature of each co-author[1]

Author	Extent of contribution (4-0) per element (A.-F.)						Signature of the author[2]
	A. Research Idea	B. Theory	C. Research Design	D. Data Collection	E. Data Analysis	F. Writing	
Parwinder Singh	4	4	4	4	4	4	Parwinder Singh
Nirvana Meratnia	1	1	1	N/A	1	1	Nirvana Meratnia
Michail Beliatis	N/A	N/A	1	N/A	1	1	
Mirko Presser	N/A	N/A	1	N/A	1	1	

[1]More rows can be added for additional authors.
[2]All authors must confirm the declaration either by signature or email.

If relevant, you may add more information on the work and collaboration such as open science practices or more detailed specifications of authors' contributions here: