DESIGN OF ARCHITECTURE PRINCIPLES FOR A LOGISTICS FOCUSED, MULTI-SIDED DATA MARKETPLACE ARCHITECTURE

A **Master Thesis** submitted to **Delft University of Technology** in partial fulfilment of the requirements for the degree of

Master of Science

in

Management of Technology

Faculty of Technology, Policy and Management

by

Saga Kassa

Student Number: **4977114**

To be defended publicly on **30/06/2023**.

Graduation Committee

Chair	**Dr. A.M.G (Anneke) Zuiderwijk**
	Section of Information and Communication Technology
First Supervisor	**Dr. A.M.G (Anneke) Zuiderwijk**
	Section of Information and Communication Technology
Second Supervisor	**Dr. ir. C. (Carlos) Hernandez Ganan**
	Section of Organization and Governance

Preface

My educational journey at TU Delft comes to an end with the completion of this master's thesis. Since I commenced my studies in September 2019, I have had the privilege of meeting remarkable individuals, including fellow students and esteemed researchers, who have enriched my academic experience and contributed to my personal growth. This journey has also seen surprises and twists. Early in the studies, the pandemic started with no end in sight. Life changed as we knew, and it was difficult to adapt. I was also confronted with personal obstacles that impeded my progress. The thesis journey extended over a span of 2.5 years, and during this time, I met moments of self-doubt and setbacks. Yet, with perseverance to not accept defeat, I managed to overcome these obstacles, convincing myself of the feasibility of reaching the finish line.

This thesis focuses on the captivating domain of large-scale business-to-business data sharing in the logistics industry, specifically examining the design of architecture requirements and architecture principles for a logistics multi-sided data marketplace. In a world where data plays an increasingly critical role in driving decision-making and innovation, enabling open and secure access to logistics data can unlock countless opportunities. It appears, however, that while design is an inhibiting factor, we also have not yet managed to quantify the gains of these opportunities. Most discussions on this topic are conceptual or in early stages. Throughout my research, I therefore encountered various challenges that tested my spirit, such as navigating the diverse definitions encountered during the literature review. Conducting interviews has therefore been a significant learning experience which allowed me to piece the different concepts. I am therefore deeply indebted and grateful to all the interview participants in this study. Although I may have had difficulties during the interviews, analyzing the interviews provided an invaluable learning opportunity, enabling me to greatly broaden my comprehension of the subject matter.

I wish to give my deepest appreciation to my thesis supervisor, Anneke Zuiderwijk, whose unwavering support and guidance have been pivotal throughout this entire endeavor. Her invaluable feedback and mentorship have played a critical role in shaping the outcome of this work. I would also like to express my heartfelt gratitude to my second supervisor, Carlos Hernandez Ganan, and Hosea Ofe, for their availability and insightful contributions during our discussions. I can not be more grateful to have had such support. I also wish to extend my appreciation to my friends and family. I am forever indebted for their unwavering support. In particular, I would like to extend special thanks to my mother, Hafid and Kerim, who stood by me, helping me believe in myself and continuously motivate me to improve myself. Finally, I wish to acknowledge the motivational words of Les Brown, whose reminder that "it is not over until I win, it is not over until I get through" resonates profoundly with me. These words served as a constant source of inspiration.

With immense satisfaction and a sense of accomplishment, I present you this thesis. Please enjoy reading the report.

Saga Kassa, June 2023.

Executive Summary

In today's increasingly interconnected and digitized world, the logistics industry plays a critical role in global trade, enabling the seamless movement of goods across essential supply chains (Reinsel et al., 2020). The rapid evolution of digital technologies has led to an unprecedented proliferation of data generated from various aspects of logistics operations. However, despite the potential of this data to drive efficiency, cost reduction, and supply chain optimization, the logistics industry has been slow to embrace large-scale data sharing practices, resulting in a fragmented data ecosystem (PwC, 2019). Large-scale data sharing in the context of the logistics industry refers to the efficient exchange of information across various stakeholders, including shippers, carriers, freight forwarders, and other logistics service providers. By enabling access to a vast pool of data, logistics service providers can gain a more comprehensive understanding of the overall supply chain, leading to numerous benefits for the industry. Multi-sided data marketplaces (MSDM) are a promising concept that can play an important role in addressing this lack of data sharing on a large scale among businesses for any domain (Koutroumpis et al. (2020b)). They take the role of the neutral middleman by providing a platform to facilitate the data transactions, the ability to search for data sets and provide complementary services. Despite their perceived benefits, however, many initiatives remain in early stages (Spiekermann, 2019) and little research is published that demonstrates the usability of the concept for the logistics industry. As no open data sharing platform currently exists for the logistics industry (Bastiaansen et al., 2020), conceptualizing a multi-sided data marketplace architecture which meets the domain-specific multi-stakeholder context may contribute with novel insights.

A crucial aspect of the design of a multi-sided data marketplace architecture for logistics data is the development of architecture principles. These principles serve as the foundation for designing a robust and effective solution that addresses the unique challenges and requirements of the logistics industry. Incorporating architecture principles as part of the design process is essential for at least three reasons. First, architecture principles provide high-level guidance for making design decisions throughout the development process. They help maintain a shared understanding among stakeholders, ensuring that the resulting solution aligns with the overarching goals and objectives of the project. Second, while generic architecture principles exist, they . Therefore, developing domain-specific architecture principles ensures that the design of the multi-sided data marketplace takes into account the unique challenges and opportunities present in the logistics industry. Third, by adhering to a set of architecture principles, the design process becomes consistent and coherent. For these reasons, this master thesis aims to address the lack of data sharing in the logistics industry by proposing a set of domain-specific architecture principles to guide the design and development of multi-sided data marketplaces, tailored to the unique needs and requirements of the logistics industry.

This study employed the Design Science Research (DSR) methodology to design the domain-specific architecture principles. DSR is a method that establishes and operationalizes research when the desired goal is an artifact or a recommendation (Dresch et al., 2015). The proposed architecture principles are grounded in a set of domain-specific architecture requirements to ensure that the resulting principles address the unique needs and challenges of the logistics domain. These architecture requirements encompass key aspects such as data security and access control, data quality and integration, real-time data processing and analysis, transparency and fairness, user-centric design and scalability, legal and regulatory compliance, and flexibility and adaptability through open standards and interoperability. To develop the domain-specific architecture requirements, a series of seven interviews were conducted with domain experts in the logistics industry. These interviews helped identify user concerns and requirements, which were then formulated into domain-specific architecture requirements. These requirements were subsequently validated against existing research. Afterwards, the formulated domain-specific requirements were grouped into categories based on common themes, and overarching concerns were generalized to formulate high-level, actionable architecture principles. Although the current study is based on a small sample of interviewees, the findings suggest the importance of:

1. Prioritizing data security and access control to maintain trust and encourage participation among stakeholders by protecting sensitive data and allowing granular control over sharing.
2. Focusing on data quality and integration by supporting diverse sources, standardization, and contextualization to facilitate seamless data analysis and utilization.
3. Enabling real-time data processing and analysis to empower users with timely insights and informed decision-making.
4. Fostering transparency, fairness, and collaboration through transparent policies, pricing mechanisms, and impartial governance structures to promote trust and equal opportunities.

5. Designing user-centric, cost-effective, and scalable solutions that adapt to stakeholders' diverse needs and resources, encouraging widespread adoption and usage.
6. Complying with legal and regulatory requirements to ensure responsible data management practices and maintain stakeholder trust.
7. Promoting flexibility and adaptability by embracing open standards and interoperability, enabling seamless integration and easy transitions between platforms, ultimately reducing vendor lock-in.

During the validation interviews, participants provided their perspectives on the proposed architecture principles, acknowledging their relevance while also identifying potential areas of disagreement. One participant questioned the value of open principles, emphasizing the priority businesses place on practical benefits. Additionally, some interviewees highlighted the blurred distinction between principles and requirements, suggesting that certain principles, like security, are often seen as fundamental prerequisites. However, an alternative viewpoint suggests that security can be approached as a flexible guideline, adapting to specific needs. Conflicting principles were also observed, such as the tension between high security and resource availability. These insights offer a more comprehensive understanding of the proposed architecture principles, emphasizing the significance of diverse perspectives and the need to address potential conflicts when applying these principles in the design of a logistics data marketplace.

Our study has notable limitations, including a limited scope in the Systematic Literature Review due to scarce data on the research question and the rapidly evolving nature of the domain. The extended research duration also led to the exclusion of recent publications. The semi-structured interview methodology had limitations as well, with a small sample size affecting generalizability and a sole focus on expert sampling, which may have overlooked insights from other stakeholders. Lastly, the validation of architecture requirements and principles might not be sufficiently rigorous, potentially leading to gaps or inaccuracies in the proposed architecture principles due to unaddressed domain-specific nuances.

A natural progression of this work is, therefore, to focus on enhancing and validating the proposed architecture principles. Further investigation into the business models and monetary value of these marketplaces can also further drive their adoption and maximize value for stakeholders while considering the industry's competitive landscape. By refining the architecture requirements and exploring the economic aspects of data sharing, future research can contribute to a more efficient and secure data-sharing environment, fostering collaboration and innovation in the logistics sector.

List of Figures

1 Relationship between Architecture Types (adapted from Wulfert & Schütte (2022)) 8
2 Design Science Research Methodology Process Model (Peffers et al. 2008) 11
3 Research Flow Diagram 13
4 A model of electronic marketplaces discerning between three ownership types (Stahl et al., 2016) 16
5 A generic business-to-business multi-sided data marketplace ecosystem (adapted from Spiekermann (2019)) 17
6 Elements of Platform-Centric Ecosystems (Tiwana et al., 2010) 19
7 IDS Soft Infrastructure and neutral building blocks (adapted from Nagel et al. (2021)) 21
8 Coding approach executed in this study 26
9 Excerpt from Coding Process: Two-Column Table 27
10 Overview Coding Process 28
11 Requirement Elicitation Process 40
12 Principle Elicitation Process (adapted from (Greefhorst & Proper, 2011a) 47
13 Interdependencies between Architecture Principles 55
14 Data-2-Value Framework (adapted from Lim et al. (2018) 57

List of Tables

1 Overview of the Logistics Industry . . . 3
2 Relationship between Logistics Service Providers and Key Components of the Logistics Industry 4
3 Overview Data Types (adapted from Helo & Hao (2022) . . . 5
4 Regulations included in the European Strategy for Data and their potential impact on data marketplaces. . . . 6
5 Keywords and related terms . . . 15
6 Domain Agnostic Requirements (adapted from Nagel et al. (2021)) . . . 20
7 Interview Participants . . . 24
8 Logistic Platforms . . . 31
9 Use cases . . . 32
10 Elicited data provider concerns . . . 33
11 Elicited data buyer concerns . . . 35
12 Shared concerns . . . 36
13 Non-Functional Requirements . . . 42
14 Functional Requirements . . . 43
15 Domain Agnostic Requirements (adapted from Nagel et al. (2021)) . . . 45
16 Comparison domain agnostic requirements . . . 46
17 Architecture Principles . . . 48
18 Architecture Principle 1: Secure and Controlled Data Sharing . . . 49
19 Architect Principle 2: Data Quality and Management . . . 50
20 Architect Principle 3: Data Processing and Analysis . . . 51
21 Architect Principle 4: Transparency and Fairness . . . 52
22 Architect Principle 5: Usability and Scalability . . . 53
23 Architect Principle 6: Compliance And Legal Considerationss . . . 53
24 Architect Principle 7: Interoperability and Open Standards . . . 54
25 Interview Participants . . . 56
26 All elicited concerns . . . 77

Contents

Preface i

Executive Summary ii

List of Figures iv

List of Tables v

1 Introduction 1

1.1 Delineation of Data Sharing 2
1.2 Data Sharing in Logistics 2
1.2.1 Industry Structure 2
1.2.2 Data Types and Standards 4
1.2.3 Data Sharing Arrangements 4
1.3 Regulatory Framework 6
1.4 The Value of Architecture Design 6
1.5 Research Objective 8
1.6 Relevance Of The Study 9
1.7 Outline of Thesis 10

2 Research Methodology 11

2.1 Main Research Question 11
2.2 Research Strategy 11
2.3 Research Methods 12

3 Research Background 15

3.1 Search Process 15
3.2 Data Marketplaces 16
3.2.1 Definition and Relevance 16
3.2.2 Scope 16
3.3 Multi-sided Data Marketplaces 17
3.4 Concerns With B2B Data Sharing 18
3.4.1 Technical Challenges 18
3.4.2 Governance Challenges 18
3.5 Development Scenarios 19
3.6 Platform Architecture and Governance 19
3.7 Conclusion 21

4 Formulation of Architecture Requirements **22**

4.1 Interview Design 22

4.1.1 Interviewee Selection 23

4.1.2 Interview Approach 24

4.2 Interview Analysis 26

4.2.1 Coding Approach 26

4.2.2 Operationalization of Results 28

4.2.3 Results: Data-to-Value 29

4.2.3.a Barriers to large-scale data trading 29

4.2.3.b Monetization opportunities 30

4.2.4 Results: Data Sharing Concerns 32

4.2.4.a Data provider concerns 32

4.2.4.b Data buyer concerns 35

4.2.4.c Shared concerns 36

4.2.5 Results: Platform Design 37

4.2.5.a Architecture Components 38

4.2.5.b Governance Components 39

4.3 Requirement Elicitation 40

4.3.1 Formulating Requirements 40

4.3.2 Requirements Specification 41

5 Formulation of Architecture Principles **44**

5.1 Comparison with Domain-Agnostic Requirements 44

5.2 Generalization of Findings 47

5.2.1 Architecture Principles 47

5.2.1.a Architecture Principle 1: Secure and Controlled Data Sharing 48

5.2.1.b Architecture Principle 2: Data Quality and Management 49

5.2.1.c Architecture Principle 3: Data Processing and Analysis 50

5.2.1.d Architecture Principle 4: Transparency and Fairness 51

5.2.1.e Architecture Principle 5: Usability and Scalability 52

5.2.1.f Architecture Principle 6: Compliance And Legal Considerations 53

5.2.1.g Architecture Principle 7: Interoperability and Open Standards 53

5.2.2 Interdependencies between Architecture Principles 54

5.3 Evaluation of the Architecture Principles 56

6 Discussion **57**

6.1 Information Marketplaces, not Data Marketplaces 57
6.2 Architecture Implications on Value Proposition 58

7 Conclusion **60**
7.1 Sub-Research Questions 60
7.1.1 Sub-Research Question 1 60
7.1.2 Sub-Research Question 2 61
7.1.3 Sub-Research Question 3 62
7.2 Main Research Question 63
7.3 Scientific Contributions 64
7.4 Managerial and Societal Contributions 65
7.5 Suitability of this research to the MoT program 65
7.6 Limitations of the study 66
7.7 Directions for future research 66

A Overview of the sources included in the literature review **73**

B Invitation E-mail Interview Participation **75**

C Interview Questions **76**

D Data Sharing Concerns **77**

E Architecture Requirements Derived Objectives **78**

1 Introduction

Data has become an independent and strategic asset as the world transitions towards the digital economy (Reinsel et al., 2017). The amount of data created over the next three years will already surpass the total size of data created over the past thirty years (Reinsel et al., 2020). Advancements in wireless networks, computing power and data storage are vital drivers enabling businesses to collect and store this vast amount of data. Data is a crucial component for data-driven products and services, and it plays a central role in fact-based decision-making for process improvements, innovation and prediction of future events through data analysis (European Commission, 2017). Clever use of this data, therefore, can have transformative effects on all aspects of the economy and create opportunities for economic growth. However, although these data sets are abundant and growing daily, they are not being traded openly on a large scale (Spiekermann, 2019).

Generating value from data requires the ability to find, access and make sense of large sets of unstructured data (Chapman et al., 2019). Surveys such as that conducted by McKinsey (2017) have shown that small- and medium sized enterprises (SME) often lack the expertise, knowledge and/or resources to engage in data trading. In contrast, businesses that do share their business data, generally only do so with trusted partners and within closed data sharing platforms - in so called vertical collaborations (Evofenedex, 2021). Many studies such as (Gelhaar & Otto, 2020; Koutroumpis et al., 2020b) have identified similar concerns in why businesses are reluctant to share their valuable business data. First, data is easily replicated, therefore providers fear the loss of control after disclosing it to a third-party. Second, businesses fear that the content of their data may be useful to competitors and harm their own interests. Third, data buyers often do not want to pay a high price without the data content disclosed or utility assessed.

The logistics industry plays a crucial role in the global economy, with efficient and reliable supply chains being the backbone of trade and commerce. Despite the critical nature of this sector, it also suffers from a lack of effective data sharing practices between logistics service providers (Bastiaansen et al., 2020). Current solutions are siloed, and instead of being structured, integrated, and openly accessible from a single-entry point, they still tend to be fragmented and proprietary (Bastiaansen et al., 2020). This has resulted in dispersed information, hampering the efficiency, visibility, and overall performance of logistics operations. As the complexity of logistics operations continues to grow, driven by the rapid evolution of global trade and e-commerce, there is an increasing need for large-scale data sharing to unlock the potential of modern logistics management. Large-scale data sharing refers to the efficient exchange of vast volumes of structured and unstructured data between diverse stakeholders, including shippers, carriers, logistics service providers, and other relevant parties in the logistics domain. When effectively implemented, large-scale data sharing can enable enhanced visibility into supply chain operations, improved decision-making, and increased collaboration among industry stakeholders (Ceniga & Sukalova, 2015). This, in turn, can lead to optimized routes, reduced costs, minimized delays, and better overall performance in the logistics sector.

One promising concept that can enable large-scale data sharing in the logistics domain is the emergence of multi-sided data marketplaces. In a multi-sided data marketplaces, multiple data providers offer their data sets, and multiple data consumers may request access and utilize these data sets for various purposes (Koutroumpis et al., 2020b). These platforms act as facilitators, connecting these data providers, consumers and other stakeholders in an ecosystem that facilitates the efficient exchange of valuable data. By providing a one-stop-shop for logistics service providers, multi-sided data marketplaces can streamline their (=logistics service providers) process of data exchange; fostering collaboration, and driving innovation in the logistics industry (Bastiaansen et al., 2020). Although multi-sided data marketplaces have shown great potential in various domains, such as Mobility (Soto et al., 2023), there has been limited research on their application in the context of logistics. In particular, the development of architecture principles to guide the design of logistics-focused multi-sided data marketplaces has received little attention. Such principles are essential for ensuring that these platforms are tailored to the unique requirements and challenges of the logistics industry as a whole.

This thesis fills a gap in the field by formulating a set of architecture principles specifically tailored for logistics-focused multi-sided data marketplaces. By combining insights from relevant literature and gathering primary data from logistics experts on domain-specific concerns related to the participation of logistics service providers in a logistics-focused multi-sided data marketplace, this study is able to formulate architecture requirements for the design of a logistics-focused multi-sided data marketplace. These architecture requirements are subsequently translated into architecture principles for guiding future design and implementation of such marketplaces in the logistics domain. To ensure the validity and relevance of the proposed architecture principles, an evaluation is conducted through additional round of interviews with logistics domain experts. This evaluation process enriches the study by reviewing the relevance, completeness and effectiveness of the architecture principles thereby addressing the lack of understanding of what architecture principles are applicable for a concrete logistics-focused multi-sided data marketplace architecture. By providing a foundation for the development of effective and secure logistics-focused multi-sided data marketplace architectures, these architectural principles not only address the specific needs of the logistics domain but also contribute to the broader understanding of the role and significance of multi-sided data marketplaces in a novel context.

1.1 Delineation of Data Sharing

For the purpose of analysis, it is necessary to further delineate what is meant by data sharing in this thesis. Generally, data sharing encompasses all possible forms and models underpinning business-to-business data access or transfer in the context of the data economy (European Commission, 2017). However, in this study, the focus must be narrowed down further.

1. Data sharing in this study only addresses direct sharing between logistics service providers and does not consider data exchange between logistics service providers and other stakeholders such as the government or actual consumers.
2. The main interest concerns sharing of data between logistics service providers who have had no prior business relationship with each other. As yet, most data sharing is within closed platforms, it will be interesting to delve into the concerns that arise when there is no pre-existing basis for trust.
3. This thesis also does not specify a specific focus segment within logistics, and will therefore consider all types of logistics service providers in the analysis[1]. Broadly speaking, multisided platforms (MSPs) are technologies, products or services that create value primarily by enabling direct interactions between two or more customer or participant groups (Eisenmann et al., 2006). These platforms essentially leverage their network to enhance the matching between sides and thereby create added value. For example, users derive more value from Facebook as more people are available to them. As Eisenmann et al. (2006) explains, greater scale generates more value, which attracts more participants, which in turn creates more value—thus creating a positive feedback loop or the so-called positive indirect network effects. For this reason, this thesis attempts to understand whether a single large platform can accommodate the needs of the entire logistics domain, unless there is evidence it would not work.
4. Last but not least, Europe has seen significant proliferation in recent years with research on data marketplaces, privacy and security regulations as well as a push in data sharing collaboration initiatives through heavy institutional investment (European Commission, 2020). For this reason, Europe has been at the forefront of discussions surrounding multi-sided data marketplaces, and has influenced policy discussions and developments globally. Researching multi-sided data marketplaces within this context allows for an in-depth analysis of the implications, challenges, and compliance requirements associated with data sharing, privacy, and consent in a highly regulated environment.

1.2 Data Sharing in Logistics

Logistics is defined as the process of planning and executing the efficient transportation and storage of goods between their origin and end-destination (Topolšek et al., 2018). In other words, logistics covers all the activities that are needed for planning, implementing and managing flows of goods and information, from raw materials to end products. In addition to the value and job opportunities generated by the logistics industry, it also serves an essential role in supporting all other industries by enabling and promoting international trade. By gaining a comprehensive understanding of the logistics landscape, we can better appreciate the challenges and opportunities associated with data sharing and its potential impact on the industry. Therefore, in this section, we will delve into the complexities of data sharing within the logistics industry, providing an overview of the key activities and actors (1.2.1), data type and standards (1.2.2), and existing data sharing arrangements (1.2.3).

1.2.1 Industry Structure

The logistics industry encompasses a range of activities and actors, each playing a crucial role in the movement of goods from their point of origin to their final destination (Yontar, 2022). These activities include

[1] Note that throughout this study, we may also refer to 'specifically focused for the logistics industry' as 'logistics-focused' or 'domain-specific'. In any case, they mean specifically for the logistics industry, thereby creating a distinction between 'for logistics' and 'for any industry'

transportation, warehousing, inventory management, order fulfillment, and customs clearance, among others (Table 1).

Table 1: Overview of the Logistics Industry

Transportation	**Warehousing**	**Inventory Management**	**Packaging**	**Freight Forwarding**	**Technology**
Movement of goods by trucks, trains, ships, and planes	Storage and management of goods in warehouses and distribution centers	Tracking and management of inventory levels	Design and creation of packaging materials	Coordination of shipments and transportation arrangements	Use of automation, AI, and IoT to optimize logistics operations and improve efficiency

Building upon the understanding of the diverse range of activities encompassed by the logistics industry, it is important to delve deeper into the different types of actors and their roles within the industry. By segmenting the industry, we can obtain valuable insights into the various business models and the significance of data in their operations. In general, five key groups of actors can be identified for the logistics industry based on their business model (Silva et al., 2023):

1. First Party Logistics Providers (1PL) are manufacturers or industrial actors who do not outsource transport and logistic activities to third parties. These are carried out by the company's own departments.
2. Second Party Logistics Providers (2PL) are businesses or persons who can be subcontracted for the operational execution of a clearly defined transport or logistic task transportation. The relation is only short-term and cost-driven.
3. Third Party Logistics Providers (3PL) are businesses who can be subcontracted for the organization of both in- and outbound transportation activities as well as warehousing. The execution of the delegated package of transport and logistic activities will be organized by the 3PL, however, they may hire third parties for the execution of specific tasks. The partnership between the 3PL and the customer is often long-term and cooperative.
4. Fourth Party Logistics Providers (4PL) are businesses who can take responsibility of planning, organizing and managing the required activities for a full supply chain. The key difference between 3PL is that 4PL Logistics Providers are also tasked with monitoring the logistic processes. The relation here is again a long-term partnership where quality of service plays a key role. 4PL are often also seen as consultants as they do not have any asset to execute the specific activities and therefore require third parties for the execution of the tasks.
5. Fifth Party Logistics Providers (5PL) guarantee the management of networks of supply chains. Unlike 4PL, 5PL also develop and implement solutions to achieve the best possible supply chain networks. They can subcontract 4PL for the management of specific supply chains. 5PL is still an emerging concept, however, as evident by the lack of examples in practice.

In addition to these actors, shippers are also essential. Shippers are the manufacturers or industrial actors who subcontract carriers and other logistic providers for the execution of transport and/or logistic activities. This includes, for example, manufacturers, retailers and wholesalers who are active in the business of buying and selling goods. With the growing complexity of supply chains and the increasing reliance on efficient logistics operations, shippers need to collaborate with various logistics providers to ensure smooth and timely transportation of goods. Table 2 showcases the relationship between the logistics service providers and the role they fulfil. Closer inspection of the table shows:

1. Increasing levels of outsourcing and complexity moving from 1PL to 5PL. While 1PLs handle all logistics functions in-house, 2PLs to 5PLs involve various degrees of contracting and managing logistics activities.
2. Demonstration of how each type of logistics service provider engages with a key components of logistics operation. For example, 2PLs are primarily focused on contracted transportation, while 3PLs offer a more comprehensive range of services, such as contracted transportation, warehousing, inventory management, packaging, and freight forwarding.

3. 4PLs and 5PLs taking on more strategic and coordinating roles in the supply chain, rather than just providing specific services.

Table 2: Relationship between Logistics Service Providers and Key Components of the Logistics Industry

Logistics Service Providers	**Transportation**	**Warehousing**	**Inventory Management**	**Packaging**	**Freight Forwarding**
1PL	In-house transportation	In-house warehousing	In-house inventory management	In-house packaging	In-house freight forwarding
2PL	Contracted transportation	-	-	-	-
3PL	Contracted transportation	Contracted warehousing	Contracted inventory management	Contracted packaging	Contracted freight forwarding
4PL	Coordinates and manages entire supply chain, including transportation	Coordinates and manages entire supply chain, including warehousing	Coordinates and manages entire supply chain, including inventory management	Coordinates and manages entire supply chain, including packaging	Coordinates and manages entire supply chain, including freight forwarding
5PL	Manages networks of supply chains, including transportation	Manages networks of supply chains, including warehousing	Manages networks of supply chains, including inventory management	Manages networks of supply chains, including packaging	Manages networks of supply chains, including freight forwarding

1.2.2 Data Types and Standards

Logistics data refers to the range of information generated, exchanged, and utilized by the businesses involved in the planning, management, and execution of the movement and storage of goods along the supply chain (i.e. 1-5PLs). Logistics data can be classified into several categories based on their nature, purpose, and source (Silva et al., 2023). Table 3 provides an overview of each data type based on its source. These data types can be further classified in structured, semi-structured and unstructured data. Structured data refers to information that is organized in a predefined schema or format, allowing for easy storage, querying, and analysis. In contrast, unstructured data is information that does not have a predefined structure or format, making it more challenging to analyze and manage. Semi-structured data on the other hand is a type of data that does not conform to a fixed schema but still has some level of organization or structure. These formats are often used for data interchange between logistics partners, such as exchanging shipment details or tracking information (Helo & Hao, 2022).

The logistics industry commonly uses data formats and standards such as EDI (Electronic Data Interchange) and XML (Extensible Markup Language) – (Park et al., 2012). EDI is a standard format for exchanging business documents between different computer systems, while XML is a markup language used for encoding documents in a format that is both human-readable and machine-readable. These formats are used to facilitate communication and data exchange between different stakeholders in the logistics ecosystem. However, according to Park et al. (2012), the adoption of these formats also has disadvantages, including the complexity of implementation and the need for standardization across different systems.

1.2.3 Data Sharing Arrangements

Logistics service providers (LSPs) share data on a business-to-business (B2B) level in various ways to facilitate communication, improve collaboration, and optimize supply chain operations. Some of the common methods currently in use are:

1. Electronic Data Interchange (EDI) is a widely used method for B2B data exchange, allowing companies to share structured data electronically. It standardizes the format for common business documents, such as purchase orders, invoices, and shipment notices, enabling efficient communication between different systems (Park et al., 2012).

Table 3: Overview Data Types (adapted from Helo & Hao (2022)

Data Type	**Brief Description**	**Examples of Data**
Shipment Data	Information about individual shipments, such as product descriptions, quantities, weights, and dimensions.	Product name, SKU, weight, dimensions, packaging type.
Transportation Data	Details about transportation modes, carriers, and schedules.	Carrier name, vehicle capacity, transit time, route.
Inventory Data	Data related to the storage and management of goods.	Stock levels, warehouse locations, stock status.
Order Data	Information on customer orders.	Order number, order date, delivery requirements.
Tracking Data	Real-time data on the location, status, and condition of shipments during transit.	Current location, estimated arrival time, shipment status, temperature (if applicable).
Customs and Regulatory Data	Information required for compliance with customs and other regulatory requirements.	Import/export documents, licenses, permits, tariff codes.
Financial Data	Data related to the costs and revenues associated with logistics activities.	Freight rates, invoices, payment terms.
Performance Data	Metrics and KPIs used to measure the efficiency, effectiveness, and quality of logistics operations.	On-time delivery rate, order accuracy, warehouse utilization.

2. Application Programming Interfaces (APIs) enable real-time data exchange between different software applications and systems. LSPs can use APIs to share data, such as shipment tracking information, inventory levels, and order status updates, directly with their business partners' systems (Ofoeda et al., 2019).
3. Cloud-based platforms and portals: LSPs use cloud-based platforms to centralize and share data with their business partners. These platforms often include web-based portals, allowing partners to access relevant data and analytics, submit orders, and monitor shipment status. This approach improves visibility and collaboration across the supply chain (Otto et al., 2019).
4. Electronic Document Management Systems (EDMS): LSPs may use EDMS to store, manage, and share electronic documents, such as bills of lading, customs declarations, and waybills, with their business partners. These systems streamline the documentation process, reduce manual errors, and accelerate data exchange (Malekani et al., n.d.).
5. Data sharing via industry-specific platforms or networks: In some cases, LSPs participate in industry-specific platforms or networks designed to facilitate data sharing and collaboration. Examples include the Digital Container Shipping Association (DCSA) and CargoSmart ((CargoSmart, n.d.). These platforms establish common standards and protocols for data exchange, enhancing interoperability between different stakeholders in their logistics ecosystem (Spiekermann, 2019).
6. File Transfer Protocol (FTP) and Secure File Transfer Protocol (SFTP): LSPs may use FTP or SFTP to transfer data files securely between their systems and their partners' systems. This method is typically used for transferring large data files or when real-time data exchange is not required.

While these methods have been effective in sharing data between logistics service providers and their business partners, there is still room for improvement. The data infrastructure of the Dutch logistics industry is characterized by data silos, meaning data sets of individual organizations are often inaccessible to external organizations (Bastiaansen et al., 2020). Furthermore, whereas larger logistic service providers increasingly share their data with supply chain partners (Spiekermann, 2019), little data is shared with competitors or across industries. Data transactions at present thus occur in low volumes within closed data sharing platforms or bi-lateral transactions. According to a study by Duparc et al. (n.d.), regarding the openness of platforms, almost 60 percent of the logistics platforms are closed source, meaning that no external modifications can be made to the platform's underlying code. 32 percent of the logistics platforms used open-source APIs to enable data exchange and interoperability. Solely 6 percent of the analyzed platforms have open-source projects beyond open source APIs and allow external developers to participate.

1.3 Regulatory Framework

Regulations for data sharing ensure compliance with data protection laws and provide a framework for handling sensitive data, which encourages trust and innovation. The EU introduced a new strategy to harness the potential of data and to create a European data economy; the European Strategy for Data European Commission (2020). The strategy promotes data sharing and access to data by creating a single market for data in the EU. The regulations included in the European Strategy for Data can be found in Table 4.

Table 4: Regulations included in the European Strategy for Data and their potential impact on data marketplaces.

Regulation	**Summary**	**Potential Impact on Data Marketplaces**
GDPR	Rules for protection of personal data within the EU.	Increases compliance costs and legal requirements. Requires best practices for data protection and privacy. Builds user trust by protecting personal data, potentially increasing the market for data marketplaces.
Open Data Directive	Promotes availability and reuse of public sector data across the EU.	Increases data availability of public sector data. Requires compliance with rules for data reuse. Increases data availability for data marketplaces, potentially driving growth and innovation.
Digital Single Market Strategy	Aims to create a more unified and competitive digital market within the EU.	Creates new opportunities for data marketplaces to expand across the EU and offer new services. Requires compliance with rules for cross-border e-commerce and online content. Provides access to new markets and customers, potentially driving growth and innovation for data marketplaces.
Horizon Europe	Supports research and innovation in the EU, including in the field of data.	Provides funding for innovative data management solutions. Requires development of new data technologies and solutions. Increased research and innovation can drive growth and innovation for data marketplaces.
Data Governance Act (DGA)	Regulates handling of data within the EU and creates a framework for data sharing and use.	Promotes data sharing and access, particularly for SMEs and startups. Increases compliance costs for data marketplaces with stronger data protection measures. Provides a clear framework for data sharing and use, potentially increasing trust and innovation in the data marketplace.
Data Act	Clarifies who can create value from data and under which conditions.	Increases data availability. Encourages data sharing and access regardless of size of the actor, potentially driving growth of the market.

The European strategy for data is relevant for multi-sided data marketplaces because its vision aligns with the objectives of multi-sided data marketplaces, which are designed to facilitate data sharing, collaboration, and innovation among various stakeholders.

1.4 The Value of Architecture Design

A key issue in the design and development of any complex software system is its architecture, i.e. its organization as a collection of interacting elements (modules, components, services etc.) (Garlan, 2014). According to Garlan (2014), architecture refers to the overall structure and design of a system that involves both hardware and software components. It provides an abstract description of the system, which helps designers and developers to understand and reason about the system's behavior, performance, and quality attributes. The architecture of an information system consists of several interconnected components or modules, each of which performs a specific function (Nakagawa et al., 2011). These modules can be designed using various patterns and architectural styles, such as client-server, service-oriented, or event-driven architectures, depending on the requirements of the system and the problem domain. An effective information systems architecture should expose certain properties of the system while hiding others. It should provide a clear and concise representation of the system's behavior, structure, and components, while abstracting away implementation details that are not relevant to the system's functionality or performance (Garlan, 2014). Moreover, the architecture should explicitly capture the intent and architecture principles that govern the system's design, and prescribe a blueprint for system construction and composition. This can help ensure that the system is well-designed, maintainable, and scalable, and can also facilitate future modifications and enhancements to the system.

A good architecture can help ensure that a system will satisfy its key functional and quality requirements, including performance, reliability, portability, scalability, and interoperability. Architecture requirements are the high-level functional and non-functional requirements that a system must satisfy to achieve its goals (Nakagawa et al., 2011). These requirements typically focus on the system's overall performance, reliability, security, and maintainability. Effective architecture design therefore requires a deep understanding of the system's requirements, as well as knowledge of relevant design patterns, principles, and best practices. It also requires careful consideration of trade-offs between different design options, such as performance vs. maintainability or scalability vs. simplicity. A well-designed architecture can help reduce development time and costs, improve the system's quality and performance, and enable future enhancements and modifications to be made more easily. Garlan (2014) outlines six reasons to develop a system architecture, which include understandability, reuse, construction, evolution, analysis, and management.

1. Understandability: Developing an architecture provides a clear and concise representation of the system's components, interactions, and constraints. This allows stakeholders to understand the system's behavior and performance and helps facilitate communication and collaboration among designers, developers, and other stakeholders.
2. Reuse: An architecture can help promote reuse of components, modules, and other design artifacts. This can help reduce development time and costs and improve the system's maintainability and scalability.
3. Construction: Developing an architecture can help ensure that the system is well-designed and implemented, by providing a blueprint or plan for system construction and composition.
4. Evolution: An architecture can facilitate future modifications and enhancements to the system by providing a framework for system evolution and change. This can help ensure that the system remains relevant and adaptable to changing business needs and technology.
5. Analysis: An architecture can be used to analyze and evaluate the system's performance, reliability, and other quality attributes. This can help identify potential issues and improve the system's overall quality and performance.
6. Management: An architecture can help facilitate the management of the system over its lifecycle, by providing a framework for system monitoring, maintenance, and support.

In the context of designing an architecture for a multi-sided data marketplace in the logistics domain, there exists a distinct value in developing such an architecture. This value stems from three aspects, each contributing to the understanding and advancement of multi-sided data marketplaces within logistics data sharing.

1. a domain-specific implementation holds significant potential as it offers a means to demonstrate the feasibility and applicability of multi-sided data marketplaces within the logistics data sharing context. By implementing a domain-specific multi-sided data marketplace, researchers can gain practical insights into its functionality, identify potential areas for further development, and showcase how the architecture can be effectively utilized in real-world scenarios. This demonstration serves as evidence of the practical viability and potential benefits of multi-sided data marketplaces within the logistics industry;
2. the design of an architecture for sharing logistics data necessitates an in-depth understanding of the domain-specific needs, concerns, and constraints inherent in this context. Developing such an architecture requires a comprehensive assessment of these factors to derive appropriate architecture requirements. By undertaking this assessment, the research establishes a foundational understanding of the logistics data sharing landscape, which can subsequently inform and shape future research endeavors in the field of logistics data sharing platforms. The derived architecture requirements provide a valuable framework for researchers and practitioners alike, enabling the design and development of effective, efficient, and domain-specific logistics data sharing platforms;
3. this research holds the potential to serve as a guide for the design and creation of multi-sided data marketplaces in other domains or data sharing contexts. The knowledge gained from designing an architecture for logistics data sharing can be extrapolated and applied to various other industries or sectors that seek to implement multi-sided data marketplaces. This transferability stems from the identification of generalizable principles, best practices, and architecture requirements that transcend the logistics domain. As a result, the research outcomes extend beyond the immediate scope of logistics data sharing, providing valuable insights and guidance for the design and implementation of multi-sided data marketplaces in diverse data sharing contexts.

Creating a complete architecture design, is however an elaborate process that requires input from various sources (Tiwana, 2014; Tiwana et al., 2010). Given the time constraints of this study, this is deemed out of scope. **There is, however, a different approach we can take to make inference on whether the architecture design of a logistics-focused multi-sided data marketplace is viable and to an extent provide a starting point for future architecture designs in this field. One such approach is by developing domain-specific architecture requirements and principles.**

1.5 Research Objective

In the field of information systems design and development, architecture requirements and architecture principles play distinct yet interconnected roles. While both contribute to building a cohesive and efficient information system, they address different aspects of the process. Architecture requirements define the desired functionalities and qualities that the system should possess in order to meet specific needs or expectations (Greefhorst & Proper, 2011b). They are dynamic and evolve over time as the system evolves. On the other hand, architecture principles are overarching, fundamental guidelines that shape the design and implementation of the system's architecture (Nagel et al., 2021; Greefhorst & Proper, 2011b). Rather than specifying specific functions or qualities, they provide a consistent framework for decision-making throughout the system's design, development, and evolution. By offering a reference point for balancing competing concerns and requirements, architecture principles assist architects in considering the broader implications of their design decisions. Ultimately, the interplay between architecture requirements and principles ensures the development of a well-considered and adaptable information system (TOGAF, 2023).

It is necessary here to further clarify exactly the concepts underlying the architecture, as these concepts will be repeated often through this study (Figure 1). The red square in Figure 1 shows the focus area of our research.

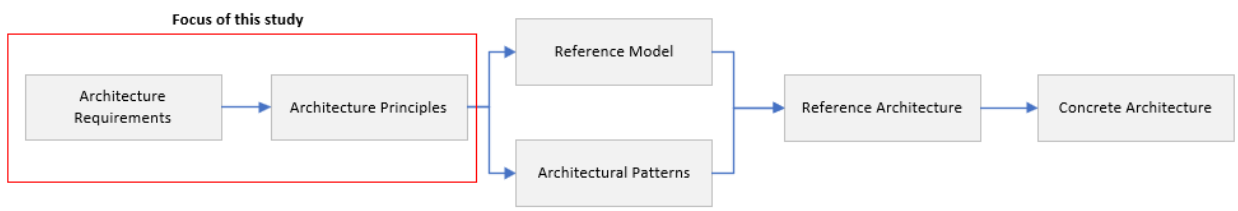

Figure 1: Relationship between Architecture Types (adapted from Wulfert & Schütte (2022))

Moving from right to left. Concrete information systems architectures deal with one particular company. However, reference architectures abstract from the company's context, therefore enabling the reuse of architecture components, providing an agreed-upon set of concepts and architectural patterns, and communicating fixed viewpoint (Nakagawa et al., 2011). The development of a reference architecture is typically inspired by concrete architectures but also refers models artifacts and thus has a "descriptive nature" (Wulfert & Schütte, 2022). As little research exists on concrete logistics-focused multi-sided data marketplace architectures; this approach of is not feasible. Developing a reference architecture based solely on existing research in a prescriptive manner is another approach that allows the creation of "a futuristic view of a class of systems". Reference architectures focus on clarification of innovative patterns and aim to convince domain architects of the architecture qualities. Consequently, concrete systems can be developed according to this research-based architecture (Wulfert & Schütte, 2022). Reference architectures are applied either to standardize existing systems to ensure interoperability or to facilitate the design and improve the quality of a concrete architecture with architectural guidelines. They can be used as a starting point for company-specific models to reduce the effort of creating them through reuse of established artifacts and constructs. As can be seen from Figure 1, however, reference architectures require architectural patterns and reference models as input. A reference architecture consists of several architectural patterns (Nakagawa et al., 2011). According to Nakagawa et al. (2011), architectural patterns are defined as a "named collection of architectural design decisions that are applicable to a recurring design problem". The patterns are reusable solutions to common architectural problems within a given domain. In turn, architectural patterns need architecture principles (Greefhorst & Proper, 2011b). Research on logistics-focused multi-sided data sharing platforms, however, have not treated the architectural principles that function as input for the reference and architectural patterns (Bastiaansen et al., 2020).

Following the information presented up to now, the three points of literature and knowledge gap could be presented as the following:

1. **Lack of guidance**: Currently, there is a lack of comprehensive guidance and principles specifically tailored to the design and development of multi-sided data marketplaces in the logistics domain. This absence of guiding principles makes it challenging for practitioners and researchers to create effective and efficient architectures that facilitate B2B data sharing among logistics stakeholders.

2. **Complexity and specificity**: The logistics domain presents unique challenges and requirements for data sharing, including diverse stakeholders, varying data types, security concerns, interoperability issues, and governance considerations. Addressing these complexities and developing architecture solutions that align with the specific needs of logistics data sharing necessitates a focused and specialized approach.

3. **Fragmented efforts**: The existing efforts in developing multi-sided data marketplaces within the logistics domain are often fragmented, lacking a cohesive and standardized approach. This fragmentation hinders interoperability, collaboration, and the realization of the full potential of data-driven innovation in the logistics industry.

To address these problems, the research objective is therefore to define architecture principles specifically tailored for multi-sided data marketplaces in the logistics domain. These architecture principles will provide a guiding framework and best practices for the design, development, and implementation of data marketplaces that facilitate efficient B2B data sharing within the logistics industry. The need for architecture principles stems from the understanding that a well-defined architecture is crucial for creating effective and sustainable multi-sided data marketplaces. By establishing architecture principles, this research aims to promote consistency, interoperability, and collaboration among logistics stakeholders. Furthermore, these principles will aid in addressing the complexities and challenges inherent in logistics data sharing, ensuring the development of secure, scalable, and adaptable marketplaces. By achieving the research objective of defining architecture principles for multi-sided data marketplaces in the logistics domain, this study will contribute to a better understanding of the design and development of these marketplaces.

1.6 Relevance Of The Study

This section highlights the academic, managerial, and societal relevance of studying the design of architecture principles for multi-sided data marketplaces.

The study of the design of architecture principles for a multi-sided data marketplace holds **academic relevance** in several ways. First, it can contribute to the field of information systems and data management by exploring the unique challenges and opportunities posed by multi-sided data marketplaces. By examining the architectural aspects, such as data integration, interoperability, security, and governance, researchers can build on the knowledge from this study to develop theoretical frameworks and models that enhance our understanding of the complex dynamics within these marketplaces. Secondly, this study fills a research gap by addressing the need for comprehensive guidelines and best practices for building multi-sided data marketplaces. While there is a growing interest in data marketplaces, much of the existing literature predominantly focuses on closed marketplaces (Hagiu & Wright, 2015) or specific technical aspects from a generic perspective (Spiekermann, 2019; Koutroumpis et al., 2020a; Otto et al., 2019). By investigating the architecture principles specifically tailored for multi-sided scenarios in the logistics domain, this research can contribute to the advancement of knowledge and provides practical insights for researchers and logistics practitioners.

The managerial relevance for marketplace owners of this study lies in its ability to inform organizations and managers about the design and implementation of multi-sided data marketplaces. As the demand for data-driven insights and collaboration continues to rise, managers are seeking effective strategies to harness the value of data assets (Midway et al., 2022). By examining the architectural principles for multi-sided data marketplaces, this research may provide valuable insights into how organizations can create and operate successful data ecosystems. Managers may benefit from understanding the key considerations involved in establishing a multi-sided data marketplace, such as identifying and engaging relevant stakeholders, defining data governance frameworks, ensuring data privacy and security, and establishing pricing and monetization models. This knowledge may enable managers to make informed decisions regarding platform design, business models, and ecosystem strategies that maximize value creation for all participants. Moreover, by exploring the technical and organizational challenges associated with multi-sided data marketplaces, this research empowers managers to address issues related to data quality, standardization, and integration. It provides guidance on selecting appropriate technologies, implementing data sharing protocols, and fostering trust and collaboration among participants. By leveraging these insights, logistics managers may better navigate the complexities of multi-sided data marketplaces and develop effective strategies to unlock the full potential of the available data assets in the logistics industry.

The societal relevance of studying the design of architecture principles for multi-sided data marketplaces stems from its vision to drive innovation, economic growth, and societal well-being. Data has become a critical resource for addressing societal challenges and driving advancements in various domains, including healthcare, transportation, energy, and education (Reinsel et al., 2020). Multi-sided data marketplaces offer a framework for unlocking the value of data by facilitating collaboration, knowledge exchange, and data-driven decision-making. By enabling diverse stakeholders, including distinct but interrelated groups of shippers and logistics service providers to participate in a data ecosystem, multi-sided data marketplaces promote inclusive innovation. They encourage the sharing of data assets, research findings, and insights, leading to novel solutions, improved services, and evidence-based policy-making. These marketplaces can contribute to addressing current societal issues such as pollution and traffic. Furthermore, the study of multi-sided data marketplaces has societal implications in terms of privacy, ethics, and transparency.

1.7 Outline of Thesis

This study is composed of seven chapters, including this introduction. Chapter two delves into the research methodology employed in this study (Chapter 2). It describes the approach taken to gather and analyze data, including the use of literature review and interviews. The chapter also discusses the rationale behind each chosen methodology and its suitability for achieving the research objectives. Chapter three presents a literature review that explores existing knowledge and research related to multi-sided data marketplaces (Chapter 3). The objective of the literature review is to establish a foundation and build on existing knowledge of multi-sided data marketplaces. Chapter four serves as a design specification, outlining the key architecture requirements for the development of logistics-focused multi-sided data marketplaces (Chapter 4). It includes a detailed account of the interview process conducted with domain experts to gather the concerns which informed the domain-specific architecture requirements. Building upon the architecture requirements derived from the interviews, Chapter five formulates architecture principles specifically tailored for logistics-focused multi-sided data marketplaces (Chapter 5). This chapter translates the gathered insights into a framework that guides the design and implementation of future multi-sided data marketplaces in the logistics domain. Chapter six engages in a discussion of the findings, addressing the implications and significance of the formulated architecture principles (Chapter 6). It explores the practical implications and potential challenges in applying these principles within the logistics domain. The final chapter concludes this thesis by summarizing the key findings, highlighting the contributions made, and discussing potential avenues for future research (Chapter 7). It offers a conclusion that consolidates the study's outcomes and their relevance to the field of logistics data sharing and multi-sided marketplaces.

2 Research Methodology

In this chapter, the research methodology employed to address the main research question and its sub-questions are outlined. The selection of an appropriate research methodology is crucial, as it ensures the validity, reliability, and generalizability of the study's findings, allowing for a rigorous and systematic approach to the development of a multi-sided data marketplace architecture for the logistics industry. The chapter starts with a brief overview of the main research question (2.1). This is followed by a discussion of the chosen research strategy (2.2). In the third part of the chapter the specific research methods employed to gather and analyze data are discussed (2.3).

2.1 Main Research Question

As an open logistics data sharing platform does not exist yet, this thesis will set out to investigate what an architecture for logistics data could look like by developing architecture principles. Achieving the research objective requires the formulation of research questions. Answering these questions results in the knowledge necessary in reaching the purpose of a research (vom Brocke et al., 2020). Based on the research objective, the overall question of the research is framed as follows:

"What multi-sided data marketplace architecture principles can facilitate business-to-business data sharing in the logistics industry?".

2.2 Research Strategy

As the objective of this research question is to design and evaluate the usefulness of an IT artifact, it is considered to be a design-oriented study (Dresch et al., 2015). These type of studies are primarily executed according to the principles of Design-Science Research (DSR). Design science research is a method that establishes and operationalizes research when the desired goal is an artifact or a recommendation (Dresch et al., 2015). A DSR study can be operationalized in many different ways, however, two design strategies are generally followed (Iivari, 2015). The first strategy begins with a general problem to be addressed and builds, demonstrates and evaluates a conceptual IT meta-artifact as a DSR contribution. The second strategy drives the researcher to solve a problem encountered in practice by one or more specific client(s) with a concrete solution artifact. For this specific context, strategy one will be adopted as there is no client and the main research questions as discussed in 2.1 relates to a general problem.

Design Science Research (DSR) has the potential to bridge the gap between theory and practice for data marketplaces, as it not only focuses on problem-solving but also generates knowledge that can serve as a reference for improving theories (Abbas et al., 2021; Dresch et al., 2015). Several DSR research frameworks in the literature prescribe the necessary design activities for strategy 1, including those by A. R. Hevner et al. (2004); A. Hevner & Chatterjee (2010); Peffers et al. (2007); Vaishnavi et al. (2015), A. Hevner & Chatterjee (2010). For this study, we have selected the DSR research framework proposed by Peffers et al. (2007), as it offers a well-established, synthesized model that builds upon the guidelines of A. R. Hevner et al. (2004) – (Figure 2).

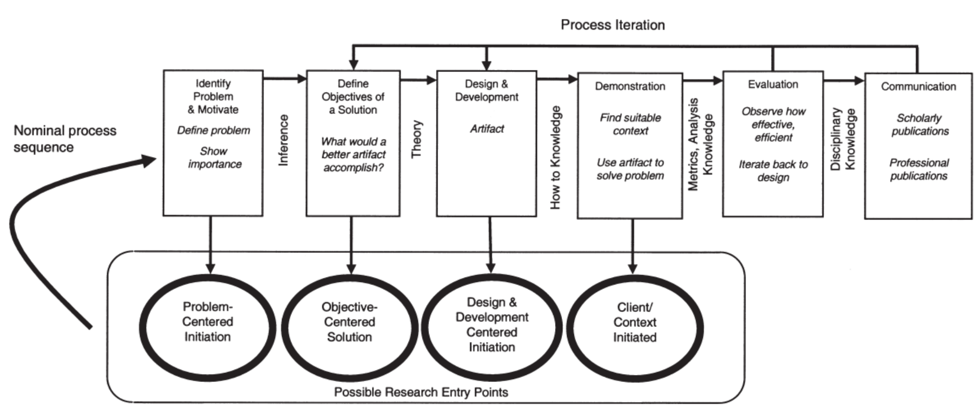

Figure 2: Design Science Research Methodology Process Model (Peffers et al. 2008)

The first activity of the DSR is to identify a problem or challenge that requires a solution or improvement. The objective is to provide a solid foundation and rationale for the subsequent research activities. In the first chapter of this study, desk research was conducted to identify current challenges and limitations in business-to-business data sharing within the logistics industry. This activity aligns closely with the first activity of the DSR framework Peffers et al. (2007). Based on this problem definition, the main research question was formulated. To guide the research process and provide a comprehensive answer, the main research question was subdivided into three sub-research questions. These sub-research questions have been formulated in accordance with the DSR model of Peffers et al. (2007):

1. What prior knowledge is available on the desired features and characteristics of a multi-sided data marketplace architecture for the logistics industry?
2. What are the key architecture requirements necessary to design a multi-sided data marketplace for the logistics industry?
3. How do the proposed multi-sided data marketplace architecture requirements address the challenges and desired features identified in the problem and objectives stages?

The first sub-research question seeks to build a foundation of understanding by exploring existing literature and knowledge related to multi-sided data marketplaces and their potential applications within the logistics sector. The second sub-research question aims to identify the essential elements, constraints, and considerations that must be taken into account when designing an effective and relevant MSDM tailored to the specific needs of the logistics industry. The third sub-research question evaluates the proposed solution by examining how well it aligns with the identified challenges and objectives, ultimately assessing its potential impact and effectiveness within the logistics domain.

By following the DSR framework, this study aims to systematically develop a solution that addresses the identified challenges and meets the requirements of the logistics industry. This structured approach ensures a rigorous exploration of existing knowledge, a comprehensive identification of essential architecture requirements, and a thorough evaluation of the proposed solution's potential effectiveness. The resulting insights will contribute to both the practical application of multi-sided data marketplaces in the logistics industry and the broader theoretical understanding of this emerging field.

2.3 Research Methods

This section clarifies how the design framework of Peffers et al. (2007) guides the development of architecture principles for a logistics-focused multi-sided data marketplace. Design science research adds a valuable dimension to this thesis by providing a practical and applied approach to problem-solving. In the context of this thesis, design science research offers a framework for developing and evaluating artifacts that address specific problems or challenges. By applying design science research, the thesis can therefore go beyond theoretical discussions and literature reviews to propose and validate practical solutions. According to Dresch et al. (2015), Design science research emphasizes the creation of artifacts, such as in our case with architecture requirements and principles, that can be implemented and tested in real-world settings. Therefore, this approach enables the thesis to contribute not only theoretical insights but also deliver actionable and tangible outcomes. Furthermore, design science research facilitates a systematic and iterative process of design, development, and evaluation. It allows for the incorporation of feedback from stakeholders and/or experts, ensuring the relevance, effectiveness, and practicality of the proposed artifacts. This iterative nature aligns well with the goal of developing architecture principles for multi-sided data marketplaces that address the specific challenges and requirements of the logistics domain.

The Research Flow Diagram (RFD) illustrates the complete overview of how the sub-research questions relate to the research process (Figure 3). As each sub-research question is aligned with a phase in the DSR framework, the output of each phase forms the input data for the subsequent phase. It is necessary here to clarify the RFD in more detail, in particular the plan of approach for each sub-research question and their integration in the selected DSR framework. In order to develop a well-rounded solution for the central research question, the approach for each sub-question is individually discussed in the subsequent paragraphs. For each sub-question, we define its objective, what data is required, what research method will be used to gather this data, and finally how the data is analyzed and processed.

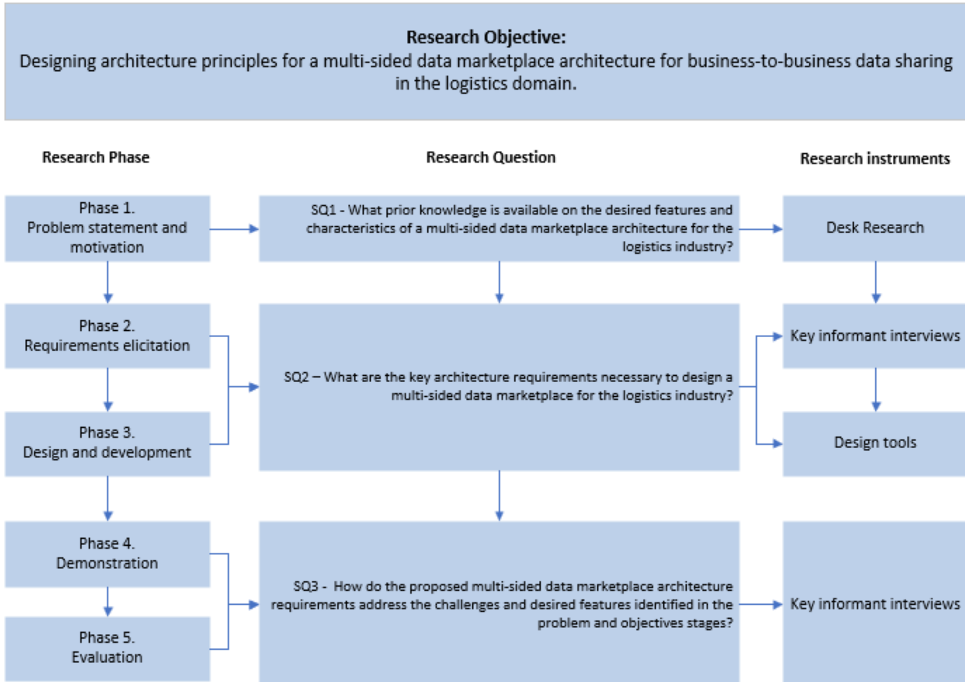

Figure 3: Research Flow Diagram

SRQ1 - What prior knowledge is available on the desired features and characteristics of a multi-sided data marketplace architecture for the logistics industry?

The primary objective of the first sub-research question (SRQ1) is two-fold. First, it aims to draw upon existing research by finding specific details about the building blocks and principal components necessary for developing a generic decentralized multi-sided data marketplace (MSDM). Secondly, it seeks to uncover the underlying domain-specific challenges within the logistics industry that are relevant for determining the functional and non-functional requirements of the architecture. By completing these objectives, we will gain a deeper understanding of what a logistics MSDM platform should achieve and how it can add value for market participants. SRQ1 is related to the second activity of the Design Science Research (DSR) framework, which involves establishing the objectives and desired outcomes of the artifact by understanding prior knowledge Peffers et al. (2007). As the objective of SRQ1 is to comprehend the theoretical background and uncover relevant domain-specific challenges, it is directly linked to this DSR framework activity.

To address SRQ1 comprehensively, we will conduct a literature review as our research method. This process involves collecting and analyzing information related to decentralized data marketplace architectures, their applicability to the logistics domain, and any specific challenges or opportunities in facilitating business-to-business (B2B) data sharing. The literature review will help ground the research in existing knowledge and identify relevant gaps in the current understanding of MSDMs in the logistics industry. A primary advantage of conducting a literature review is that it enables a comprehensive understanding of the current state of research on the topic (Greefhorst & Proper, 2011b). However, potential drawbacks are the risk of incompleteness due to vast amounts of literature, or a limited scope caused by scarce data availability. By performing a literature review, we will be able to identify knowledge gaps in the existing literature and establish the context for the remainder of the research. This approach will contribute to advancing our understanding of multi-sided data marketplaces in the logistics industry and lay a solid foundation for the subsequent research questions and activities within the DSR framework.

SRQ2 - What are the key architecture requirements necessary to design a multi-sided data marketplace for the logistics industry?

The primary objective of the second sub-research question (SRQ2) is to identify the essential architecture requirements for designing a multi-sided data marketplace (MSDM) specifically tailored to the logistics industry. By accomplishing this objective, we will gain a better understanding of the critical elements and considerations that must be taken into account when designing such a marketplace. This knowledge will guide the design process, inform decision-making, and contribute to the development of an effective and relevant solution. SRQ2 is related to the third activity of the Design Science Research (DSR) framework, which involves creating an artifact that addresses the problems and theories found in the first two activities Peffers et al. (2007). As the requirements elicited from SRQ2 address the issues identified in the previous research question, SRQ2 is directly linked to this DSR framework activity.

To address SRQ2, we will employ semi-structured interviews as our research method, which involves conducting guided conversations with experts and target users to gather domain-specific knowledge (Adams, 2015). We aim to conduct ten interviews to develop an understanding of the unique challenges within the logistics industry. The requirements for the MSDM will be inferred from both generic concerns and domain-specific needs, constraints, and considerations. One key benefit of this method is that it allows for the inclusion of multiple perspectives from potential users and experts, which could reveal new issues and insights about the topic (Navarrete, n.d.). However, a drawback is that the collected information often represents a surface layer of knowledge, making it difficult for participants to envision the ideal needs of users. Additionally, small sample sizes cannot be generalized to the larger population and may introduce biases in the results. To minimize researcher bias, a systematic and transparent qualitative data analysis technique will be employed. This approach will help ensure that the insights gathered from the interviews are accurately represented in the resulting architecture requirements, ultimately contributing to the development of an MSDM that effectively addresses the unique challenges and opportunities within the logistics industry.

SRQ3 - How do the proposed multi-sided data marketplace architecture requirements address the challenges and desired features identified in the problem and objectives stages?

The objective of this third sub-research question (SRQ3) is also two-fold. First, it involves gathering insights from logistics industry stakeholders to demonstrate, validate and evaluate the adapted or extended architecture principles. Through these second round of discussions, we aim to assess the effectiveness, and determine the practicality of the principles in facilitating business-to-business data sharing within the logistics industry. Subsequently, we are able to synthesize the validated and refined architecture principles into a comprehensive framework specifically tailored for facilitating business-to-business data sharing in the logistics industry. This includes consolidating the findings, identifying inter-dependencies between the architecture principles, and proposing a practical framework for future endeavors. Through this approach, we are able to both demonstrate and evaluate the proposed architecture principles, thus connecting SRQ3 to the fourth and fifth DSR framework activity (Peffers et al., 2007).

To accomplish the objective of demonstrating, validating and synthesizing the architecture principles into a framework for multi-sided data marketplaces in the logistics industry, we will again employ the key informant research method with with logistics industry stakeholders. This method involves in essence a second-round engagement with logistics experts, practitioners, and key stakeholders from research institutes. The structure of the interviews should be designed interviews aim to gather feedback, validate the adapted or extended architecture principles, and synthesize them into a comprehensive framework. Given the time constraints of this thesis and the fact that interviews can be a time-consuming and resource-intensive process, the aim is to conduct two to three interviews in total for demonstration and evaluation purpose. Building on the insights from these key-informant interviews, we will employ informed argumentation research method to demonstrate the usefulness of the proposed architecture principles by constructing logical arguments based on existing evidence and knowledge (Wulfert & Schütte, 2022). This method helps demonstrate the value and strengths of the developed architecture principles and clarifies their contributions to addressing the problems within the domain. We do this by comparing the developed underlying architecture requirements with domain-agnostic requirements found in the literature to gauge the comprehensiveness and novelty of our proposed architecture principles. A key benefit of informed argumentation is its ability to provide a persuasive rationale for the usability of the proposed architecture. However, a potential drawback is that it relies on the quality and depth of existing evidence, which may not always be comprehensive.

To communicate the usability of the architecture, along with the lessons derived, this thesis will include a detailed account of the research process, adopting a structure similar to the one proposed by (A. R. Hevner et al., 2004).

3 Research Background

In this section, we present the research background that underpins this study, providing an overview of the existing literature, theories, and models that inform the research questions and objectives. The chapter begins with the search process employed to identify relevant literature and sources (3.1), followed by the definition, relevance, and scope of data marketplaces (3.2). We then explore the characteristics of multi-sided data marketplaces (3.3), discuss the concerns associated with business-to-business (B2B) data sharing (3.4), present development scenarios for the evolution and growth of multi-sided data marketplaces in the logistics industry (3.5), and examine the considerations for designing and managing effective data-sharing platforms in terms of architecture and governance (3.6). The chapter concludes by summarizing the main findings and discussing their implications for this study (3.7). The following sub-research question is answered by this chapter:

SQ1 - What prior knowledge is available on the desired features and characteristics of a multi-sided data marketplace architecture for the logistics industry?

3.1 Search Process

An approach consisting of three iterative steps was followed to ensure a structured analysis of the relevant literature (Jesson & Lacey, 2011). The first step included systematically collecting potentially relevant publications based on the keywords and related terms listed in Table 5. These keywords were combined using Boolean operators (AND, OR, NOT) to form search strings that are specifically relevant to multi-sided data marketplace architectures in the logistics industry. The search was conducted using TU Delft thesis repository, Web of Science (WoS), and Google Scholar. The abstract, introduction, and conclusion of each paper were read to understand if the described work is of interest before proceeding to skim through the findings and figures. Based on this approach, seemingly relevant publications were saved as input papers. After identifying an input paper, forward and backward snowballing techniques were used to find additional references and/or keywords. As many relevant sources were scarcely cited, forward snowballing was often ineffective. However, backward snowballing paved the way for various enlightening articles. The second step focused on processing the identified input papers and selecting candidate papers that were relevant to read based on inclusion and exclusion criteria. In the third and final step, the candidate papers were read and synthesized to identify a research direction from the current state of research. By combining the keywords in Table 5 and using the search process outlined, the review was able to identify a comprehensive set of articles that provided information on multi-sided data marketplace architectures (Appendix A).

Table 5: Keywords and related terms

Keywords	Related Terms	Keywords	Related Terms	Keywords	Related Terms
multi-sided data marketplace	multi-party data exchange, data platform	data governance	data management, data stewardship	data privacy	data protection, data confidentiality
architecture principles	design principles, architectural guidelines	data security	data integrity, data safety	data quality	data accuracy, data reliability
business-to-business data sharing	B2B data sharing, inter-organizational data sharing	interoperability	data exchange, data integration	trust	trustworthiness, credibility
logistics industry	supply chain, transportation industry, freight industry	data monetization	data commercialization, data value	blockchain	distributed ledger technology
architecture requirements	architectural constraints, system requirements	smart contracts	self-executing contracts, blockchain contracts	data sharing models	data exchange models, data collaboration
high value data sharing	valuable data exchange, data monetization	logistics data	supply chain data, transportation data	data-driven logistics	digital logistics, intelligent logistics

3.2 Data Marketplaces

The purpose of this section is to present the findings of the literature review for data marketplaces that can enable large-scale data sharing. In the first part of this subsection, the definition and relevance of data marketplaces are defined (3.2.1). The second part discusses the scope of data marketplaces (3.2.2).

3.2.1 Definition and Relevance

Historically, the term 'marketplace' has been used to describe the concrete place of exchange that brings together the supply and demand of a commodity by facilitating transactions between market participants (**?**). In the context of data marketplaces, data becomes the commodity for trade and all transactions occur through an online infrastructure. Data, however, has different characteristics than physical goods; making it challenging to trade in online marketplaces, especially if it is highly valued (Koutroumpis et al., 2017, 2020b):

1. data needs to be processed with complementary ideas and assets to provide utility and thus requires a suitable match,
2. data is a non-rivalrous good meaning it can be replicated at low-cost and is re-usable by many individuals;
3. data has weak intellectual protection regime therefore it can be difficult to appropriate its full value and specify control rights .

Although varying definitions exist (Stahl et al., 2016; Spiekermann, 2019), there appears to be some agreement that a data marketplace refers to: 'a digital platform enabling and simplifying the transactions of machine-readable data by matching market participants, facilitating data sharing services and providing a pricing mechanism'. It is necessary here to clarify what is meant by digital platform. Ghazawneh and Henfridsson (2015) define digital platforms as: "software-based external platforms consisting of the extensible codebase of a software-based system that provides core functionality shared by the modules that interoperate with it and the interfaces through which they interoperate". Stahl et al. (2016) developed one of the first taxonomies using ownership and orientation, derived from economics theory on electronic markets, as dimensions to better understand and classify types of data marketplaces. Ownership indicates whether the data marketplace is owned by a private company, a consortium or whether it is operated without any connection to the market participants. Orientation indicates the extent to which pricing of the data and transactions are determined by the data marketplace owner (hierarchy) or users (market). Using these dimensions, a total of six distinct business model types were identified and defined through the number of market participants they have on each side (Figure 4).

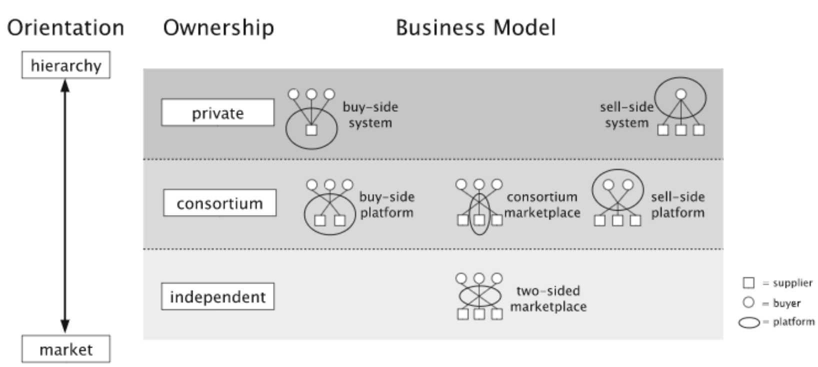

Figure 4: A model of electronic marketplaces discerning between three ownership types (Stahl et al., 2016)

3.2.2 Scope

Rather important for understanding relevant market developments is the question of what value propositions drive the emergence of these data sharing platforms (Figure 4). While the term 'platform' might sound neutral at first, they are set up and owned by particular businesses with different incentives and strategic development perspectives in mind. Stahl et al. (2016) points out that the level of orientation is shaped by how much the platform owner has an interest in the outcome of the transactions on the platform. Privately owned data marketplaces typically control the flow and price of the data sets to facilitate their interests, while independent marketplaces do not influence pricing and focus on minimizing transaction costs. 20pt Previous surveys such as that conducted by **?**van de Ven et al. (2021) have shown that privately- and consortium-owned data marketplaces were prevalent compared to independently operated data marketplaces (e.g. DAWEX). As private and consortium platform owners exhibit strategic behavior to protect their valued data sets and advance their self-interest, they often operate in 'closed communities' (Otto, 2020). As a result, other businesses, often SME's, not part of these communities and also lacking the capabilities and/or resources to discover, access and process external data remain sidelined (Rehof, 2020; Commission, 2020). This outlines a critical need for an independent data marketplace that can offer the infrastructure and facilitation services to enable a large number of users to engage in data trading while addressing the inherent generic and contextual challenges in monetizing high value business data.

3.3 Multi-sided Data Marketplaces

A digital platform facilitating the interaction and transaction between two groups of users, such as data buyers and sellers, is denoted as a two-sided platform (Spiekermann, 2019). A two-sided platform is considered multi-sided if third party service developers (complementors) can also offer complementary services to buyers and sellers on the platform. De Reuver et al. (2017) found that two-sided platforms naturally developed to multi-sided over time as complementors become increasingly attracted by the growth of the platform. Multi-sided platforms are not a new concept. For example, AirBnb started as a two-sided marketplace, connecting homeowners and travelers but now operates as a multi-sided platform. Multi-sided platforms generate value through enhanced market efficiency by stable matching, facilitating high transaction volume and resource allocation efficiency (Koutroumpis et al., 2020b).

Spiekermann (2019) identifies the following key actors participating in a generic multi-sided data marketplace (MSDM) as: the platform owner, data provider, data consumer and complementors. The platform, actors, their roles and interdependencies is also referred to as a data ecosystem (Figure 5). Although these platforms can be costly to maintain, a sustainable business model based on economies of scale can be achieved if a critical mass of users join the platform (Evans and Schmalensee, 2010). For this reason, MSDM have been an interesting point of debate in the context of enabling data sharing on a large scale (Rehof, 2020). Despite the alleged potential of MSDM's, however, Spiekermann et al. (2019) found that high value data was rarely traded via such marketplaces.

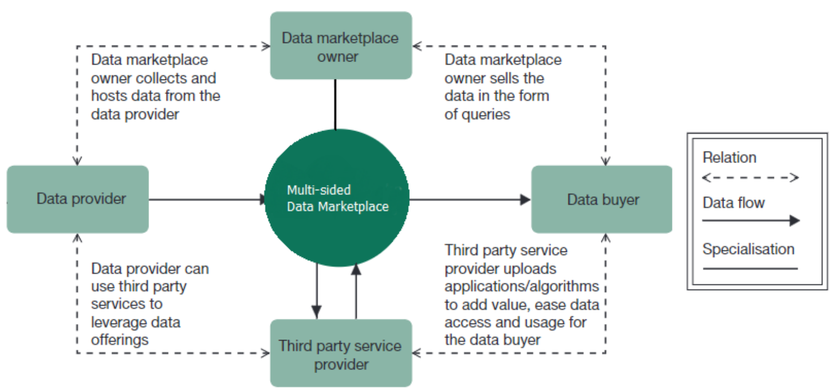

Figure 5: A generic business-to-business multi-sided data marketplace ecosystem (adapted from Spiekermann (2019))

Different authors have shown that MSDM's can be designed and classified in a variety of ways (Spiekermann, 2019; Fruhwirth et al., 2020; van de Ven et al., 2021). Fruhwirth et al. (2020) provides the most complete taxonomy by also identifying three archetypes for MSDM's based on platform design: centralized, decentralized and personal data trading. As this literature review set out to to study MSDM's in B2B context, personal data trading archetype will be disregarded because it relates to a consumer-to-business (C2B) relationship. The key difference between the remaining design approaches is whether the data is stored and accessed from a central storage space managed by the platform owner (centralized) or whether data is stored decentralized at the data provider and transactions are peer-to-peer (decentralized). Decentralized approaches share many basic attributes with centralized approaches but alleviate some of its limitations (Koutroumpis et al., 2020b). In particular, safe and controlled data transactions are now supported without a central structure, thereby technically ensuring data providers retain ownership and usage control over their data sets. Various publications have already established that centralized data storage is not commercially viable for high value B2B large scale data sharing in a multi-sided context (Rehof, 2020; Otto et al., 2019). Even so, decentralized

MSDM approaches constituted only a small part in both Spiekermann (2019) and Fruhwirth et al. (2020) data sets.

3.4 Concerns With B2B Data Sharing

In the section we address the challenges associated with business-to-business data exchange. In the first part the focus will be on techincal challenges (3.4.1). The second part will focus on governance challenges (3.4.2).

3.4.1 Technical Challenges

B2B data sharing is becoming an essential part of the digital economy, with data marketplaces providing a platform for different organizations to share data (Sharma et al., 2020). However, there are several technical challenges that must be addressed to ensure the success of these data marketplaces. In this section, we will explore the technical challenges of B2B data sharing and the measures that can be taken to overcome them.

Firstly, data providers fear the misuse of their data due to a lack of trust and security controls (Koutroumpis et al., 2020b). Data providers invest significant time, effort, and resources in collecting and processing high-quality data, and they need assurance that their data will not be misused. To address this challenge, data marketplaces must implement appropriate trust and security controls to ensure that data is used only for its intended purposes. These controls can include access controls, encryption, data anonymization, and monitoring tools to track data usage. Additionally, data marketplaces must ensure that all participants adhere to strict data sharing agreements.

Secondly, data consumers often do not recognize the value of a dataset until after acquisition, which reduces their willingness to pay for the data (Gelhaar & Otto, 2020). To address this challenge, data marketplaces can provide a preview or sample of the data to potential buyers. They can also establish a pricing model that is based on the value that the data provides to the buyer. This approach will help data consumers understand the value of the data set and be willing to pay a fair price for it.

Thirdly, the creation, processing, storage, and distribution of high-quality data is a significant cost factor for the data provider, and it is difficult to assess expected returns up front (Gupta et al., 2021). This challenge can be addressed by offering a cost-sharing model. For instance, the data marketplace can offer a shared infrastructure for data processing and storage. Additionally, data marketplaces can offer rewards for data providers who contribute high-quality data and discounts for data consumers who purchase a large volume of data.

Fourthly, interoperability is essential for data sharing between different systems (Agahari et al., 2022). It is challenging to ensure interoperability due to the variety of data formats and data schema used by different data providers. This challenge can be addressed by adopting standard data formats and APIs. Additionally, data marketplaces must ensure that all data providers adhere to the same data schema.

Finally, building trust between ecosystem actors and commitment to the growth of the platform is critical (Gelhaar et al., 2021). This can be achieved by providing incentives for data providers and consumers to participate in the platform. For instance, the data marketplace can offer rewards for data providers who contribute high-quality data, and discounts for data consumers who purchase a large volume of data. The data marketplace must also establish a governance framework that ensures the fair and transparent distribution of data and revenue.

3.4.2 Governance Challenges

B2B data sharing has become increasingly common in today's digital landscape, with businesses seeking to gain insights into their customers, markets, and operations. However, this practice raises several concerns, including governance challenges that need to be addressed to ensure a fair and sustainable ecosystem for all actors involved (Lis & Otto, 2021; Hagiu & Wright, 2015; Gelhaar & Otto, 2020). One of the primary governance challenges in B2B data sharing is building trust between ecosystem actors. To share sensitive information, businesses need to have confidence that their data will be kept secure and used appropriately. This requires clear communication, transparency, and accountability between data providers, platform owners, and data buyers.

To incentivize high-quality data and promote data sharing, suitable mechanisms must be put in place to ensure data quality. Incentives such as rewards or discounts for data providers can encourage the sharing of accurate, reliable data, while data buyers can be incentivized to provide feedback and validate data quality. Another critical aspect of B2B data sharing is ensuring that the value generated is shared fairly among all ecosystem actors. This includes data providers, platform owners, and data buyers, who all contribute to the creation of value. To promote fairness, clear revenue-sharing models and compensation schemes need to be established. However, large-scale multi-sided data marketplaces present unique governance challenges, as pointed out by Koutroumpis et al. (2020). In such contexts, platform owners may struggle to retain sufficient control to ensure the integrity of the platform and mitigate strategic behavior by data buyers. As a result, specific governance arrangements need to be identified to accommodate large-scale data trading in multi-sided contexts.

3.5 Development Scenarios

With its strategy for Data (European Commission, 2020), the European Commission (EC) promotes the development of European data spaces to foster higher availability of data pools in strategic sectors and public interest domains. According to a definition provided by EC, a data space refers to a decentralized infrastructure enabling trustworthy B2B data sharing and use based on commonly agreed principles (Nagel et al., 2021). This definition of a data space as used by the EC is similar to our notion of a MSDM decentralized platform design, making it of great interest to this study. While data spaces may be designed in line with domain-specific challenges and legislation, use-cases may also require data sharing across industry domains.

The International Data Space Association (IDSA) proposed a Reference Architecture Model (RAM) that specifies how a data space should look like in order to ensure that the key requirements for data sharing and data spaces are met (International Data Space Association, 2019). These key requirements are trust among participants, interoperability of the data and sovereignty when it comes to re-use of the data through other members in the ecosystem. Nine domain specific data spaces were initially proposed in 2018: industrial, green deal, mobility, health, financial, energy, agriculture, public administration and skills (Scerri et al., 2022). Nevertheless, many data space initiatives are still in early stages and little research is available that evaluates the usability of the RAM in a detailed case study.

3.6 Platform Architecture and Governance

MSDM's are sociotechnical constructs; they are both technical platforms and market intermediaries. Thus, describing an MSDM requires both a technical architecture and a governance structure (Tiwana et al., 2010; Bianco et al., 2014). According to Tiwana et al. (2010); Baldwin et al. (2009), the technical architecture constitutes a conceptual blueprint that describes how the platform is partitioned into a relatively stable core, a complementary set of modules that are encouraged to vary, and the design rules binding on both (Figure 6). Platform governance refers to the mechanisms through which a platform owner exerts influence over app developers participating in a platform's ecosystem (Tiwana, 2014). Whereas architecture helps reduce structural complexity through modulerization, governance helps reduce behavioral complexity. Governance matters as it helps the platform owner shape and influence the behavior of users in a way that their varied contributions form a harmonious whole.

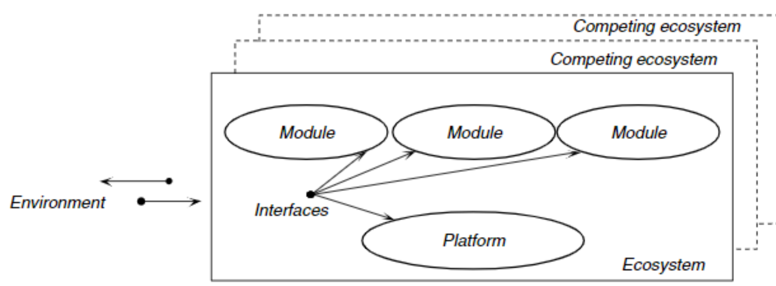

Figure 6: Elements of Platform-Centric Ecosystems (Tiwana et al., 2010)

Studies have shown that MSDM's can be designed and classified in a variety of ways (Spiekermann, 2019; Fruhwirth et al., 2020; van de Ven et al., 2021). From the perspective of archetypes, two types exist based on platform architecture: centralized and decentralized (Fruhwirth et al., 2020). The key difference is whether the data is stored and accessed from a central storage space managed by the platform owner (centralized) or whether data is stored decentralized at the data provider and transactions are peer-to-peer (decentralized). Decentralized approaches share many basic attributes with centralized approaches but alleviate some of its limitations (Koutroumpis et al., 2020b). In particular, safe and controlled data transactions are now supported without a central structure, thereby technically ensuring data providers retain ownership and usage control over their data sets. Various studies have already established that centralized data storage is not commercially viable for high value B2B large scale data sharing in a multi-sided context due to the inherent security risks (Rehof, 2020; Otto et al., 2019). Even so, decentralized MSDM approaches constituted only a small part in both Spiekermann (2019); Fruhwirth et al. (2020) data sets.

Nagel et al. (2021) describes the Reference Architecture Model (RAM) of IDS, which includes the most comprehensive work on architecture principles for a generic MSDM platform infrastructure. The RAM is built on four principles that are derived from a wide range of industry stakeholders concerns (Otto et al., 2019; Otto & Jarke, 2019):

1. a decentralised soft infrastructure enabling direct bilateral as well as multilateral data sharing;
2. data providers should have exclusive self-determination with regard to their data sets (data sovereignty);
3. low entry barriers and a level playing field which ensures fairness and competition on quality of service;
4. a two-tier governance structure. Each data space will have a platform owner that governs the platform, while an overarching central governance authority, i.e. IDS, oversees aspects with interoperability between data spaces.

Table 6: Domain Agnostic Requirements (adapted from Nagel et al. (2021))

	Requirement	**Description**
1	Data Provider Empowerment	About ensuring that decisions can be made by appropriate stakeholders. This means that tools and practices are available for the possibility to: define and monitor policies in data sharing (1), govern the use of your data (2), and connect several data platforms with each other with each retaining control of own operations (3).
2	Data sharing trustworthiness	About ensuring that data sharing operations run according to expected (baseline) requirements. This means that the development of data sharing applications must support: security-by-design (1); privacy-by-design (2) and assurance-by-design (3).
3	Data sharing publication	About enabling data to be published so it can be easily located by data consumers.
4	Data sharing economy	About creating the conditions for data sharing and exchange, requiring: non-financial incentive mechanisms (1), financial incentive mechanisms, including models to determine the value of data as well as to monetise (2) and agreement mechanisms (3).
5	Data sharing interoperability	About providing the ability for all applications on data spaces to create, use, transfer and effectively exchange data. This requires the definition of data exchange API's and data models supporting: semantic interoperability (1), behavioral interoperability (2) and policy interoperability (3).
6	Data space engineering flexibility	About providing the ability to add customised features in data processing applications and platform to enable flexibility in terms of interoperability (i.e. specific interoperability capabilities) - (1), trustworthiness (i.e. specific security, privacy and assurance capabilities) - (2) and data processing (i.e. data processing capabilities) - (3).
7	Data space community	About fostering maximum reuse of data space solutions. This includes building on open solutions (1), ensuring features and solutions can be easily replicated (2), allowing free access to data and marketplace components developed by communities (3) and assurance that solutions will be available and maintained long in the future (4).

Table 15 provides an overview of the domain-agnostic requirements. The generic set of technical and governance building blocks, adhering to these architecture principles, are to be implemented as core elements of any MSDM architecture (Figure 7). ADVANEO is an example of a multi-sided data marketplace built from the principles of IDS RAM (ADVANEO, 2022). Nevertheless, many IDS initiatives remain in early stages. Examining other academic architecture publications reveals that they are mostly conceptual or generic (Parra-Moyano et al., 2021; Banerjee & Ruj, 2019; Sharma et al., 2020; Mohanta et al., 2019). This view is supported by Abbas et al. (2021) who writes that many architecture publications in the domain for data marketplaces failed to establish a link between real-world problems and solutions, as is common in Design Science Research.

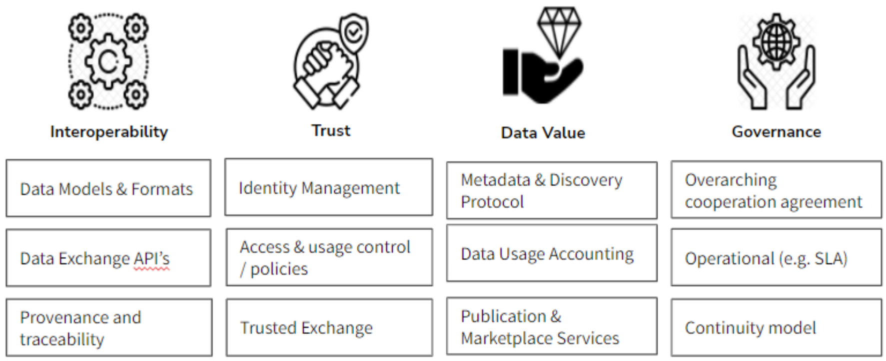

Figure 7: IDS Soft Infrastructure and neutral building blocks (adapted from Nagel et al. (2021))

3.7 Conclusion

This chapter set out to review the current state of research regarding the topic of multi-sided data marketplace architectures. In this conclusion the following SRQ is answered:

SQ1 - What prior knowledge is available on the desired features and characteristics of a multi-sided data marketplace architecture for the logistics industry?

In conclusion, the analysis of the prior knowledge on the desired features and characteristics of a multi-sided data marketplace architecture for the logistics industry has revealed several key insights that are crucial for understanding the current state of research and potential areas for further exploration:

1. First, we have identified that different types of data marketplaces exist, each with its own specific focus and purpose. However, multi-sided data marketplaces stand out as suitable for large-scale data sharing within the logistics industry. These platforms enable multiple stakeholders to engage in data exchange, fostering collaboration and innovation across the sector.
2. Second, we found that there is limited empirical research available on domain-specific multi-sided data marketplace architecture requirements, indicating a gap in the existing literature that should be addressed to develop the architecture principles in this study. This scarcity of research underscores the need for more in-depth investigation and analysis of the architecture principles and features that contribute to effective multi-sided data marketplace architectures.
3. Nevertheless, we discovered domain-agnostic requirements from international data spaces that can be applied as a starting point for designing the domain-specific architecture requirements. These general requirements offer a valuable baseline for creating a robust and flexible architecture that can accommodate various data types and use cases.
4. However, to achieve a comprehensive understanding of the architecture and its suitability for the logistics industry, domain-specific requirements must be considered. These requirements will help ensure that the architecture principles reflect the unique challenges and opportunities within the logistics domain, such as specific data formats, privacy concerns, and regulatory compliance.

4 Formulation of Architecture Requirements

This study set out to investigate what multi-sided data marketplace architecture can facilitate large-scale B2B data sharing for logistics data. We defined large-scale data sharing as: *the process of exchanging, distributing, or making available large volumes of data across multiple organizations, stakeholders, or systems*. There is a global trend and consensus that data should be shared more freely among businesses, and this trend is echoed in the decisions of major funding bodies such as the European Commission (European Commission, 2020). A specific example is the European Strategy for Data, which, among other things, proposes a European way of data governance to facilitate data sharing across sectors and Member States (1.3). While the Data Governance Regulation (Tabel 4) creates the processes and structures to facilitate data, the Data Act clarifies who can create value from data and under which conditions. However, in certain fields, data sharing behavior is still observed at low levels. Logistics is one of the fields where data sharing and re-use can be especially valuable yet seems to be lacking behind (Ceniga & Sukalova, 2015). Multi-sided data marketplaces (MSDM) are a promising concept that can play an important role in addressing this lack of data sharing on a large scale among logistics businesses (Spiekermann, 2019). As no open data sharing platform currently exists for logistics data, conceptualizing a basic data sharing architecture which meets the specific multi-stakeholder context may contribute with novel domain-specific insights as well as provide insight into the possibilities for such platforms in the logistics domain.

Due to the practical constraints as discussed in Chapter 1.5, establishing an architecture design is beyond the scope of this study. The analysis of the research question is thus based solely on domain-specific architecture principles that guide the design and development of the architecture. We formulate these architecture principles based on the domain-specific architecture requirements. A requirement is a property of an artifact designed to meet the need of its stakeholders and one that is used to guide the design and development (vom Brocke et al., 2020). In other words, a set of architecture requirements describe what properties and/or qualities the system should have to meet the needs and/or concerns of its stakeholders. Requirements are diverse in the sense that they can cover a wide range of aspects concerning the architecture; such as desired functionalities, its structure, environment and even effects from using the platform. However, not all requirements are equally important, therefore, classifying them into different categories helps to better understand and prioritize when evaluating conflicting objectives. Through the analysis of the architecture requirements, we are able to make informed arguments to answer the main research question.

In Chapter 3, we conducted a Systematic Literature Review (SLR) to examine existing research on data marketplaces. This allowed us to identify domain-agnostic requirements that can act as a foundation for guiding the design and development of the architecture principles. Domain-agnostic requirements, however, can be applied to any domain or industry. Therefore, while these requirements provide a good starting point, they are not sufficient to address the full domain-specific needs, concerns, challenges and characteristics of the logistics domain. For example, as domain-agnostic requirements are generic, they might therefore not account for specific nuances, terminology and concepts. To address this limitation, it is essential to complement the domain-agnostic requirements we identified in the literature with domain-specific requirements that account for the unique needs, challenges and characteristics of the logistics domain.

This chapter aims to investigate and understand these domain-specific requirements in order to enhance our understanding on what is required to guide the design and development of a MSDM architecture for logistics data. By identifying the concerns and needs of the stakeholders in the logistics domain, it is possible to make a first iteration of the requirements. The remaining sections of this chapter are therefore concerned with elicitation of the domain-specific challenges and concerns from domain experts and transforming their concerns into domain-specific requirements. First, we introduce and substantiate the interview design that served as the framework to guide the interview process (4.1). A total of seven interviews were conducted to elicitate domain-specific concerns and challenges. Second, the resulting interview data is analysed using a systematic and rigorous approach to allow a thorough and reliable analysis (4.2). Third and last, after having conducted, summarized and scrutinized the interview data, the concerns are then translated into a preliminary set of domain-specific architecture requirements (4.3). The following sub-research question is thus answered by the end of this chapter:

SQ2 - What are the key architecture requirements necessary to design a multi-sided data marketplace for the logistics industry?

4.1 Interview Design

Requirements are derived from various sources that reflect the needs and constraints of the platform operator, its users and the specific contexts in which the system will operate. According to (Nagel et al., 2021), the main sources of requirements include domain knowledge, industry best practises and stakeholder concerns. We already gathered data on domain knowledge and industry best practises through our desk research (Chapter 1). Moreover, we built a basic understanding of potentially relevant requirements through our Systematic Literature Review (Chapter 3). Interviews are an important tool for the next step of the requirement elicitation process as they allow stakeholders to share their perspective, needs and concerns in a setting that allows for further clarification of any ambiguity or misunderstanding (Adams, 2015). Therefore, by conducting interviews, we can identify stakeholder concerns that may not have been captured by the domain-agnostic requirements we identified in our Systematic Literature Review and also gain a deeper understanding of the specific nuances, terminology, best practices and concepts unique to the domain. In this section, we present the process and rationale behind the interviewee selection (4.1.1) as well as the adopted interview design methodology for the elicitaton of the domain-specific concerns (4.1.2).

4.1.1 Interviewee Selection

Selecting interview participants for the elicitation of stakeholder concerns to formulate architecture requirements can be a complex process. Specifically, determining the right sample size as well as ensuring the selected participants are representative of the relevant stakeholder groups require careful consideration as they impact the rigor and generalizability of the findings to the various segments. For this reason, this study adopts the purposive sampling approach, or more specific the expert sampling type of the purposive sampling approach. The use of purposive sampling is a well-established approach for requirement elicitation. Purposive sampling, also known as judgmental or selective sampling, is a non-probability sampling method in which researchers deliberately choose participants based on specific characteristics, criteria, or qualities that are relevant to the research question (Palinkas et al., 2015). Purposive expert sampling means focusing on participants who are considered experts or knowledgeable in the topic and domain being studied, ensuring that their input is well-informed and relevant. According to (Palinkas et al., 2015), purposive sampling is particularly useful in research where the aim is to explore, understand, and interpret complex phenomena, rather than making statistical generalizations.

We adopt the use of purposive expert sampling primarily for two reasons. First, experts have a broader understanding of the domain and can provide a more comprehensive perspective on the system's requirements. As the objective of the interviews is to understand the concerns and challenges for a domain-specific reference architecture, which should be applicable for all logistics segments; they can consider the bigger picture, taking into account the various aspects of the system, while actual stakeholders may only focus on their specific needs and concerns. This can enhance the credibility and validation of the research findings as experts may offer nuanced understanding, validate existing theories and/or anticipate on future developments. Second, experts can also be more efficient in providing relevant and concise information due to their extensive experience and knowledge. This can help streamline the elicitation process and reduce the effort to formulate architecture requirements. In line with our reasoning, we specified the following inclusion and exclusion criteria for potential interviewees:

1. Experts should have at least five years experience in logistics, supply chain management or related fields. This could include professionals who have worked in various aspects of logistics, such as transportation, warehousing, inventory management, procurement or distribution.
2. Since the focus is on a data marketplace, experts should have some understanding of data management, analytics and data-driven decision making in logistics. This might include experience in data integration and data quality.
3. Those who are no longer actively engaged in the domain or have no kept up with recent developments are excluded as they may not provide accurate or relevant information.

The aim was to conduct approximately seven to ten interviews with subject matter experts, i.e. domain experts. To recruit the interviewees, a draft email was prepared, which can be seen in Appendix B. At first, a list of thirty-two potential interviewees was drafted using e-mail contact information from websites of research institutions (e.g. university, research groups, relevant research papers). The potential interviewees were contacted in batches of ten at first. However, the response rate was low as only four responses were returned after batch two. For this reason, the list was expanded with twenty additional potential interviewees and all participants were contacted simultaneously. In the end, a total of seven interviews were successfully conducted (Table 7). The column Relation to LSP's refers to the perspective of the logistics service providers, discussed in Chapter 1.2, the interviewees cover. This allows us to understand from which perspective they engage the interview questions.

Table 7: Interview Participants

	Experience	Relation to Logistic Service Providers's
I-1	Assistant professor with over sixteen years of experience in logistics operations research. In particular focused on developing data-driven optimization approaches for logistics transportation.	2PL and 3PL
I-2	Researcher with over than ten years of experience in large-scale multi-stakeholder innovation projects in the area of international trade.	3PL, 4PL
I-3	Professor in freight transport and logistics with thirty years of experience in multimodal freight transportation at urban, national and global level.	2PL and 3PL
I-4	Professor in Information and Communication Technology with experience on IT innovation to facilitate international trade. In particular focused on service engineering and governance as well as multi-agent modeling to develop business procedures in international trade.	3PL, 4PL
I-5	Researcher with over eight years experience in the Dutch maritime sector. In particular focused on Scenario Analysis, Port Community Systems and Information Architectures.	2PL, 3PL
I-6	Practitioner with over ten years of experience in the Dutch maritime sector. Currently specialized in developing practical applications for present-day challenges with data sharing and digitisation in logistics.	1PL, 2PL, 3PL, 4PL, 5PL
I-7	Researcher with over twenty four years of experience in logistics with a focus on information systems architecture. In particular focused on topics related to creating economies of scale and scope through inter- and intra- supply-chain collaboration.	4PL, 5PL

4.1.2 Interview Approach

Interview questions were formulated using the five common guiding principles synthesized from the literature (Rowley, 2012). First, questions should be carefully formulated accounting for any terms the interviewee might not understand or interpret differently. Understanding the terms the interviewee is unlikely to be acquainted with and defining them is essential to avoid confusion during analysis of the data. Second, the questions should not contain leading or implicit assumptions. Third, the questions should not include two questions into one. Fourth, the questions should not invite yes or no answers. Fifth, questions should not include two in one. Based on these principles and the objective of the interviews, a total of seven questions [2] were carefully formulated. The questions are arranged in three distinct sections: general outlook (1), data-sharing concerns (2) and holistic view of the desired platform (3). The full interview questions can be seen in Appendix C.

To get interviewees familiar with the research context, a brief introduction to the working concept of multi-sided data marketplaces and related definitions is provided after the personal introduction with the interviewees. The first section of the interview 'general outlook' provides two broad questions. The first question adopts a helicopter view and inquires the interviewees about their view of sharing and monetizing logistics data through open, multi-sided data marketplaces. Depending on the interviewees response, the second question is more focused and inquires the interviewees on their views of data sharing between logistic service providers who are competitors. The purpose of starting with general questions is two-fold. First, it is to reduce the risk of leading or biasing the participants responses. Highly specific questions may accidentally prompt the interviewee to focus on certain aspects or issues, potentially skewing the data. Second, it is to help establish rapport and trust. By allowing greater flexibility in how the question can be addressed, the conversation can flow more naturally which can yield richer, more nuanced data and can reveal insights that might not emerge through more narrowly focused questions.

[2] excluding the introduction and closing questions.

For the second section of the interview, the interview questions focus on the domain-specific needs, concerns and constraints from the perspective of the platform participants. Three questions were formulated for this section. Before these questions were asked, however, participants were first briefly informed on the type of platform participants that are present on the logistics MSDM. We provided the definitions we used in Chapter 3. This is to reduce ambiguity and misunderstandings. The questions followed a simple structure. First, the interviewee was asked to describe what the concerns of a logistic service provider interested in participating in open, multi-sided data marketplace for logistics data would be. The second question inquired on the same topic, but this time from the perspective of the data buyer. The third and final question asked the participants to prioritize the concerns that were discussed based on relative importance.

In the third and final section, the focus was set on how the concerns discussed in section two of the interview could be addressed through an architecture design. Two questions were formulated for this section. First, interviewees were asked to describe their ideal version of an open, multi-sided data marketplace for logistics data. By encouraging interviewees to envision their ideal platforms, they might share innovative ideas, features or solutions that have not yet been considered. Moreover, it also allows for verification and validation of our understanding of the concerns that were stated previously. The second question asked the participants what type of use-cases could be interesting as a demonstration for the architecture. After having asked the interviewees ideal version of the platform, subsequently asking about interesting use-cases for the demonstration of the architecture can help bridge the gap between technical details and real-world applications. Moreover, understanding which use cases are deemed interesting or important by interviewees can help prioritize development efforts. This can ensure that the most impactful features and functionalities are addressed first.

Initially, the interview questions were refined on paper through feedback from TU Delft research advisors. This was a useful process because it revealed the importance of starting with general questions before narrowing the questions down to concerns. After refinement on paper, the interview questions were first tested in a mock-interview with a TU Delft MSc. student with a background in Transport Infrastructure and Logistics. Afterwards, the questions were also tested with a research assistant from TU Delft who was knowledgeable on the subject of IS platform design and governance. From this process, we learned it might be too demanding for every interviewee to come up with (new) concerns from both the data providers and buyers perspective. Additionally, the question on what their ideal version of a platform might be based on the concerns they stated appeared to be restrictive and inhibit creativity. This is because the participants were more occupied in ensuring they covered the concerns they shared and less about the potential design solutions. For this reason, the questions related to data sharing concerns were complemented with pre-defined follow-up questions related to the domain-agnostic concerns we found in the literature (see Chapter 3). These questions, such as "*What about concerns related to interaction with existing systems?*" would be asked to supplement the interview conversation if the initial response and follow-up question did not yield sufficient depth. The question related to the ideal version of the platform was also adjusted and now participants were not explicitly asked to match the concerns with the potential design solutions. Instead, this information can still be inferred from the data by comparing the stated concerns and envisioned platform functionalities and features.

The interviews with the subject matter experts were conducted in the period of February to April, 2022. The interview lengths varied from 30 minutes to 50 minutes. With the exception of the first interview, all other interviews were conducted in an online setting using TEAMS software. The interview data was recorded and processed based on informed consent. All interviewees were carefully informed of this process and had agreed to sign a informed consent declaration. The transcripts generated from the TEAMS software as well as the audio recording was used to create the anonymous interview summaries, which were subsequently used as input for the interview analysis.

4.2 Interview Analysis

This section presents the challenges discovered during the analysis of the interview data. A necessary precondition to a useful and effective analysis is ensuring that the data documented is organized, accurate and consistent (4.2.1). This allows for the data to be analyzed in a thorough and reliable manner, thereby reducing the risk of errors or misunderstandings, and/or enabling the identification of patterns and trends. The next section presents the operationalization process we used to objectively measure the user concerns from the interviews (4.2.2). From a high level, we identified three themes of information from the data set using our coding approach. This first theme informs on the general outlook of interviewees on the concept of large-scale data sharing in logistics (4.2.3). The second theme explores the concerns of potential users with respect to engaging on a large-scale multi-sided data sharing platform for logistics data (4.2.4). The third and final theme incorporates a holistic perspective and zooms in on the potential design solutions that can serve to address the identified concerns (4.2.5).

4.2.1 Coding Approach

Coding is a central component of qualitative analysis. Codes are created as a means to understand the topic of concern and/or interviewees and their views using a systematic and rigorous approach (Skjott Linneberg & Korsgaard, 2019). Coding distills the data, sorts and bundles them into related segments and allows the researcher to infer meaningful conclusions. In essence, it is about attaching labels to segments of data in order to depict what each segment is about. A working definition given by (Skjott Linneberg & Korsgaard, 2019) describes coding as: "*a word or short phrase that symbolically assigns a summative, salient, essence-capturing, and/or evocative attribute for a portion of language-based or visual data*". Advantages of coding qualitative data are that they offer an effective way of acquiring deep insights, making the data retrievable, and ensuring transparency and research validity (Skjott Linneberg & Korsgaard, 2019). It is therefore important for any researcher to delineate the specific approach to the coding approach and subsequent analysis used in the study.

Several qualitative data analysis approaches exist but they generally differ in their level of abstraction and focus (Skjott Linneberg & Korsgaard, 2019). The question of what coding analysis approach this study needs to use depends on the research objective, type of data collected, available resources and the researcher's own experiences. As discussed, the interview questions were designed with the objective of understanding concerns of prospective data providers and buyers in the context of large-scale B2B multi-sided data marketplaces for logistics data. According to (Mayring, 2023), Qualitative Content Analysis (QCA) approach is one of the more practical ways of identifying and analyzing the content of the qualitative data, such as words, phrases, and themes that are present. Other advantages of QCA are its reliability, validity as well as its simplicity over other alternatives such as Grounded Theory. Figure 8 depicts an overview of the coding approach adopted for the analysis of the interview transcripts.

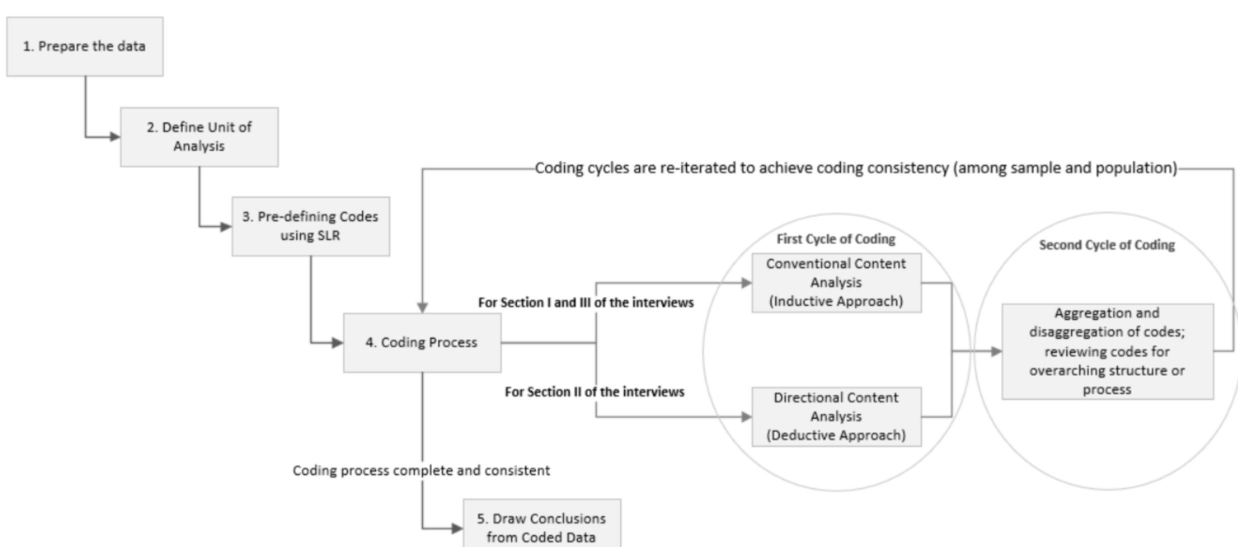

Figure 8: Coding approach executed in this study

First, the data is collected and documented in a form that allows systematic analysis. Linnebeg & Korsgaard (2019) emphasize that the most important part in this condition is ensuring the data is documented in such a way that it allows line-by-line labelling of textual segments in a straightforward and efficient manner. As discussed previously in Chapter 4.1, both an audio recording as well as a direct transcript of each interview audio serve as the basis for creating the interview summaries. Each interview summary is then positioned on the left of a two-column table in MS Word. The second column on the right is then used to specify the codes to segments of text (Figure 9).

So, I mean there probably needs to be some kind of organization that governs it but in a centralized platform, the owner is the most threatening one as they have all the data and therefore the market power. So, Poort8 develops several services and we made a big deal out of it that the dataspaces we are part of are set-up in such a way that any role, including ours, is federated. You should imagine that we as a new start-up setting up these many new data products can be threatening to others as well. Because they may be afraid that we are the one taking the data, so it is also in our interest to take away this threat so they do not have a vendor lock-in with us, but at the same time we want to prevent them from being in a vendor lock-in with any of our competitors, let's say.	

Figure 9: Excerpt from Coding Process: Two-Column Table

Second, the unit of analysis was defined in order to understand the basic unit of text to be classified during coding. In general, the unit of analysis refers to the smallest unit of data that is relevant to the research question and can be meaningfully coded and analyzed. As suggested by (Skjott Linneberg & Korsgaard, 2019), it is important to select a unit of analysis before coding qualitative interview data because it helps to define the boundaries of what is being analyzed and ensure consistency in the coding process. During the interview design, we specified that the objective for the interviews is to understand what concerns data providers and data buyers may have when participating in an open, large-scale multi-sided data marketplace for B2B logistics data. Therefore, the kind of information that is needed to formulate architecture requirements are expressions of concerns, needs and desired solutions of logistics service providers. In other words, our unit of analysis are the individual logistics service providers whose perspective we will collect through domain experts. By analyzing the concerns provided by the domain experts, we attempt to understand the overall concerns and viewpoints of the broader population of logistics service providers regarding their participation in a multi-sided data marketplace.

Third, we reviewed the findings of our systematic literature review to develop a pre-defined list of codes before starting to code the data. Understanding the relevant, existing research literature provides guidance to delimit the area of research and focus effectively on knowledge gaps. During the systematic literature review (SLR) as discussed in Chapter 3, we identified research papers on domain-agnostic concerns in the context of large scale data sharing platforms, however, we found information specific on the logistics domain to be lacking. Due to the lack of information with respect to data sharing requirements and underlying concerns in logistics, we decided to focus the interviews on understanding what the domain-specific concerns and constraints may be. This approach is favoured as it prevents the whole coding process from becoming too complicated and lacking in focus, compared to an open coding approach.

Fourth, creating and adhering to an analytic procedure or a coding scheme increases trustworthiness and validity of the study (Noble & Smith, 2015). Trustworthiness refers to the extent the research process and its outcomes are consistent, rigorous and reliable. Validity refers to the extent to which the findings accurately reflect the phenomenon being studied. It is therefore important for any researcher to delineate the specific approach to the analysis that is going to be used in the study before beginning data analysis. The question of what analysis methodology this study needs to use is answered by matching the specific research objective with the appropriate analysis technique. For this study, Qualitative Content Analysis (QCA) is chosen as the most suitable coding approach (Figure 10). According to (Lindgren et al., 2020), QCA involves a process designed to condense raw data into categories or themes based on valid inference and interpretation. QCA is adopted for the analysis the interview data because it is a well established approach that can be used in conjunction with other qualitative methods, such as interviews, to provide a more comprehensive understanding of the research question Mayring (2023). The goal is to identify underlying meanings and interpretations that participants or authors express in their responses or text.

The fourth step also includes executing the coding cycles, which can consist of two or more cycles based on how extensive the research is (Skjott Linneberg & Korsgaard, 2019). In this study, we re-iterated two coding cycles to analyze the interview data (Figure 10). The first coding cycle consists of formulating and refining the initial codes. Multiple approaches exist for executing QCA in the first cycle of coding. For our analysis, we chose to adopt two different QCA styles for the first cylce of coding (Mayring, 2023). We applied conventional QCA for Section I and III of the interview, and directed QCA for Section II of the interview. According to Assarroudi et al. (2018), directed qualitative content analysis involves using a pre-existing theoretical or conceptual framework to guide the analysis of the data. Conventional qualitative content analysis, on the other hand, involves analyzing qualitative data in a more open-ended manner, without a pre-existing theoretical or conceptual framework The second cycle focuses on the aggregation and abstraction to a higher category based on similarities and differences. Both coding cycles are then re-iterated until appropriate coding consistency within one interview and between interviews is achieved. In contrast to the first and third section of the interview, we already have a built a solid understanding of the various concerns that could arise from existing research. For this reason, it made sense to begin the coding process of the data sharing concerns in Section II with a pre-defined list of codes. This helped speed up the coding process and allowed us to compare and contrast with existing research more easily.

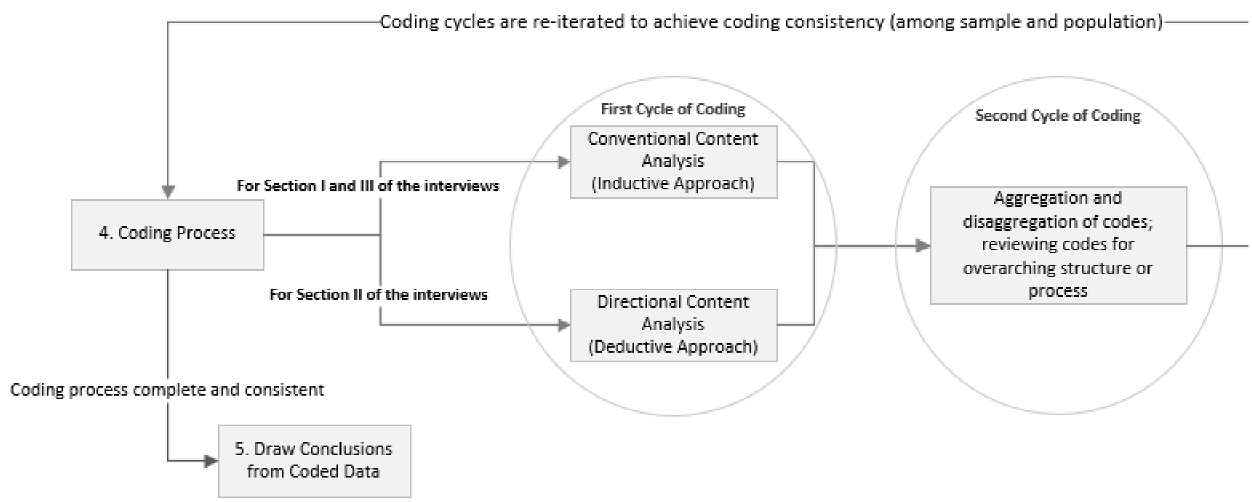

Figure 10: Overview Coding Process

4.2.2 Operationalization of Results

Conceptualization and operationalization are two crucial processes in empirical social research that create a connection between concepts and data. Concepts represent abstract ideas, encompassing the (social) phenomena being studied in the research. Since concepts can have varying meanings, conceptualization is a process where researchers strive to reach a consensus on the meanings of the concepts being studied. Saunders et al. (2018) suggests that the research process may involve conceptualization through formulating a research problem, defining concepts, reviewing relevant literature, and constructing nominal definitions for concept meanings. In this study, we have already completed the conceptualization process. For instance, we presented specific definitions for our research context, such as our definition of large-scale multi-sided data sharing in Chapter 1. We also outlined the research problem and the scope of research activities in Chapter 2, and identified domain-agnostic requirements in the context of open, multi-sided data marketplaces for logistics data sharing through a systematic literature review in Chapter 3. As a result, we established concept definitions to pinpoint the study's focus and describe the research construct. Additionally, we conveyed these nominal definitions to interviewees by providing brief explanations during the interviews.

What concerns this subsection of the report is the second vital process in empirical research: operationalization, which is creating an operational definition for the concept as well as the steps and procedures that can be used in measuring the concept. In the case of understanding logistics service provider' concerns for participating in a large-scale multi-sided data marketplace, operationalization, however, is not necessary for the interview data. The focus of the research is on exploring the concerns and perspectives of the participants, rather than on measuring or quantifying abstract concepts or variables. The data collected through the interviews should already provide rich and detailed distinctive information about the concerns, barriers, and potential design solutions from the participants' perspectives. As we designed the interview questions in such a way that the participants had to list the concerns according to the respective user, we are able to identify them without difficulty. However, in order to differentiate in importance, we did count the number of times a concern was explicitly mentioned. Moreover, we also added a note whenever an interviewee explicitly mentioned the concern was important or crucial. The total count for each concern is later used as input to define the requirement priority in 4.3.2.

4.2.3 Results: Data-to-Value

This first section of the results explores the general view of interviewees on the concept of large-scale data sharing in logistics. In order to identify any topics that might otherwise not be discussed after specifying the context into concerns related to the architecture, the interviewees were asked to describe their thoughts on whether and/or how large-scale data sharing could develop in the logistics domain. By keeping the questions open-ended, interviewees were encouraged to describe their experiences and use examples of specific situations. This provided insight into the feasibility and concerns related to the business model of large-scale data sharing in logistics. Based on the responses and the coding approach discussed, a total of twenty seven labels arranged in four categories were gathered through re-iteration of the coding cycles[3]: barriers to data trading (1) and monetization opportunities (2), examples of platforms (3) and potential use-cases (4).

4.2.3.a Barriers to large-scale data trading We identified four domain-specific barriers that hinder the development of open, large-scale B2B multi-sided data marketplaces for logistic data: unclear value proposition, high degree of stakeholder differentiation, traditional roles and lack of necessary trust. First, it appears that the majority of interviewees identified the absence of a clear value framework as a major barrier to enabling large-scale data sharing in logistics (I1-3 and I6). As I3 states: "*If the whole value framework is not clear, I do not think anything will change*". A value framework provides a structured approach to understanding and evaluating the benefits, costs, and trade-offs associated with different options or initiatives. What the added value of open, large-scale MSDM for logistics data can be for LSP's seems not yet clear according to the interviewees. I3 emphasized that overcoming this challenge should be given utmost priority: "*These things (=architecture considerations) matter of course, so the word 'trusted', 'usage conditions' and 'authorization' could be important but I think first of all the value story should be clear.*". I2 also discussed a similar narrative in relation to an unclear value framework for providing data to interviewees outside the supply chain. As I2 mentioned: "*For logistics it is clear; it (=data sharing) is all used for logistics optimization. So, most of the time a certain party wants to enrich (their) data sets in order to, for example, achieve better predictability. Before it was also interesting for the importers to get this data, but now they receive this data, for example on TradeLens, through their own supply chain channels*". If a logistics service provider can fulfil their optimization desires by engaging with partners from within the supply chain, then what is the benefit of engaging with businesses outside the supply chain? At the same time, perceived commercial risk plays as a demotivating factor due to the many unknowns of this new endeavor. All in all, it comes down to economics according to I3: "*I think there is kind of an efficiency question also. This kind of comes down to cost and prices. This whole system has to be efficient for the data provider, e.g. it simply requires enough revenues and benefits for the costs that are incurred. Behind this 'efficiency', there are questions about connectivity, interoperability, messaging data standards, cost and revenue models et cetera. All these aspects that relate to cost and prices need to be cleared; therefore, if the economics are OK, it would happen*". Next to direct financial incentives, I1 points out that there are also other ways to compensate data providers: "*Returning to the motivation for data providers, the main driver for businesses is monetary incentive. However, in the Netherlands; you should also keep in mind that there are other ways to incentivize, such as joining a club with social value*". In summary, these results show that the value proposition of multi-sided data marketplace are not yet clear for logistics; which may explain the reluctance to share data.

Second, despite large-scale data sharing being a benefit for the social good, allowing innovation and experimentation, the interests of data owners are simply not aligned as they have much to lose while unsure about their gain. As I1 stated: "*Private companies and municipalities all have their own way of thinking so their goal and interests are also not aligned together, 90 percent of the time*". Not only private companies and municipalities; LSP's, as the data providers, also do not fully own the data in question. Therefore, the data owner, e.g. the owner of the goods in transportation; also serves their own interests which may not be aligned with the LSP providing the data to the marketplace. This makes the situation complicated. I2 experienced this hands-on in a project on aggregating logistics operations data for experimentation, the importer (= owner of the goods in transportation) became a major barrier unwilling to supply its data: "*When we had discussions with a fruit importer, they said that they did not want the platform to even see what their data is on an aggregate level as they were concerned it could disclose their import routes to competitors.*

[3] We included the discussion of the codes arranged in examples of platforms and potential use-cases in monetization opportunities

So it is not only about the specific transaction but even if you aggregate it and mention that they import through Africa instead of Asia, it already discloses a lot of information. Therefore, there is a lot of commercial sensitivity in supply chains which they don't want to share". In logistics, it appears that the owner of the goods in question play a significant role in whether the data sharing initiative project can be completed at all. As I1 put it: "*You need to think about this in data marketplaces: who owns the data. And of course, if traders (= LSP's) have all the data; they could start selling it left and right, that will be really nice. However, they don't own the data and also what you see is that for the part that they (=data owners) agreed to join, they said. Well, I joined to get the benefits for myself (=logistics optimization), but they don't want their data to be aggregated, made available, sold etc.*". Overall, this shows that, regardless of whether the data provider is willing, the data owner's interest may be completely different and it is necessary to make clear how the owner of the goods stands to gain and how its assets are protected.

Third, while there is an on-going shift in the industry to adapt to new technologies due to increased competition and changing consumer demands, interviewees also pointed out that LSP's, especially SME's, mostly still operate according to traditional roles. According to I1, several factors explain this. One explanation for this is due to the characteristics of the industry: high volume and low margins. As I1 experienced first-hand on a data sharing project in the on-demand delivery sector for a large food delivery business and logistic service provider: "*..traditional logistics remains old-fashioned, so not much data or advanced analytics there. For new ones, emerging businesses in logistics, they work based on data and they can easily collect data, so they have a lot of data. But these ones (=new logistic businesses), they don't want to share. Moreover, also because the margin is low*". This experience seems to be similar with I7 who worked with various SME's at the Port of Rotterdam: "*All things we can think of in theory about data, the value of data sharing etc. is really far from practice in logistics, especially for small and medium enterprises. They usually have few employees that work really hard to solve today's and yesterday's problems but nobody is looking ahead basically*". So what appears to be an unresolved question is that if traditional LSP's are not into data trading or whatever data they need to optimize their operations originates from their supply chains; then how is it natural for LSP's to start trading data outside their supply chains? This view was echoed by I4 who suggested that new-comers may be what just what the industry needs: "*Many people argue like Flexpoort or Amazon Logistics that they say well, if there will be a big data-driven organization that is really making steps; then the other parties in the chain will also start moving. For instance, in making standardization, using platforms et cetera. Probably these are the two main elements (= standardization + new-comer). If there are really data driven companies that show you can really do more efficient and effective business using data platforms, that will definitely have huge impact*". I4 pointed out, however, that there may be parties such as freight forwarders already that perform a similar role: "*On the other hand, you know, for instance freight forwarders, they do a lot of matching and planning. This group might feel threatened by such initiatives so you have to think how to overcome such concerns and have them participate*". The establishment of collaborative relationships among participants in the marketplace is critically dependent on trust. Without trust, logistics service providers may be hesitant to participate in the data sharing platform, leading to limited availability and diversity of data offerings, and ultimately slowing down adoption.

Fourth and final, another major challenge is overcoming the lack of trust from engaging with businesses whom you not had earlier engagements with. I4 points out that transparency in data within supply chains is already met with intense reservations. So how would the situation be for sharing data outside the supply chain? I3 explains that LSP's are unwilling to engage with what they see as blind trust: "*With blind trust, you do not know anything or little. Moreover, there are also no standards and it is hard to anticipate the return. Therefore, you just have to trust and be careful. I think that we are in this situation currently, as there is little known about the real business benefits and risks of sharing logistics data. Unless it would be a bilateral exchange, these are the reasons why open platforms fail*". I1 pointed out, however, even in bilateral exchanges, gaining access to data is not a straight-forward task: "*I think the value and the insights you can obtain from this data, for some cases, is high. It has a commercial benefit for whoever uses this data. But, even for our project, however, they didn't share this data with us directly. We have access to the data through their data center. They give us a laptop to connect to their system, so it was not straightforward. Moreover, they anonymized the data and even then the access was restricted and they did not want to share anything from their business*". The establishment of collaborative relationships among participants in the marketplace is critically dependent on trust (Stahl et al., 2016). Without trust, logistics service providers may thus not only be hesitant to participate in the data sharing platform; they may also implement erect barriers to extract value, leading to limited availability and diversity of data offerings, and ultimately slowing down adoption.

4.2.3.b Monetization opportunities Despite the challenges mentioned above, a recurrent theme in the interviews was a sense amongst interviewees that monetization opportunities should exist. From a high level perspective, value creation in logistics optimization can come in one, or a combination, of three ways (Karam et al., 2021). **First, through horizontal integration of data sets**. This refers to the process of combining data from different sources or systems that operate at the same level as the logistics service provider. For example, combining data from multiple warehouses or distribution centers. **Second, through vertical integration of data sets**. This refers to the process of combining data from different levels within an organization or industry. In this context, this means combining data from different stages of the supply chain. For example, combining data from production, distribution, and retail stages of a product's supply chain. The third and final way is external integration. This refers to the process of combining data from external sources such as weather and traffic data to gain an understanding of the factors influencing a certain logistics process. This type of data is also considered open-data – See Chapter 1.2 for a more detailed review. Many recent studies such as Spiekermann (2019) have shown that closed data sharing platforms related to logistics currently focus on the vertical and external integration of data sets. Horizontal collaboration is a relatively new field of research that is gaining attention in recent years. However, despite its perceived benefits, it has yet to see much success (PwC, 2019). To better understand what use-cases might be more relevant. Interviewees were inquired on what use-cases could be interesting to explore as a demonstration for an open, large-scale MSDM platform. Through the interviewees responses, several similar existing platform initiatives also came to light that we could draw lessons from, e.g. TradeLens and PortBase (Table 8).

Table 8: Logistic Platforms

Use case	**Description**	**Integration of Data**
TradeLens[4]	Blockchain-based digital platform for global trade and supply chain management developed by IBM and Maersk, providing real-time access to shipment information, documentation, and other essential data for all stakeholders involved in the supply chain to improve efficiency, transparency, and security in the shipping industry.	A, B, C
Port Base[5]	Digital platform that enables efficient, transparent, and secure management of port activities by facilitating data sharing and collaboration among stakeholders in the port community related to cargo handling, terminal operations, and hinterland transportation.	A, B, C
National Single Window[6]	Digital platform that simplifies cross-border trade by providing a single point of entry for businesses to submit all required regulatory documents and interact with multiple government agencies, streamlining customs processes, reducing administrative burden, and enhancing transparency in international trade.	A, C

Legend: A = Horizontal, B = Vertical, C = External

What these systems have in common is that they revolve around a specific context or logistic chain of operations, e.g. TradeLens based on global Maritime shipping or Port Community Systems based on port sectors (e.g. general cargo or in-land bulk shipping). However, they are not interconnected; meaning the data streams in these platforms are silos and can not be integrated. As a result, a trucking company would likely need to find ways to participate on multiple platforms at once. I2 sees an opportunity in how the MSDM could be positioned versus existing platforms: "*So, I do see a lot of possibilities of course. As it hasn't really been conceptualized somewhere, it might be interesting to see how it can be linked and positioned between these big platforms such as TradeLens. I think it is still interesting to conceptualize it somewhere, it is not a typical marketplace; it is something else. It will probably have possibility to provide an infrastructure on top of which you could build marketplaces or something else, but it conceptually is a bit different I think*". On the question of what use-cases could serve as an interesting demonstration for the architecture, interviewees proposed various use-cases that can be interesting to explore (TABLE 9). The relation of these use-cases to this study is later discussed in the Chapter 6.

[4]https://www.tradelens.com/about

[5]https://www.portbase.com/en/about-us/

[6]https://taxation-customs.ec.europa.eu/eu-single-window-environment-customs_nl

Table 9: Use cases

Use case	Description	Integration of Data
CO2 accounting	The process of measuring and tracking the carbon footprint of a supply chain to promote eco-friendly practices, reduce emissions, and make informed decisions on reducing environmental impact	A, B, C
Final mile delivery	The last stage of the delivery process where goods are transported from a transportation hub to the final destination, and optimizing this process with real-time data can improve delivery times and customer satisfaction.	A, C
Counterfeit monitoring	Multi-sided data marketplace architecture to detect and prevent counterfeit products from entering the supply chain, ensuring product authenticity and protecting brand reputation through data analytics and tracking systems.	A, B, C
Synchromodality	Logistics approach that optimizes transport modes and routes in real-time based on cost, time, and environmental impact through a data-driven marketplace architecture.	A, C
Port Community Systems	Digital platforms that enable stakeholders in a port community to exchange information and collaborate, streamlining port operations, enhancing efficiency, and improving cargo handling processes.	A, C
Carbon trading	Market-based approach to reduce greenhouse gas emissions, where businesses buy and sell carbon credits based on their emission allowances, and a data-driven marketplace architecture can facilitate this process by providing accurate emissions data and ensuring transparency in the trading process.	A, B, C
Circular economy	Economic system that aims to eliminate waste and promote the sustainable use of resources by implementing practices such as recycling, remanufacturing, and product-as-a-service models, and a data-driven marketplace architecture can support this by providing data on resource flows, waste management, and product lifecycles to enable sustainable logistics operations.	A, B, C
Zero-Emissions Logistics	Reducing greenhouse gas emissions in the logistics sector by adopting clean energy sources, eco-friendly vehicles, and efficient transportation practices, and a data-driven marketplace architecture can help monitor and manage emissions data, identify opportunities for improvement, and support the transition to sustainable logistics operations.	A

Legend: A = Horizontal, B = Vertical, C = External

What seems apparent from these use-cases is that they revolve around transforming specific data sources into information that provides added value. It might not be data that will be traded on logistics data marketplaces, but rather information. As I6 also experienced when attempting to monetise data sets: "*Commercializing data itself is pretty difficult. I tried this myself in my previous capacity when I was an employee. It is pretty challenging and it is not really developing fast, as fast as I would have hoped. At the same time, there are quite a few applications that would fit this. Not by selling data directly but creating additional services, such as federative ways of booking, track & trace, CO2 accounting which is becoming big in the next years*".

4.2.4 Results: Data Sharing Concerns

This second section of the interviews explores the concerns of logistics businesses that would engage in a large-scale multi-sided data sharing platform for logistics data. Based on our knowledge from previous research (see Chapter 3), we noted several domain-agnostic concerns ahead of the interviews that could be discussed by the interviewees. Interviewees were encouraged to freely discuss any concern that may come to mind within context. The inquiry only made specific distinction between concerns of data providers and buyers. **During stage two of the coding cycle, however, we identified concerns shared by both the data provider and the data buyer**. Therefore, based on the responses and the coding approach as discussed, a total of twenty-two labels arranged in three categories were gathered through re-iteration of the coding cycles: data provider - (1), data buyer - (2), and shared concerns (3).

4.2.4.a Data provider concerns The first category covers the concerns related exclusively to data providers. In a multi-sided data marketplace, a data provider is an individual or organization that offers and provides data to other interviewees in the marketplace, such as data buyers or other data providers. In order to identify potential concerns, the interviewees were initially asked to describe what concerns a data provider may have if they would engage in a large-scale MSDM. Upon mentioning concerns, follow-up questions were asked to extract more information on the context. The concerns that were brought to light can be found in Table 10.

Table 10: Elicited data provider concerns

	Concern	**Mentions**	**Description**
1	Loss of competitive advantage	I5, I4, I2, I1, I7,	Concerned with losing competitive advantage over sharing data that may inadvertently disclose sensitive information to competitors.
2	Shipper's ownership rights	I6, I3, I2, I1,	Data providers do not always have ownership of the data they provide, so there is a concern that data owners will feel they have no control or insight in what is happening with the data.
3	Data confidentiality	I3, I2, I1, I6, I7,	Even if the data is shared, data providers are still concerned of unauthorized access by malicious actors.
4	Usage control	I5, I6, I3, I7,	Concerned with the data being used in a way that was not agreed upon.
5	Legal concerns	I6, I1,	Unaware of the relevant compliance procedures on laws and regulations that govern the collection, use, and sharing of logistics data.
6	Large investment for providers	I2, I1, I6,	Concerned with the high cost aspect of the initial investment. As the logistics is a cost-sensitive industry, with tight margins, logistics service providers must minimize their costs to remain competitive.

As noted by the majority of the interviewees (I1, I2, I4, I5, I7) and corroborated by Chapter 3, the biggest concern that was echoed by the interviewees is the fear of losing unique competitive advantage. According to I4, there is always a built-in tension between the data provider and data buyer. The concern of losing competitive advantage arises when a logistics data provider shares their business-sensitive data, such as unstructured operational route data for example, it may lead to information that can be used by others to gain an advantage in the market. Data providers fear that sharing this information can lead to imitation or copying by competitors, resulting in a loss of competitive advantage. According to I3, even sharing data with customers can have negative implications: "*What we have seen from various pilots is that there is a danger when you signal that you have overcapacity. What may happen then is that customers start negotiating down the prices*". This may relate to any type of information such as pricing, customer preferences and supply chain strategies. Therefore, it is important to carefully consider which data to share and with whom to maintain a competitive edge. However, according to I5, this seems to be improving: "*I think that more and more the industry is becoming aware that data can be shared without necessarily providing all (=full access), because I think you can also choose which data to share and which not to share. Even if the data is there, it could be prepared in such a way that not necessarily anyone will gain competitive advantage or lose competitive edges by means of sharing data. But it is definitely still one of the main hurdles to overcome in systems such as you are describing (=multi-sided data marketplaces)*".

At the same time, even if data providers are willing to share data; ownership rights are also a concern to address (I1, I2, I3, I6). As I3 noted: "*Another failed example that I forgot to mention is a sandbox project for researchers of open data in the port of Rotterdam. There was a plan to create a big data lake of all kinds freight data which researchers were able to use against favorable conditions. This, however, never materialized because companies just did not want to do it. The one who had the data was not the owner; the owner of the data was too far away and had no relationship at all with the data lake*". At the heart of logistics, LSP's adopt the responsibility of executing a specific part of the customers supply chain. LSP's as data providers, therefore, have access to valuable data related to their customers, partners, or suppliers that do not want their data to be disclosed to third parties without their consent. Informing them adequately and receiving consent to use their data is both sensible from the perspective of customer relationship management as well as required by data protection regulations, i.e. Data Governance Act. Receiving consent from the data owner can be more complicated than it appears, however. As noted by I2: "*You need to find the reason to why, for example, an importer will be willing to share his data when there will be some (logistics) service provider that will be making money on top of his (=importer) own data*". It is a valid remark. Specifically in the logistics domain, an important concern for LSP's is how to secure that the data owner sees a benefit of sharing considering their position. What I2 was able to share based on an earlier experience with a project related to fruit farmers is: "*They (=fruit farmers) were careful in sharing their data. They did not want even aggregate data to be shared or sold in any way*". Taken together, these results suggest that incorporating ownership rights and interests will be a crucial but a multi-dimensional aspect to address.

Another important consideration is the perceived risk of third parties receiving access to the data in any way (I3, I2, I1, I6, I7). Even in cases where access to sensitive data is restricted to authorized users, there is still a relatively high perceived risk of data confidentiality breach through either unauthorized access or the disclosure of data contents to third parties. In Chapter 3, we identified that data can easily be replicated or shared at little to no additional cost. The concern that was echoed by the interviewees is how to ensure data remains inaccessible by unauthorized parties. As I1 mentioned, even in a project where third-party researchers were invited to work on optimizing the logistics capacity of PostNL for the delivery of food, businesses still want to minimize any risk of data leakage: "*They(=PostNL & Thuisbezorgd) didn't share this data with us directly. We have access to the data through their data center. They give us a laptop to connect to their system, so it was not straightforward. Moreover, they anonymized the data and even then the access was restricted and they did not want to share anything from their business*".

Another important consideration, which is also corroborated by our findings in Chapter 3, is about losing control of the data set after having shared it. Interviewees pointed out two primary reasons. First, data providers may fear that once they share their data, they may not be able to control who accesses it, how it is used, or whether it is shared with third parties. Not to mention, if data providers are not able to control the data and its usage, they may also be unable to ensure its accuracy or completeness, which can lead to incorrect business decisions. This loss of control can lead to unintended or undesirable consequences, with reputation or financial loss as a result. Second, if the data provider is unable to exert control over the data set if they share it, then how can the enforceability of contractual obligations be guaranteed? In such cases, the data provider would feel vulnerable and hesitant to share their data. As I1 put it: "*When businesses share their data, they become vulnerable. What happens in the future? It is not only related to the business but, possibly, also other things such as law aspects. By exposing their business data, however, perhaps it is possible, for a lawyer to sue them based on the data. This might be problematic. It is possible to anonymize everything, but businesses still do not know for sure what else can arise. This may bring them legal issues in the future*".

Specific to the logistics industry is the vast number of applicable regulations LSP's must manage such as the General Data Protection Regulation (GDPR), Dutch Customs Act (DCA) and the Dutch Transportation of Dangerous Goods Act (DTDGA). Although the specific regulations and data requirements vary depending on the nature of the logistics operations and the types of goods being transported, LSP's are subject to various regulations regarding data sharing with the government (I6, I1). For example, the government may require access to certain data in order to comply with the applicable regulations. This data can include information about the bill of lading[7] or the environmental impact of the operation. As I6 said: "*LSP's are not really concerned for a data platform which they can use for a specific need. They are interested in sharing data with them as the center. They deal with so many different partners that it becomes cumbersome to develop a tailored approach for each endeavour*".

Potentially the most impactful concern shared by the majority of the interviewees is the cost-sensitive aspect of the logistics industry (I1, I2, I3, I6). Logistics is a cost-sensitive industry, with tight margins; meaning LSPs must minimize their costs to remain competitive. Participating in a large-scale multi-sided data marketplace requires significant investment in technology infrastructure, data management and security, as well as human resources. According to I1, money dictates rule in logistics. LSPs are therefore concerned about the cost of acquiring, processing, and sharing data, as well as the cost of maintaining their IT systems. These costs can be especially large for smaller LSPs, which may have limited resources and capabilities to participate in such marketplaces. As I6 put it: "*All things we can think of in theory about data, the value of data sharing etc. is really far from practice in logistics, especially for small and medium enterprises. They usually have few employees that work really hard to solve today's and yesterday's problems but nobody is looking ahead basically*". Thus, the cost sensitivity concern of logistics data providers relates to the potential financial burden and risks of participating in large-scale multi-sided data marketplaces.

[7] A bill of lading (BL or BoL) is a legal document issued by a carrier (transportation company)

4.2.4.b Data buyer concerns Data buyers in essence refer to the businesses who make use of the insights extracted from the data set. During the interviews six concerns were highlighted, which can be found in Table 11:

Table 11: Elicited data buyer concerns

	Concern	Mentions	Description
1	Low data quality	I6, I3, I2, I1,	Concerned with inaccurate, incomplete or outdated data resulting in misinformed decision making.
2	Low data coverage	I6, I2, I1,	Concerned there will be an insufficient variety of logistics service providers offering data thereby limiting the application of the data assets.
3	Pricing of data sets	I6, I1,	Concerned with overvaluation and inconsistent pricing of data assets.
4	Real-time data processing	I2, I7,	As time sensitivity is a key aspect of logistics, the need for real-time processing and analysis of data was expressed.
5	Insufficient data supply	I6,	Concerned there will be inadequate amounts of data in the same format and context to conduct comprehensive and rigorous analysis.
6	Understanding data sets	I2,	Concerned that understanding and generating insights from the data sets is difficult due to the sheer volume and lack of context.

As a data buyer, interviewees have expressed that data quality is considered to be one of the bigger concerns (I1, I2, I3, I6). Data quality ensures that data is tailored for use by data consumers (Adams et al. 2019; Lawrenz et al. 2019;). As I2 said: "*Getting it (=data) with the right quality is a challenge. Because we have also seen in projects where I worked, the data quality was horrible and you could hardly do anything. You (=data buyer) also don't understand the (full context of the) data well so I mean monetizing even on these data sets was difficult*". Inaccurate, incomplete, i.e. missing fields, or outdated data does not only negatively impact operations, it also complicates the returns of the data acquisition. Furthermore, data cleaning and processing would also require additional resources, which adds to the overall cost of the data acquisition and value extraction process. I1 experienced similar challenges with getting the right quality from projects for a large LSP: "*But, even for them, the problem that we had was that the data was scattered everywhere; ten to fifteen databases that they would try to merge together in order to retrieve the data. Data quality was a major issue*".

Several interviewees indicated that low data coverage is also a concern (I6, I2, I1). Low data coverage may occur when there are a limited number of LSP's providing data in the marketplace, or when the participating providers only cover a small geographic area or a specific type of service. As I6 said: "*What about the data coverage? Typically information is dispersed throughout the market so if you want to have information on vessels, let's say, you may need to combine data from 10 to 15 different sources. If you buy a data set about something like this, then, an important question becomes whether the dataset has sufficient coverage*". Therefore, even if there are LSPs operating in the area of interest, they could not be collecting or providing data on the specific type of product or service that the data buyer is interested in. This means that there is a gap in the available data, even though there may be some data available. I1 experience of attempting to retrieve full data coverage When data coverage is low, the data may not accurately represent the entire logistics ecosystem and may not provide a complete picture of trends, patterns, and insights. This can lead to incorrect or incomplete decision making, and can result in missed opportunities or increased risks. Data buyers may also be hesitant to invest in data platforms that has low coverage, as it may not provide sufficient value for their needs. Therefore, it is important for logistics data providers and platform owners to ensure that data coverage is comprehensive and representative of the focus areas as a whole.

Two interviewees also expressed the pricing of data sets to be a concern (I1, I6). I1 noted: "*Logistics businesses typically run on low marginal profits. Therefore, if they want to buy data; it is relatively high in comparison to their overall profit. For this reason, they do not have a lot of money to acquire external datasets*". To make matters worse, according to I1, it appears that there can be a large discrepancy between the perceived value of the data between the perspective of a data provider and buyer: "*..there is a company who sells data related to cities. For example, traffic data or Google data on which areas are lively but also which restaurants are attractive based on user reviews and online engagement. What you can see here is that they (=data provider) charge a high price for this data. I don't exactly remember the exact amounts but in the line of €50,000 to €100,000. A huge amount of money just to give access to this data.*". Pricing of data sets is be influenced by various factors such as data quality, data source, demand, and supply. This variability can hinder data buyers' ability to plan and budget effectively for data acquisition. This issue is further complicated when the data sharing arrangement involves multiple stakeholders each contributing a part of the full data set. As I1 experienced first-hand, it results in the overvaluation, rather than a undervaluation, of the data assets, leading to a loss of trust from the data buyers: " Moreover, I1 emphasized that traditional LSP's are already tight on resources to acquire external data sources, therefore

One interviewee argued that inconsistent data supply may also be another concern for data buyers. It shares similarities with the low data quality concern, but they both refer to slightly different issues. Insufficient data supply refers to the concern whether the data is received consistently in the same format rather than whether the available data sets are representative of the logistics operation. As I6 said: "*Is the definition clear? There is usually a lot of different concepts in logistics where same things can be called differently or different things are called the same in between different parties*". If the same logistics data differs significantly across providers or over time, data buyers will only struggle to assess the reliability and accuracy of the data.

Data buyers with limited technical know-how will also struggle to analyze and extract valuable insights from the acquired data. Even if all other aspects are covered, there will still be challenges with using advanced analytics tools, machine learning algorithms, or even with the integration of the insights into own systems. According to the experience of interviewee 2 who received large amounts of unstructured data from an importer, making sense of the large volume of unstructured data is a major task . As I2 put it: "*..cleaning the data and trying also understand the data is a huge challenge. Also the whole idea was to link datasets. So we were buying data from different providers to match the data, find discrepancies and potentially create value for risk analysis for example. But it's almost impossible to link these datasets*". Specifically as data buyer with no or little connection to the data source, making sense of the sheer volume and variety of the data can be overwhelming. Making sense of large amount of data sources is a complex process which requires integrating multiple data sources to get the necessary information. How to effectively and efficiently extract information from different data sources is therefore an urgent concern for the data buyer.

4.2.4.c Shared concerns There were, however, also shared concerns shared between data providers and buyers. These four concerns can be found in Table 12

Table 12: Shared concerns

	Concern	Mentions	Description
1	Heterogeneous IT systems	I6, I4, I1, I7,	Many different IT systems such as ERPs, CRMs, TMS, WMS, SCM amd GPS generate relevant data in logistics and therefore will require interoperability to fetch data and upload insights.
4	Lock in effect	I6, I4, I7,	Concerned that using the platform causes dependency to the point it becomes difficult or costly to switch to an alternative platform or third-party service provider.
2	Lack of in-house technical capability	I6, I2,	Concerned that there is inadequate technical capability to setup and maintain intensive data-sharing initiatives.
3	Neutral platform ownership	I6, I5	Concerned of conflict of interest from the platform owner resulting in unfair market competition if the platform owner is competing in the market.

An repeated and explicitly emphasized shared concern is the need for integration with existing systems (I6, I4, I1, I7). As noted by interviewees, LSP's employ a variety of software tools, platforms, and technologies for their day-to-day operations. These systems, as discussed in Chapter 1, include among others Enterprise Resource Planning (ERPs), Customer Relationship Management (CRM), Transport Management (TMs), Warehouse Management Systems (WMS), Supply Chain Management Systems (SCMS) and IoT devices and sensors. All these systems generate different types of data that require integration in order to develop a complete picture of the logistics operation. The intensity of this task is worrying. Not only is the way data is documented different between the various systems, data standards also differ between different providers and legacy systems. According to I7, this is changing slowly, such as with the Open Trip Model, which has gained popularity in recent years. In the past, planning algorithms like those from ORTEC or Quintiq were not interoperable due to their proprietary semantic standards. However, adopting a common standard like the Open Trip Model makes it easier for logistics service providers to switch between planning algorithms and collaborate with competitors, regardless of the software they use. According to I7 own words: "*This facilitates the development of new business models and enhances collaboration. The key factors driving this change are trust and the ability to quickly establish interoperability between different systems*". There is no consensus yet of how this integration needs to be achieved, however, interviewees agreed that a logistics MSDM architecture must have some adaptable solution to incorporate this concern. As I6 put: "*It is line with what I said at first. Any role in such a scheme, not only the marketplace but I would recognize other requirements as well; they need to be set-up in such a way that they are interoperable*". If not, the costs associated with the additional effort to map, validate and transform data between various formats as well as the inefficiency and complexity related to this task create a significant barrier.

Regardless of whether common data models or procedures exist for seamless integration with existing systems, a concern echoed by the interviewees for both data providers and buyers is the lack of in-house capability, specifically for medium- and small enterprises (I2, I6). According to I6: "*All things we can think of in theory about data, the value of data sharing etc. is really far from practice in logistics, especially for small and medium enterprises. They usually have few employees that work really hard to solve today's and yesterday's problems but nobody is looking ahead basically*". This means that platform users, such as SME's, may face potential challenges in levering the platform effectively due to their limited technical expertise.

Neutral platform ownership is a concern related to when the platform users come from different industries or competing organizations (I6, I5). As I6 said: "*..in a centralized platform, the owner is the most threatening one as they have all the data and therefore the market power. So, ..we made a big deal out of it that the data spaces we are part of are set-up in such a way that any role, including ours (third-party service provider), is federated.*". Neutral platform ownership refers to the concern that a data marketplace or platform should be owned and operated by an unbiased, impartial entity to ensure fair and transparent access, management, and distribution of data and services for all users. This concern arises due to the potential for conflicts of interest or anti-competitive behavior when the platform is owned by an organization with vested interests in the industry, particularly if it is a dominant player or has direct influence over the data being exchanged. As I5 put it: "*platform owner may misuse their position...A discussion that I had recently with ...from the Port of Rotterdam about the platform owner being a public function: it does not necessarily need to be a private company that wants to maximize its profit. It could also be done by a nonprofit organization, which, for example, is also Port base. Port base is a subsidiary of the Port of Rotterdam and, of course, they need to make enough money to survive but even if they try to maximize, it will be reasonable and likely with more careful consideration*".

Several interviewees agreed that a shared concern appears to be whether the resources invested are platform-specific and may not benefit other data-sharing projects (I6, I4, I7). The lock-in concern refers to the apprehension that interviewees may become overly reliant on a specific platform or technology, making it difficult or costly to switch to alternative solutions or vendors in the future. This applies not only to marketplace providers, but also to third-party service developers. As I4 experienced: "*..switching, from one provider data to the other often requires different interfaces because they all use their own IT systems. This also has a commercial aspect to it, however. An incentive is that if you make it non-standardized, then you achieve lock-in effect so there is a kind of built-in incentive not to standardize. But, what you see is, under the pressure of the market, more and more messages are standardized. So, in the future the expectation is that there will be much more standardization*". I6 shares a similar perspective: "*If I would change to a different marketplace provider, what will it mean? Would I lose my technical connections or can I reuse them?*". While vendor lock-in is not the biggest worry of platform users, it is somewhere up there I6 states: "*First, if you have the trust network, you know who you are dealing with. Let's say this is the pre-condition for starting the data sharing. After this, however, you start to think about how you can do it in such a way that you are not locked-in with the vendors*".

4.2.5 Results: Platform Design

The first section addressed the concept of large-scale data sharing in logistics from a business problem perspective. The second section focused on understanding what concerns future stakeholders might have and what problems must or should be addressed by the platform design. In this third and final section of the results, interviewees were encouraged to describe their ideal version of the platform. Building on the challenges and concerns discussed previously in the interview, interviewees responded by describing the platform in terms of functionalities and features. Based on the responses and the coding approach discussed, a total of twenty-five labels arranged in two categories of design considerations were gathered through re-iteration of the coding cycles: architecture - (1) and governance components (2).

4.2.5.a Architecture Components Architecture components refer to the structural elements that make up the system or platform, which includes preferred functionalities for example. Based on the inquiry, a total of five architecture related design considerations were: access management (1), federation (2), tokenization (3).

Access management received the highest attention among all interviewees to alleviate the concerns of access and usage control. As I6 emphasized: "*I think the first thing you would need to have in place is a trust network. This is a combination of three of those twelve blocks referred to in OpenDEI – a combination of identification (authentication) and access management (authorization) building blocks. Connect this to the data exchange that is being made. It allows for more scalability and switching between providers. Moreover, it ensures you know who you are dealing with. This takes away few of the concerns*". Access management refers to controlling and managing user access to data and information systems. While we wish to highlight the specific related mechanisms here, it is a set of policies, processes, and technologies that govern the access of users to specific resources and functionalities of the system. The goal of access management is to ensure that only authorized users are granted access to sensitive data and functionalities, while unauthorized users are prevented from accessing them. Access management typically involves the use of authentication and authorization mechanisms, such as usernames and passwords, access control lists, and role-based access control. According to I7, authorization and authentication are also the basic functionalities which must be in place: "*I think if we have functionality in place that smaller medium enterprises have the ability to have a fair level playing field that would be already really beneficial and from my perspective it's all about identification, authentication and authorization. If you have that in place in the general term, then the market should be able to compete with each other*". In conclusion, as I4 said: "*What is key here is that you can identify and verify the other party on the platform, thus the identity management is important.*

Three interviewees also suggested that some type of federated design is likely the way various key concerns such as retaining data ownership may be alleviated. As I6 said: "*First of all, the word platform is tricky as it comes with the idea that there is a centralized storage of data, which should not be the case I think*". I7 also shared this view: "*I think that we should make a distinct because in in my definition a platform you have data platforms. It's fine with me, but it in my attention is data platform is always data at the source. So it's a Federated or decentralized initiate initiate the platform, so there's not one platform where all the data is being as a silo is being put over there even if it's in the cloud or private cloud or whatsoever*". In the context of Information Systems (IS), federation refers to the approach of distributing data storage and processing across multiple systems or nodes, rather than centralizing it in a single location. In this approach, each node in the federation maintains control over its own data while also contributing to a larger network of data resources that can be accessed by other nodes in the system. I2 has seen that federation has been a solution for competitors also : "*For Maersk, it was challenging how they could set-up the blockchain and governance structures because they wanted to have the competitors. However, the competitors (=carriers) of course do not want others to see their data. So, the solution became to silo the data streams of the carriers despite them using the same platform. This was possible due to the hyper ledger*". One of the main benefits of federation is that it can help to address concerns around data ownership and control, as each node can maintain its own security protocols and access controls. It can also improve scalability, as additional nodes can be added to the federation to handle increasing amounts of data. However, federation can also be challenging to implement and maintain, as it requires coordination across multiple systems and the development of standardized protocols for data sharing and communication. It may also result in increased complexity and cost due to the need for additional hardware and software resources.

When inquired on the ideal version of a platform, one interviewee suggested an architecture with tokenization of data assets. Tokenization involves the process of substituting sensitive data with a non-sensitive equivalent (token). This token can then be used to conduct various transactions without revealing the original information. This can be useful in ensuring the security of the sensitive data, as tokens can be stored and transmitted more securely than the original data. Incorporating tokenization as a core design consideration, I3 envisioned: "*And I think the data kind of thing could be also something like this; where data is flowing around. You know, we have agreed some standards because I think a general logistics platform is just too big, it would have to be about something. What is logistics data right? It could be about anything. In the vision of the physical* *internet we have adopted the long-term vision for logistics. It is not the digital internet as we know it but like an analogy to the physical world where goods flow freely. It's like I push something on the internet and it will find its way. Nobody knows exactly knows how its auctioned, every step of the way to take the best possible route; where there is also an alignment with other data packages and other physical packages that are running around. So, the way that goods move now; somebody fixes for me that the thing I ordered from China will follow exactly that route. In the physical internet, this could all be changing every step of the way so it is not really certain. There is some kind of tokenization of all this going on where the ownership of goods, shipping contracts can call be exchanged freely. Now, this requires also some kind of open data space as well. So, there would be NFT's floating around of different goods being shipped all over the world. And, I could take responsibility for one of those, if I'm in that business, whenever it comes my way. And I would specialize in just getting things quickly from A to B as possible and whenever it comes around, I will pick up and perform the action at the lowest cost possible. This type of thing would allow another round of specialization in the freight service market. I think this kind of open pool of bills of lading that can freely be transferred is something advanced that could eventually happen. So, this is something in the operational logistics world where the use of the data is clearly confined to logistics, where the business model is clear. Where the information is the bill of lading; the data around it could be changed to another bills of lading but it is about getting things from A to B*".

4.2.5.b Governance Components Governance components, on the other hand, refer to the policies, processes, and organizational structures that guide the decision-making, planning, and management of the system or platform. Based on the inquiry, a total of three governance related design considerations were discussed: trust (1), rules based membership (2), phased-onboarding (3).

First, trust remains a big denominator on whether participants are enticed to join the platform. Trust must be achieved both on a technical level and governance level. I4 also only sees a chance when the governance is done by a neutral party: "*Well, there are some pilots in the port of Houston for instance, but one important aspect is that it is done by a trusted party. And yeah, kind of neutral party that is like semi-public, similarly as the Port of Rotterdam despite it formally being a company but it has kind of semi-public role. And, if they are kind of the trusted party and that they more or less control the platform, decide who can provide the data, who can generate algorithms. This has a chance*".

Second, a rules-based membership to go about governance appears to be a success factor. As I2 explained: "*..they (=TradeLens) have a governance structure for how you could join. In this sense, it is open for other parties to join. The only thing is that you need to comply to the governance rules of how you could become a member. There is an-boarding process and as long as you comply with their rules; you can become a member and use their services. I have to check the right terminology but they distinguish between the roles of some players. Some of the players are more like data providers, such as the terminals who provide data like the carriers, whereas other players are more like users of the data. they want to grow so the on-boarding process has been one of the major things. Of course, when you are a global platform, you want to grow and even on-board your competitors. In principle, they have a governance structure with a governance board which allows competitors to join, e.g. other carriers, and the whole thing is that they want to make an industry-solution. And they also want competitors to join, so that is why they setup this industry board where they also invite them into the decision-making process. But also, of course, they want to on-board more supply chain partners. First of all, the main goal was to get content on the platform and they needed this data from Maersk. However, later, they also wanted other carriers to join so they get a lot of data (now I believe they have more than 50 percent of container shipment worldwide). The next strategic decision was when they started on-boarding port terminals because then you start getting data that has value for the other*".

Third, as I6 mentioned, if there is little interest to share data, it is always possible to add third-party service providers to build information-intensive services on top of the available data: "*Commercializing data itself is pretty difficult. I tried this myself in my previous capacity when I was an employee. It is pretty challenging and it is not really developing fast, as fast as I would have hoped. At the same time, there are quite a few applications that would fit this. Not by selling data directly but creating additional services, such as federative ways of booking, track & trace, CO2 accounting which is becoming big in the next years etc. These type of services use the data but do not commercialize the actual data let's say*". This appears to be the strategy of TradeLens also according to I2: "*we had discussions with the TradeLens team about a third-party who wanted to offer value-adding services on top of that. So, this third party went and got consent to use the data from the parties that hold the specific information. This consent was so that the company could bundle the information and offer an improved planning application, for example, when you can pick up your container from the terminal. So, I think what happens is that TradeLens provides like the basic layer. It is like an infrastructure and then other parties could enter, and of course they need to get the consent to use the* *data, to build value added applications on independent layers. However, from what I understood they simply approached the parties and asked for consent. Based on that, this app provider – who is independent of TradeLens – could then offer the value added services and charge for that*".

4.3 Requirement Elicitation

This chapter set out to identify what the domain-specific architecture requirements are for a multi-sided data marketplace architecture for logistics data. So far, we identified a total of sixteen distinct user concerns from our interviews that need to be translated into architecture requirements. The focus of this section is therefore dedicated to the development of architecture requirements that can address these concerns. For this reason, we begin this subsection by elaborating on our approach for formulating these requirements (4.3.1). The second part provides an overview of the architecture requirements developed in line with the formulation approach (4.3.2). The requirements originating from this subsection form a basis for the architecture design for a multi-sided data marketplace for logistics data.

4.3.1 Formulating Requirements

Documenting architecture requirements involves providing clear, concise, and comprehensive information about each requirement, its purpose, and any associated constraints or dependencies (Kruchten, 1995). If the architecture requirements are not well formulated, it can lead to various issues such as incomplete or incorrect implementation and the inability to meet the stakeholder concerns (Wulfert & Schütte, 2022). Architecture requirements should therefore be formulated as specific, well-defined statements that describe what a system must do and how it must behave in order to meet a (set of) specific need(s) or concern(s) of a stakeholder. The overall process we followed in this Chapter for the requirement elicitation process is depicted by Figure 11.

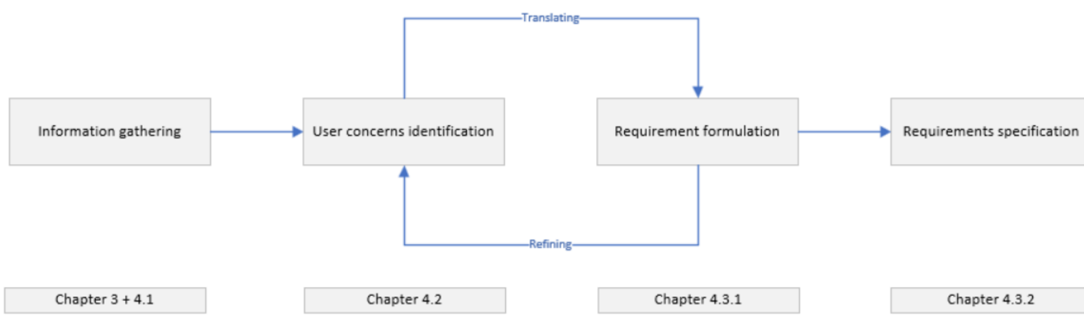

Figure 11: Requirement Elicitation Process

At this stage, we already completed the information gathering step about the users, stakeholders and the logistics operations that take place in Chapter 1. We reviewed state of the art literature to identify applicable requirements for the research question, but we found domain-specific requirements to be lacking in the context of multi-sided data marketplaces for logistics data. For this reason, we gathered additional information from domain experts using interviews. Through rigor and systematic analysis of the interview data, we were able to identify sixteen user concerns (Chapter 4.2). The identification of user concerns marks the completion of the second step in our requirement elicitation process. Regarding the third stage of the requirement elicitation process, it involves transforming as well as refining the identified user concerns into a preliminary list of requirements. As depicted by the figure, the third step is an iterative process between formulating requirements (translating) and ensuring they address the identified user concerns (refining). The translation of user concerns into requirements consists of two steps. First, each concern is translated into a specific objective that the architecture must achieve to alleviate the corresponding concern. The translation is derived by following the expression: *As a <role>, I want <goal> so that <reason>*. An overview of the concerns and corresponding objectives to alleviate them is found in Appendix D. Second, a requirement is formulated as a normative statement specifying how the architecture should be designed to meet this objective. After all concerns are translated, the corresponding requirements are subsequently refined based on how well they address the specified user concern.

To avoid misinterpretation of the requirements, the formulation of requirements should follow certain guidelines. Following these best practices helps to ensure that the architecture requirements are well-defined, comprehensive, and aligned with the concerns identified in our interviews. According to (TOGAF, 2023), five best practices for a well-formulated requirement are:

1. Clear and concise: The requirement should be stated in simple and straightforward language that is easy to understand and interpret.
2. Specific: The requirement should clearly state what the system must do or what it must not do.
3. Measurable: The requirement should be defined in such a way that it can be objectively measured and evaluated.
4. Traceable: The requirement should be traceable to a specific stakeholder need or expectation.
5. Consistent: The requirement should be consistent with other requirements and with the overall architecture of the system.
6. Relevant: The requirement should be relevant to the system and to the needs and expectations of its stakeholders.

4.3.2 Requirements Specification

Documenting the requirements is the final step in the requirement elicitation process. This phase is initiated after translating and refining the initial list of requirements. It is necessary that the documentation of the requirements specification employs structured format to avoid any misunderstanding in the interpretation. For this reason, we adopt the following elements for the specification of each requirement:

1. Requirement ID: a unique identifier to each requirement for easy reference and tracking.
2. Title: a brief, descriptive title or name for the requirement that conveys its purpose or intent.
3. Description: a concise description of the requirement, detailing what it entails, its purpose, and any associated functionality or behavior.
4. Rationale or Justification: relevance of the requirement to the stakeholder concerns.
5. Priority: priority level to the requirement (i.e., standard, emphasized) based on its importance to the project's success or the urgency of addressing the associated concern or need. Any concern with a count higher than 3 is classified as emphasized. This priority of the requirements is in line with the priority of the concerns as discussed in Chapter 4.2.2.

Based on the identified user concerns, we have formulated a total of sixteen corresponding requirements to address the specified concerns. Table 14 and 13 present the results obtained from the requirement formulation phase. We recognize that the formulated requirements address different objectives and may therefore have different characteristics which must be understood to get an overall understanding of whether the requirements are comprehensive. In the literature, there seems to be no universal way to classify requirements (Garlan, 2014). This study utilises a distinction between functional and non-functional requirements. Functional and non-functional requirements are two distinct categories of architecture requirements that serve different purposes in defining a system's behavior and characteristics. Functional requirements describe the features, capabilities, and services that the system must provide to fulfill its intended purpose. They define what the system is supposed to do, such as executing specific tasks, providing particular outputs, or supporting certain user interactions. Non-functional requirements, on the other hand, describe the characteristics or qualities of the system, rather than its specific features or capabilities. They define how well the system performs its functions and address aspects such as performance, reliability, scalability, security, and usability. What emerges from applying our classification scheme to the formulated requirements is a total of ten functional - and six non-functional requirements.

Creating this distinction has not been a straight forward process, however. In some cases, requirements appear to have both functional and non-functional aspects. For example, F-7, establishing transparent and consistent pricing mechanisms for data sets, can indeed be seen as having elements of both functional and non-functional requirements. As a functional requirement, it involves the implementation of specific pricing algorithms, tools, or features that allow data providers to set pricing and data buyers to compare prices across different data sets. For example, a function that allows data providers to easily set prices based on various factors, such as data volume or data complexity, and a function that displays the prices to data buyers in a clear and easily comparable manner. As a non-functional requirement, it focuses on the overall architecture principles and characteristics related to pricing, such as fairness, transparency, and consistency. These characteristics are crucial for ensuring that the data marketplace operates effectively and meets the needs of the logistics service providers (users). While these aspects are related and, in hindsight, may require additional the formulation of requirements to address the underlying aspects to the concern, the interview data we gathered at this stage only emphasized the functionality aspects rather than the principles. For this reason, the requirement has been formulated in such a way that system functionalities result from it. Similarly, F-2, guaranteeing that only authorized users can access the shared data through secure user authentication, role-based access control, and monitoring for suspicious activity, is also a mix of both functional and non-functional requirements. The main priority is on describing the specific functionalities that the architecture should posses, however, it indirectly also implies necessary aspects related to security, which are non-functional requirements.

A finding that stands out from the list of architecture requirements is the lack of non-functional requirements originating from the concerns of data buyers. This discrepancy may be attributed to differing priorities between data providers and data buyers. Data providers may prioritize the proper functioning of the system, while data users may prioritize their immediate data needs. The division in the main concerns expressed by interviewees reflects this divergence. Conducting further interviews may shed more light on this matter. It is important to note that users may not always immediately reiterate concerns they have previously expressed unless they are of significant importance. Although this issue could have been addressed more consistently throughout the interviews, additional clarification could be gained through further investigation.

Table 13: Non-Functional Requirements

	User Concern	**Requirement Title**	**Requirement**	**Priority**
NF-1	Shipper's ownership rights	Data Ownership and Attribution	The architecture should provide data ownership and attribution policies within the marketplace that specify the rights and responsibilities of data providers concerning the data they share. This includes establishing guidelines for data usage, licensing, and intellectual property protection.	Emphasized
NF-2	Large investment for providers	Low-Cost Entry and Scalability	The architecture should be designed to be cost-effective for data providers with limited resources while providing scalability as needed.	Emphasized
NF-3	Legal concerns	Legal and Regulatory Compliance	The architecture should include a compliance management module to assist users in adhering to relevant laws and regulations governing data collection, use, and sharing in the logistics industry.	Standard
NF-4	Lack of in-house technical capability	User-Friendly Interface and Documentation	The architecture should develop an intuitive user interface and comprehensive documentation, catering to users with varying levels of technical expertise.	Standard
NF-5	Neutral platform ownership	Neutral Platform Ownership	The architecture should ensure that the platform is governed by a neutral party or by a collaborative governance model that doesn't favor any specific stakeholder.	Standard
NF-6	Lock in effect	Avoid Vendor Lock-in	The architecture should support interoperability and ease of switching between platforms or third-party service providers, minimizing dependency and vendor lock-in.	Standard

Legend: Color schema: Blue = Data Provider, Green = Shared

Table 14: Functional Requirements

	User Concern	Title	Requirement	Priority
F-1	Loss of competitive advantage	Data Control and Selective Sharing	The architecture should incorporate encryption and access control mechanisms, enabling data providers to selectively share data and regulate the granularity of the disclosed information	Emphasized
F-2	Data confidentiality	Data Security and Access Control	The architecture should implement access control and authentication mechanisms, guaranteeing that only authorized users can access the shared data through secure user authentication, role-based access control, and monitoring for suspicious activity.	Emphasized
F-3	Usage control	Usage Agreement Enforcement	The architecture should monitor and enforce usage agreements, mitigating unauthorized or non-compliant use of the data.	Emphasized
F-4	Heterogeneous IT systems	Interoperability and Data Integration	The architecture should support integration with existing logistics systems, streamlining the exchange of data across diverse formats, protocols, and systems.	Emphasized
F-5	Low data quality	Data Validation and Quality Assurance	The architecture should implement mechanisms to validate data accuracy, completeness, and timeliness.	Emphasized
F-6	Low data coverage	Discoverability	The architecture should be designed to enable access to a wide array of data sources, ensuring extensive data coverage to satisfy a variety of needs	Emphasized
F-7	Pricing of data sets	Transparent and Fair Data Pricing Mechanism	The architecture should establish transparent and consistent pricing mechanisms for data sets.	Standard
F-8	Real-time data processing	Real-Time Data Processing and Analysis	The architecture should enable real-time processing and analysis of data to accommodate data buyers' requirements for prompt insights and decision-making.	Standard
F-9	Insufficient data supply	Data Standardization and Contextualization	The architecture should facilitate data standardization and contextualization processes, ensuring that shared data sets conform to unified formats, schemas, and contexts for simplified aggregation, comparison, and analysis.	Standard
F-10	Understanding data sets	Data Processing and Analytical Tools	The architecture should provide tools and resources that assist in processing, analyzing, and deriving insights from extensive volumes of unstructured data.	Standard

Legend: Color schema: Blue = Data Provider, Yellow = Data Buyer, Green = Shared

5 Formulation of Architecture Principles

In this chapter, we delve into the key aspects of our study on the design of multi-sided data marketplace architecture principles for logistics data. Having formulated the domain-specific architecture requirements in Chapter 4.3, our focus now shifts to synthesize the collected findings and contemplate its implications for our research. Additionally, we will explore the interconnections between our architecture requirements and the domain-agnostic requirements identified in Chapter 3. Building on these insights, we develop architecture principles derived from the architecture requirements, thereby highlighting the importance of the proposed architecture requirements in shaping multi-sided data marketplaces for logistics data. This chapter therefore serves as a platform for in-depth analysis and reflection on this study's results, paving the way to the discussion (Chapter 6) and conclusion on this study's main research question (Chapter 7).

This chapter is composed of three sections. First, we will extend the discussion by exploring the connections between our architecture requirements and existing research. This will not only strengthen our comprehension of the domain-specific distinctions in logistics but also help us identify the gaps that our study has addressed (5.1). Second, building on these insights, we propose architecture principles derived from our architecture requirements, showcasing their importance in creating a resilient and effective multi-sided data marketplace (5.2). These principles aim to serve as guidance to platform architects in design considerations during trade-offs, ensuring they meet stakeholder needs and priorities. Third, and last, we evaluate these architecture principles with domain experts to gather feedback, validate the adapted or extended architecture principles, and synthesize them into a comprehensive framework (5.3). The following sub-research question is thus answered by the end of this chapter:

SRQ3 - How do the proposed multi-sided data marketplace architecture requirements address the challenges and desired features identified in the problem and objectives stages?

5.1 Comparison with Domain-Agnostic Requirements

As mentioned in Chapter 3, little was found in the literature on the question of what domain-specific architecture requirements are suitable for the design of a logistics-focused multi-sided data marketplace architecture. Instead, we found a comprehensive list of seven domain-agnostic requirements which we used as a starting point to get an understanding of the functioning of a generic multi-sided data marketplace. However, while these requirements provide a good starting point (Table 15), they are not sufficient to address the full domain-specific needs, concerns, challenges and characteristics of the logistics domain. In search for answers, we conducted interviews with domain experts to identify domain-specific concerns and translated a list of the initial requirements (see Chapter 4.3.2). In this section, we present our comparative analysis between the domain-specific architecture requirements we developed in Chapter 4 and the domain-agnostic architecture requirements we found in the literature (Chapter 3). The purpose of this comparison is to validate our requirements, identify gaps, and ensure that the system design satisfies the broader system-level requirements.

Table 16 presents our comparison between the domain-agnostic and domain-specific requirements. Before we discuss the analysis, however, it is necessary to clarify exactly what is meant by the numbering in the table. The numbers in brackets that can be seen behind the domain-agnostic requirements on the horizontal axis refer to the sub-conditions defined in the description of the requirements (see Table 15). Whenever a cell contains a number, it means that the architecture requirement is related to the specific sub-condition of that domain-agnostic requirement. The color of the cell then informs on whether there is a match, similarity or no match. Closer inspection of the similarities and no matches in the table reveals the following:

1. Architecture requirements NF-2 Low-Cost Entry and Scalability, Architecture requirement, NF-4 User-Friendly Interface and Documentation and NF-5 Neutral Platform Ownership are new requirements compared to the existing domain-agnostic requirements. These requirements originate from the low resource availability and lack of technical capabilities concerns in logistics service providers.
2. Architecture requirement NF-3 Legal and Regulatory Compliance shows similarities with data-sharing trustworthiness but extends the requirement by proposing a compliance module to provide flexible guidance on adhering to the sector-specific regulations and laws. This originates from the uncertainty on the regulation and laws that are applicable.
3. Architecture requirement F-6 Discoverability shows similarities with the data sharing publication requirement but extends the requirement through a higher abstraction formulation. In contrast to the data sharing publication requirement which specifically implies functionalities for finding available

data, F-6 discoverability also implies the formulation of additional policies to incentivize participation of broad range logistics data providers.

4. In contrast to the domain-agnostic requirements, our study emphasizes the importance of data quality through a specific, separate requirement: the Data Validation and Quality Assurance requirement. Whereas the data sharing economy touches upon models to assess the value of data and monetise; it falls short of proposing functional mechanisms to ensure high value data in any transaction.
5. Data space engineering flexibility touches upon the ability to extend the architecture with specific capabilities based on context-specific data sharing conditions. While several architecture requirements we proposed may F-1, F-2, F-4 and F-8 look be similar depending on interpretation, they do not fully capture the adaptability essence of this requirement. The added value of this requirement indirectly relates to NF-2 Low-Cost Entry and Scalability. A highly flexible architecture may require complex and expensive implementations, while a cost-effective solution might sacrifice customization options and adaptability. To address this challenge, the architecture should be designed with modularity and extensibility in mind, allowing for incremental improvements and the integration of additional features as needed. This approach can help balance the need for customization and flexibility with the desire for cost-effectiveness and scalability,

According to this analysis, we can infer that all generic requirements have been met or extended to an extent, with the exception of the data space engineering flexibility requirement. Understanding how this architecture requirement should be incorporated would be a fruitful area for further work.

Table 15: Domain Agnostic Requirements (adapted from Nagel et al. (2021))

	Requirement	Description
1	Data Provider Empowerment	About ensuring that decisions can be made by appropriate stakeholders. This means that tools and practices are available for the possibility to: define and monitor policies in data sharing (1), govern the use of your data (2), and connect several data platforms with each other with each retaining control of own operations (3).
2	Data sharing trustworthiness	About ensuring that data sharing operations run according to expected (baseline) requirements. This means that the development of data sharing applications must support: security-by-design (1); privacy-by-design (2) and assurance-by-design (3).
3	Data sharing publication	About enabling data to be published so it can be easily located by data consumers.
4	Data sharing economy	About creating the conditions for data sharing and exchange, requiring: non-financial incentive mechanisms (1), financial incentive mechanisms, including models to determine the value of data as well as to monetise (2) and agreement mechanisms (3).
5	Data sharing interoperability	About providing the ability for all applications on data spaces to create, use, transfer and effectively exchange data. This requires the definition of data exchange API's and data models supporting: semantic interoperability (1), behavioral interoperability (2) and policy interoperability (3).
6	Data space engineering flexibility	About providing the ability to add customised features in data processing applications and platform to enable flexibility in terms of interoperability (i.e. specific interoperability capabilities) - (1), trustworthiness (i.e. specific security, privacy and assurance capabilities) - (2) and data processing (i.e. data processing capabilities) - (3).
7	Data space community	About fostering maximum reuse of data space solutions. This includes building on open solutions (1), ensuring features and solutions can be easily replicated (2), allowing free access to data and marketplace components developed by communities (3) and assurance that solutions will be available and maintained long in the future (4).

Table 16: Comparison domain agnostic requirements

	Requirement	Data Provider Empowerment (3)	Data sharing trustworthiness (2)	Data sharing publication (1)	Data sharing economy (3)	Data sharing interoperability (3)	Data space engineering flexibility (3)	Data sharing community (4)
F-1	Data Control and Selective Sharing	2					2	
F-2	Data Security and Access Control	2					2	
F-3	Usage Agreement Enforcementc	1						
F-4	Interoperability and Data Integration					2	1	
F-5	Data Validation and Quality Assurance				1			
F-6	Discoverability							
F-7	Transparent and Fair Data Pricing Mechanism				1-3			
F-8	Real-Time Data Processing and Analysis						3	
F-9	Data Standardization and Contextualization					1		
F-10	Data Processing and Analytical Tools						3	2,3
NF-1	Data Ownership and Attribution	3						
NF-2	Low-Cost Entry and Scalability							
NF-3	Legal and Regulatory Compliance		1-3			3		
NF-4	User-Friendly Interface and Documentation							
NF-5	Neutral Platform Ownership							
NF-6	Avoid Vendor Lock-in							1

Color scheme: Green = Match, Yellow = Similar, Red = No Match

5.2 Generalization of Findings

In this study, we delve into the design of architecture principles for a logistics multi-sided data marketplace reference architecture. Our focus has been to develop a comprehensive understanding of the complete architecture requirements that underpin the successful design and implementation of a multi-sided data marketplace specifically tailored to the logistics industry. To underpin our findings and contribute valuable insights to the growing body of knowledge in this domain, we set the focus of this sub-section to make the move towards the generalization of the research findings.

In this section, we will translate the identified architecture requirements into a set of coherent architecture principles (6.1). These principles will serve as guiding elements that inform and shape the development of a logistics multi-sided data marketplace reference architecture. By examining the relationships between these principles, we seek to provide a holistic understanding of how the individual components of the architecture interact and contribute to its overall effectiveness (5.2.2). This exploration will underscore the significance of our research in advancing the current understanding of logistics multi-sided data marketplace reference architectures and pave the way for future designs.

5.2.1 Architecture Principles

Architecture requirements and architecture principles are two related but distinct concepts in the field of information systems design and development. They both serve as guidelines to build a coherent and effective information system, but they focus on different aspects of the process. Architecture requirements specify what the system should do or what qualities it should possess in order to satisfy a specific need or expectation (Greefhorst & Proper, 2011b). In contrast, architecture principles are high-level, fundamental guidelines that drive the design and implementation of the architecture (Nagel et al., 2021). Architecture requirements are not static, they evolve with time. In order to ensure architects have a reference point for balancing competing concerns and requirements, architecture principles are necessary to help focus on the broader implications of design decisions. Therefore, instead of specifying a function or quality, architecture principles provide a consistent framework for (future) decision-making during the design, development and evolution of the system (TOGAF, 2023).

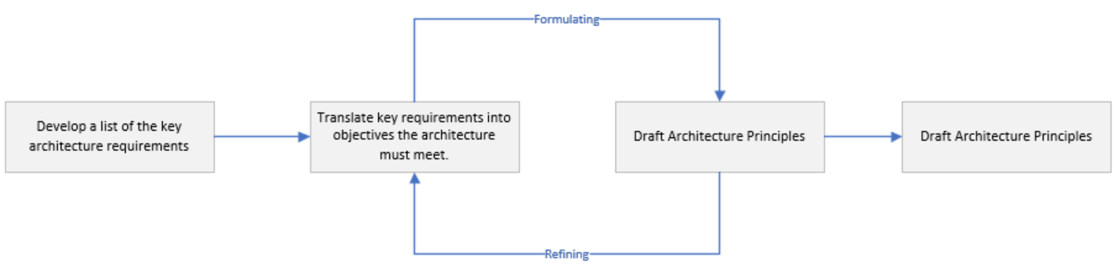

Figure 12: Principle Elicitation Process (adapted from (Greefhorst & Proper, 2011a)

Translating architecture requirements into architecture principles involves analyzing the specific needs and expectations of the system and deriving high-level guidelines that can inform the design and development process. This study follows a structured approach to generalize the situated outcomes (Figure 12). The first step involves reviewing the functional and non-functional architecture requirements and identifying the most critical ones that will have a significant impact on the system. In our case, we concluded that it is not necessary to exempt architecture requirements at this stage. We argue that any concern stated by the interviewees from which the architecture requirements originate is sufficiently important for us to look into what architecture principle it could translate to. As the architecture requirements have not yet been validated, reduction and optimization of the requirements and corresponding principles should occur in later stages. However, for now, we note the prioritization list we proposed based on our data and incorporate this in how we prioritize between the architecture principles we will propose. This is done based on the number of prioritized requirements are clustered under each principle. Moving on to the second step, this stage consists of determining the underlying objective each requirement is addressing in order to identify overarching themes. A full overview of our translation of the architecture requirements into objectives can be found in Appendix E. Based on this, we identified a total of seven distinct categories. In the third and final step, we derive high-level principles as statements that prescribe how to meet the shared objective of each category. The aim is to create principles that support the fulfillment of the requirements while providing a consistent framework for decision-making.

The generalization process resulted in a coherent set of seven architecture principles. Table 17 provides an overview of the principles. Because the lessons drawn in 6.1 also inspired their formulation, the architecture principles are based not only on the truth - the requirements - but also on what is important for designers to do to avoid design obstacles encountered during this research. The subsequent sections below explain all architecture principles. Each section describes a principle and explains the internal cohesion between the underlying architecture requirements. The code between brackets in each section refers to the architecture requirements that section 4.3.2 introduces (F = Functional Requirement, NF = Non-Functional Requirement). **Moreover, the color scheme in the tables indicate a prioritized requirement as elaborated in 4.3.2**.

Table 17: Architecture Principles

	Principle Title	**Description**
Architecture Principle 1	Secure and Controlled Data Sharing	Embed comprehensive security measures and access control mechanisms to protect data and enable selective sharing while ensuring authorized access.
Architecture Principle 2	Data Quality and Management	Foster data quality and seamless integration by supporting diverse data sources, and promoting data standardization, contextualization, and validation.
Architecture Principle 3	Data Processing And Analysis	Enable efficient data processing, real-time analysis, and versatile tools to empower users with timely insights and informed decision-making.
Architecture Principle 4	Transparency and Fairness	Maintain high transparency, fairness, and collaboration through predefined policy frameworks, consistent pricing mechanisms, and impartial governance structures.
Architecture Principle 5	Usability and Scalability	Optimize the architecture for user-centric design, cost-effectiveness, and scalability to cater to diverse needs, capability and resource availability.
Architecture Principle 6	Compliance And Legal Considerations	Adhere to relevant legal and regulatory requirements to maintain trust, protect user privacy, and ensure responsible data management.
Architecture Principle 7	Interoperability and Open Standards	Promote flexibility and adaptability by embracing open standards and interoperability, allowing seamless integration and easy transitions between solutions.

5.2.1.a Architecture Principle 1: Secure and Controlled Data Sharing The first and according to our data, the most prioritized architecture principle is "Secure and Controlled Data Sharing". LSP's often handle confidential data, making it crucial to establish a secure environment that prevents unauthorized access or misuse. This principle guides the design and development by emphasizing the importance of robust security and privacy measures when handling logistics data shared on the platform: "*Embed comprehensive security measures and access control mechanisms to protect data and enable selective sharing while ensuring authorized access.*

Adhering to this principle is critical for addressing security and privacy concerns of LSP's. It ensures that sensitive data is protected, and the risk of unauthorized access or non-compliant use is minimized. Data providers and users need to trust that the platform protects sensitive information and mitigates the risk of unauthorized access or non-compliant use. By enabling data providers to selectively share information and control its usage, a secure and controlled data-sharing environment fosters trust and encourages more logistics service providers to participate and share their data. Thus, the primary goal of this principle is to build a secure and reliable data-sharing platform that safeguards data providers' sensitive information, ensures only authorized access, and enforces usage agreements. This enhances confidence in the system and encourages more logistics service providers to participate and share their data. This principle provides guidance on the following aspects when making trade-offs between conflicting concerns during the design and development,

1. Prioritize the implementation of authentication mechanisms (e.g., multi-factor authentication) to verify user identities and prevent unauthorized access.

2. Emphasize the importance of role-based access control systems that allow data providers to define and manage access rights for different users or user groups.
3. Focus on the development of monitoring and alert systems to detect and respond to suspicious activities or potential security breaches.
4. Ensure that usage agreements are enforced and monitored, with mechanisms in place to identify and address non-compliant usage.
5. When facing trade-offs between conflicting goals, prioritize features and improvements that directly contribute to enhancing security, privacy, and control in the data-sharing environment.

By following this principle, the architecture will prioritize security, privacy, and control for all participants in the multi-sided data marketplace, fostering trust and collaboration among logistics service providers. Table 18 gives an overview of the requirements fundamental to architecture principle 1.

Table 18: Architecture Principle 1: Secure and Controlled Data Sharing

	Requirement Title	**Description**
F-1	Data Control and Selective Sharing	The architecture should incorporate access control mechanisms, enabling data providers to selectively share data and regulate the granularity of the disclosed information.
F-2	Data Security and Access Control	The architecture should implement access control and authentication mechanisms, guaranteeing that only authorized users can access the shared data through secure user authentication, role-based access control, and monitoring for suspicious activity.
F-3	Usage Agreement Enforcement	The architecture should monitor and enforce usage agreements, mitigating unauthorized or non-compliant use of the data.

5.2.1.b Architecture Principle 2: Data Quality and Management The second architecture principle relates to Data Quality and Management: "*Foster data quality and seamless integration by supporting diverse data sources, and promoting data standardization, contextualization, and validation.*

This principle is essential for addressing data quality and variety concerns in the multi-sided data marketplace for the logistics industry. In a multi-sided data marketplace for logistics data, the availability of high-quality and diverse data is crucial for driving informed decision-making, optimizing processes, and discovering new insights. As logistics operations often involve complex networks of interconnected entities, ensuring data quality and accessibility across multiple sources is essential for data buyers to leverage the platform effectively. This architecture principle addresses the need for accurate, complete, and timely data that can be easily standardized, contextualized, and analyzed, enhancing the value proposition of the data marketplace.

The primary goal of this principle is to create a robust data-sharing platform that prioritizes data quality and diversity while simplifying data integration and analysis processes. By focusing on these aspects, the architecture aims to provide a reliable source of logistics data that caters to various users' needs and enables them to derive valuable insights from the platform. By adhering to the Data Quality and Management principle, architects should focus on the prioritize the following aspects when deciding on design trade-offs between requirements:

1. Emphasize the importance of data validation mechanisms, such as automated data quality checks, data cleansing tools, and user feedback loops, to ensure that the data shared on the platform is accurate, complete, and up-to-date.
2. Encourage the integration of various data sources, including data from different logistics service providers, transportation modes, and geographic regions, to provide extensive data coverage that satisfies the diverse needs of users.
3. Prioritize data standardization and contextualization processes that facilitate data interoperability and consistency, allowing users to easily combine, compare, and analyze data sets from different sources. This may involve defining unified data formats, schemas, and context information, as well as implementing data transformation and mapping tools.

4. Encourage data providers to supply metadata or supplementary information that can help users better understand the context and quality of the data.
5. When facing trade-offs between conflicting goals, prioritize features and improvements that directly contribute to enhancing data quality, diversity, and usability, as these factors are critical to the success of the data marketplace.

By following this principle, the architecture will prioritize data quality and variety, ensuring that users have access to reliable, consistent, and diverse logistics data sets for informed decision-making and insights. Table 19 gives an overview of the requirements fundamental to architecture principle 2.

Table 19: Architect Principle 2: Data Quality and Management

	Requirement Title	**Description**
F-5	Data Validation and Quality Assurance	The architecture should implement mechanisms to validate data accuracy, completeness, and timeliness.
F-6	Discoverability	The architecture should be designed to enable access to a wide array of data sources, ensuring extensive data coverage to satisfy a variety of needs.
F-9	Data Standardization and Contextualization	The architecture should facilitate data standardization and contextualization processes, ensuring that shared data sets conform to unified formats, schemas, and contexts for simplified aggregation, comparison, and analysis.

5.2.1.c Architecture Principle 3: Data Processing and Analysis The third principle relates to the information creation aspect of the architecture: "*Enable efficient data processing, real-time analysis, and versatile tools to empower users with timely insights and informed decision-making*

The ability to process, analyze, and generate insights from data in real-time are regarded as essential for effective decision-making and responsiveness to dynamic operational needs of logistics. Furthermore, the integration of the marketplace with existing logistics systems and the ability to handle diverse data formats and protocols are critical for seamless data exchange and maximizing the value derived from shared data. The primary objective of this principle is to create a data marketplace that enables real-time processing and analysis of logistics data, supports integration with existing logistics systems, and offers tools and resources to assist data buyers in handling extensive volumes of unstructured data. This principle aims to ensure that users can efficiently generate actionable insights and make data-driven decisions that optimize their logistics operations. When making trade-offs during the design and development of the multi-sided data marketplace architecture for logistics data, this principle provides guidance on the following aspects:

1. Prioritize the development of real-time data processing and analysis capabilities, ensuring that data buyers can access prompt insights and make informed decisions in response to evolving operational needs.
2. Emphasize the importance of providing versatile tools and resources that assist data buyers in processing, analyzing, and deriving insights from extensive volumes of unstructured data, enabling them to harness the full potential of shared logistics data.
3. Focus on the seamless integration of the marketplace with existing logistics systems, ensuring smooth data exchange across diverse formats, protocols, and systems, and minimizing the need for manual data transformation or mapping.
4. When facing trade-offs between conflicting goals, prioritize features and improvements that directly contribute to enhancing agility, interoperability, and efficient data processing and analysis in the data-sharing environment.

By adhering to this architecture principle, the multi-sided data marketplace can effectively address the need for agile and interoperable data processing and analysis for logistics data. Table 20 gives an overview of the requirements fundamental to architecture principle 3.

Table 20: Architect Principle 3: Data Processing and Analysis

	Requirement Title	Description
F-8	Real-Time Data Processing and Analysis	The architecture should enable real-time processing and analysis of data to accommodate data buyers' requirements for prompt insights and decision-making.
F-10	Data Processing and Analytical Tools	The architecture should provide tools and resources that assist data buyers in processing, analyzing, and deriving insights from extensive volumes of unstructured data.
F-4	Interoperability and Data Integration	The architecture should support integration with the existing logistics systems, streamlining the exchange of data across diverse formats, protocols, and systems.

5.2.1.d Architecture Principle 4: Transparency and Fairness The fourth architecture principle relates to Transparency and Fairness: "*Maintain high transparency, fairness, and collaboration through predefined policy frameworks, consistent pricing mechanisms, and impartial governance structures.*

In a multi-sided data marketplace for logistics data, establishing transparent pricing mechanisms, well-defined data ownership and attribution policies, and flexible interoperability are critical for fostering trust and collaboration among stakeholders. These factors ensure that participants understand the value of the data they are sharing or purchasing, are aware of their rights and responsibilities concerning shared data, and have the flexibility to switch to alternative solutions if needed. This architecture principle is essential as it guides the design of the marketplace to prioritize transparency, equity, and flexibility, which are key factors influencing user satisfaction and the platform's overall success. By focusing on these aspects, the architecture can provide a solid foundation for stakeholders to engage in a fair and thriving ecosystem for data-driven decision-making in the logistics industry.

The primary objective of this principle is to create a data marketplace that offers transparent and consistent pricing mechanisms for data sets, provides clear data ownership and attribution policies, and supports open standards and interoperability to ensure a transparent, equitable, and flexible environment. This principle aims to enhance trust and collaboration among logistics service providers, enabling them to engage in mutually beneficial data-sharing activities and ultimately driving user satisfaction and the platform's success. When making trade-offs during the design and development of architecture requirements, this principle provides guidance on the following aspects:

1. Prioritize the establishment of transparent and consistent pricing mechanisms for data sets, ensuring that all participants have a clear understanding of the value of the data they are sharing or purchasing.
2. Emphasize the importance of providing well-defined data ownership and attribution policies, including guidelines for data usage, licensing, and intellectual property protection, to clarify the rights and responsibilities of data providers and users.
3. Focus on designing the architecture with open standards and interoperability in mind, enabling participants to easily switch to alternative solutions if needed and promoting a flexible, adaptable marketplace.
4. When facing trade-offs between conflicting goals, prioritize features and improvements that directly contribute to enhancing transparency, equity, and flexibility in the data-sharing environment.

By adhering to this architecture principle, the multi-sided data marketplace can effectively address the need for a transparent, equitable, and flexible data-sharing environment. Table 21 gives an overview of the requirements fundamental to architecture principle 4.

Table 21: Architect Principle 4: Transparency and Fairness

	Requirement Title	Description
F-7	Transparent and Fair Data Pricing Mechanism	The architecture should establish transparent and consistent pricing mechanisms for data sets.
F-1	Data Ownership and Attribution	The architecture should provide data ownership and attribution policies within the marketplace that specify the rights and responsibilities of data providers concerning the data they share. This includes establishing guidelines for data usage, licensing, and intellectual property protection.
F-5	Avoid Vendor Lock-in	The architecture should be designed with open standards and interoperability in mind, enabling participants to easily switch to alternative solutions if needed.

5.2.1.e Architecture Principle 5: Usability and Scalability The fifth architecture principle relates to Usability and Scalability: "*Optimize the architecture for user-centric design, cost-effectiveness, and scalability to cater to diverse needs, capability and resource availability.*

Throughout multiple interviews, we found lack of technical capability and resource availability to be a key concern for LSP's. For this reason, it is crucial to ensure that the architecture is cost-effective for data providers with limited resources, scalable to accommodate growth, and user-centric in its design. By considering the diverse needs and technical expertise of the users, the marketplace can attract a broader range of stakeholders such as SME's in last mile delivery sector. This architecture principle is essential as it guides the design of the marketplace to prioritize user-centricity, cost-effectiveness, and scalability, which are key factors influencing user satisfaction and the platform's overall success. By focusing on these aspects, the architecture can provide a solid foundation for stakeholders to engage in a thriving ecosystem for data-driven decision-making in the logistics industry.

The primary objective of this principle is to create a data marketplace that is cost-effective for data providers with limited resources, scalable to accommodate growth, and user-centric in its design. This principle aims to ensure that users with varying levels of technical expertise can effectively navigate and use the platform, while also minimizing the cost barriers for participation. By catering to these diverse needs, the architecture will drive user satisfaction and foster a successful data-sharing ecosystem in the logistics industry. When making trade-offs during the design and development of the multi-sided data marketplace architecture for logistics data, this principle provides guidance on the following aspects:

1. Prioritize cost-effectiveness for data providers with limited resources, ensuring that the marketplace remains accessible and inclusive for a diverse range of stakeholders.
2. Focus on scalability, designing the architecture to accommodate growth in data volume, user base, and feature set as needed.
3. Emphasize the importance of developing an intuitive user interface and comprehensive documentation that caters to users with varying levels of technical expertise, enhancing user satisfaction and minimizing the learning curve for platform adoption.
4. When facing trade-offs between conflicting goals, prioritize features and improvements that directly contribute to enhancing user-centricity, cost-effectiveness, and scalability in the data-sharing environment.

By adhering to this architecture principle, the multi-sided data marketplace can effectively address the need for a user-centric, cost-effective, and scalable data-sharing environment. Table 22 gives an overview of the requirements fundamental to architecture principle 5.

Table 22: Architect Principle 5: Usability and Scalability

	Requirement Title	Description
F-2	Low-Cost Entry and Scalability	The architecture should be designed to be cost-effective for data providers with limited resources while providing scalability as needed.
F-4	User-Friendly Interface and Documentation	The architecture should develop an intuitive user interface and comprehensive documentation, catering to users with varying levels of technical expertise.

5.2.1.f Architecture Principle 6: Compliance And Legal Considerations The sixth architecture principle relates to Compliance And Legal Considerations: *Adhere to relevant legal and regulatory requirements to maintain trust, protect user privacy, and ensure responsible data management*"

Complying with legal and regulatory requirements related to data privacy and protection is essential to maintain trust, safeguard user privacy, and ensure responsible data management. As various jurisdictions impose different regulations and laws governing data collection, use, and sharing, the architecture must be adaptable to accommodate these requirements, minimizing the risk of non-compliance and potential legal consequences. This architecture principle is crucial as it guides the design of the marketplace to prioritize compliance with legal and regulatory requirements, which is a key factor influencing user trust and the platform's overall success. By focusing on compliance, the architecture can provide a solid foundation for stakeholders to engage in a responsible and legally compliant ecosystem for data-driven decision-making in the logistics industry.

The primary objective of this principle is to create a data marketplace that complies with applicable legal and regulatory requirements related to data privacy and protection. This principle aims to ensure that the architecture is designed with compliance in mind, which will help maintain trust among users, protect user privacy, and promote responsible data management. By adhering to these requirements, the architecture will foster a data-sharing ecosystem that operates within the bounds of the law and maintains user trust. When making trade-offs during the design and development of the multi-sided data marketplace architecture for logistics data, this principle provides guidance on the following aspects:

1. Prioritize compliance with legal and regulatory requirements related to data privacy and protection, ensuring that the marketplace operates responsibly and maintains user trust.
2. Focus on designing the architecture to be adaptable and flexible, accommodating different legal and regulatory requirements as they evolve over time or vary across jurisdictions.
3. Emphasize the importance of incorporating privacy-by-design principles, minimizing the risk of non-compliance and potential legal consequences.
4. When facing trade-offs between conflicting goals, prioritize features and improvements that directly contribute to enhancing compliance with legal and regulatory requirements, maintaining trust, and protecting user privacy.

By adhering to this architecture principle, the multi-sided data marketplace can effectively address the need for compliance with legal and regulatory requirements related to data privacy and protection. Table 23 gives an overview of the requirement fundamental to architecture principle 6.

Table 23: Architect Principle 6: Compliance And Legal Considerationss

	Requirement Title	Description
F-3	Legal and Regulatory Compliance	The architecture should be designed to comply with applicable legal and regulatory requirements related to data privacy and protection.

5.2.1.g Architecture Principle 7: Interoperability and Open Standards The seventh and final architecture principle relates to interoperability and open standards: "*Promote flexibility and adaptability by embracing open standards and interoperability, allowing seamless integration and easy transitions between solutions.*"

Designing the architecture with open standards and interoperability in mind is crucial to enable participants to easily switch to alternative solutions if needed. By promoting flexibility and adaptability, the architecture encourages collaboration and fosters a thriving data-sharing ecosystem, allowing users to take advantage of various data sources, platforms, and tools without being locked into a specific solution. This architecture principle is essential as it guides the design of the marketplace to prioritize open standards and interoperability, which are key factors influencing the platform's overall success and the ability of stakeholders to collaborate effectively. By focusing on these aspects, the architecture can provide a solid foundation for stakeholders to engage in a flexible and adaptable data-sharing environment, enhancing the potential for data-driven decision-making in the logistics industry.

The primary objective of this principle is to create a data marketplace that embraces open standards and interoperability, enabling participants to easily switch to alternative solutions if needed. This principle aims to ensure that the architecture is designed with flexibility and adaptability in mind, promoting collaboration and fostering a thriving data-sharing ecosystem. By adhering to open standards, the architecture encourages the seamless exchange of data across various platforms, tools, and systems, enhancing the potential for data-driven decision-making in the logistics industry. When making trade-offs during the design and development of the multi-sided data marketplace architecture for logistics data, this principle provides guidance on the following aspects:

1. Prioritize open standards and interoperability, ensuring that the marketplace promotes flexibility, adaptability, and collaboration.
2. Focus on designing the architecture to accommodate various data sources, platforms, and tools, enabling users to easily switch between solutions and take advantage of diverse offerings in the ecosystem.
3. Emphasize the importance of promoting data portability and seamless integration, minimizing the risk of vendor lock-in and fostering a collaborative data-sharing environment.
4. When facing trade-offs between conflicting goals, prioritize features and improvements that directly contribute to enhancing open standards and interoperability, promoting flexibility, and fostering collaboration in the data-sharing ecosystem.

By adhering to this architecture principle, the multi-sided data marketplace can effectively address the need for open standards and interoperability. Table 24 gives an overview of the requirement fundamental to architecture principle 7.

Table 24: Architect Principle 7: Interoperability and Open Standards

	Requirement Title	Description
F-6	Avoid Vendor Lock-in	The architecture should be designed with open standards and interoperability in mind, enabling participants to easily switch to alternative solutions if needed.

5.2.2 Interdependencies between Architecture Principles

The architecture principles described in Chapter 5.2 are interdependent in multiple ways (Figure 13)[8]:

1. Data Quality and Management <-> Interoperability and Open Standards: There is a strong interdependency between fostering data quality and seamless integration, and promoting flexibility and adaptability through open standards and interoperability. Data standardization is essential for achieving seamless integration and interoperability between different systems and data sources. By promoting data standardization, the architecture ensures that diverse data sources can be easily combined, compared, and analyzed, making the data more useful and valuable to users. At the same time, embracing open standards and interoperability allows for more efficient data exchange between systems and simplifies the integration process.
2. Security and Access Control <-> Interoperability and Open Standards: Open standards allow for greater transparency in security practices. The specifications and protocols defined in open standards are openly available and accessible to the public. This transparency promotes scrutiny,

[8]the direction of the arrow describes the link of the relationship

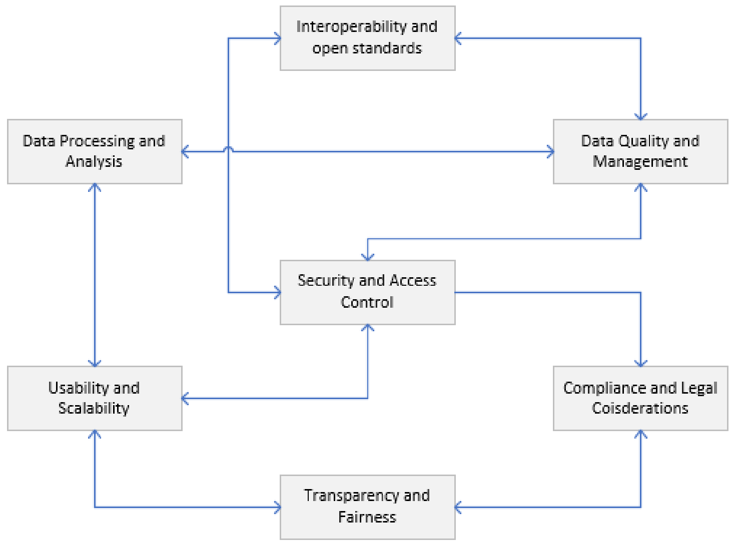

Figure 13: Interdependencies between Architecture Principles

peer review, and collaborative development, which can enhance the robustness and effectiveness of security measures.

3. Security and Access Control -> Compliance and Legal Considerations: Secure access control mechanisms helps to ensure compliance with data privacy regulations.
4. Security and Access Control <-> Data Quality and Management: Effective security measures and access controls help maintain data quality and integrity by preventing unauthorized access or tampering. For example, role-based access controls can ensure that only authorized users can modify data, preserving data quality (Sharma et al., 2020). High data quality and management, in turn, support security measures by providing a reliable, standardized, and well-structured data environment that is easier to secure and maintain.
5. Data Quality and Management <-> Data Processing And Analysis: High-quality data is essential for effective data processing and analysis. Standardization and contextualization of data sets enable more efficient and accurate analysis, leading to better insights and decision-making. In turn, efficient data processing and analysis tools can help identify and resolve data quality issues, ensuring continuous improvement of data quality.
6. Data Processing and Analysis -> Usability and Scalability: Efficient data processing and analysis capabilities contribute to an architecture's usability by providing users with timely insights and informed decision-making. As the architecture scales to accommodate more data sources and users, the importance of efficient data processing and analysis grows. Conversely, a scalable architecture should be able to handle increasing demands on data processing and analysis as the system grows.
7. Usability and Scalability <-> Transparency and Fairness: A user-centric design should facilitate collaboration between stakeholders by providing the necessary tools and features that encourage interaction and cooperation. This could include features such as communication channels, feedback mechanisms, and collaborative tools for data analysis. These features can contribute to a more positive user experience and promote the growth of the platform. Conversely, balancing cost-effectiveness with fairness is essential to ensure that the platform remains accessible to users with varying resources.
8. Transparency and Fairness <-> Compliance and Legal Considerations: Ensuring transparency and fairness in pricing and ownership policies supports better governance and helps comply with legal and regulatory requirements. Conversely, a strong governance model and adherence to regulations will foster transparency and fairness in the marketplace.
9. Security and Access Control <-> Usability and Scalability: Implementing robust security measures can be resource-intensive, which may impact the cost-effectiveness of the architecture. Likewise, as the platform scales, both in terms of the number of users and the volume of data being handled,

the security measures must also evolve to keep up with the increased risks and potential threats. Scalability considerations must be addressed alongside security measures to ensure that the platform remains secure as it grows.

These interdependencies show that the architectural principles are not isolated aspects but rather work together to create a cohesive and robust data marketplace architecture. By understanding and considering these interrelationships, researchers can therefore create a more effective and robust design.

5.3 Evaluation of the Architecture Principles

This section presents the methodology findings of the validation process conducted to demonstrate and validate the proposed architecture principles for the logistics domain. The architecture principles were formulated as a result of careful analysis and consideration of the unique architecture requirements and challenges. To ensure the robustness and effectiveness of our seven proposed architecture principles, a validation phase was conducted, consisting of three interviews with domain experts in logistics. The purpose of these interviews was to gather expert opinions and insights regarding the relevancy, effectiveness, and completeness of the proposed architecture principles. The validation process aimed to assess the practical applicability (1) and suitability (2) of the principles in addressing the specific needs and concerns of logistics data sharing. Table 25 outlines an overview of the interviewees from the second round of interviews, and presents the key findings and outcomes of the validation interviews. The column Relation to LSP's refers to the perspective of the logistics service providers, discussed in Chapter 1.2, the interviewees cover. This allows us to understand from which perspective they engage the interview questions. The insights gained from the expert interviews contribute to the overall credibility and applicability of the proposed architecture principles and enhance their relevance in the context of logistics data marketplace design.

Table 25: Interview Participants

	Experience	**Relation to LSP's**
I-8	An experienced logistics operations researcher with expertise in data-driven optimization for transportation. Specialized in studying data sharing in the Port of Rotterdam and currently focused on Sustainable Logistics at a large research institute.	1PL, 2PL
I-9	Manager of a large-scale, national Digital Infrastructure Logistics (DIL) project, focused on accelerating the digitization of the Dutch logistics sector.	3PL, 4PL
I-10	Professor with over 22 years of experience in logistics, with roles as VP Smart Logistics and Principal Business Consultant at a a large international IT consulting firm. He has extensive expertise in planning, optimization, collaboration, and applying IT technologies in transportation and logistics. Additionally, he is an Assistant Professor at a renowned university in the Netherlands, teaching and supervising students in industrial engineering and business information systems.	3PL, 4PL, 5PL

During the validation interviews, participants shared their perspectives on the proposed architecture principles, highlighting both their relevance and potential areas of contention. While all principles were deemed relevant, it was noted that not all participants agreed on their importance to the same extent. For instance, Interviewee 9 questioned the validity of open principles, suggesting that businesses would prioritize value over adherence to open standards. This viewpoint underscores the differing emphasis placed on principles by different stakeholders. Furthermore, two out of three interviewees raised the issue of the fine line between principles and requirements. They pointed out that certain principles, such as security, are often considered as baseline requirements rather than distinct architectural principles. However, an alternative perspective suggests that security can be viewed as a guideline with varying levels of implementation, ensuring higher levels of security as needed for specific cases. In addition, conflicts between principles were identified during the interviews. One example was the potential conflict between high security and resource availability. Achieving a high level of security measures could potentially impact the availability of resources, highlighting the need for carefully managing and balancing these aspects within the architecture. These insights from the interviewees contribute to a more comprehensive understanding of the proposed architecture principles. It underscores the importance of considering differing perspectives and potential conflicts when applying these principles in the design and implementation of a logistics data marketplace.

6 Discussion

In this discussion chapter, our focus now shifts to synthesize the collected findings and contemplate its implications for our research. Building on these insights, we are able to highlight the viability of the proposed architecture principles in shaping multi-sided data marketplaces for logistics data. This chapter therefore serves as a platform for in-depth analysis and reflection on this study's results, paving the way to the conclusion of this study (Chapter 7).

This chapter is composed of two sections. First, we begin by reflecting on the implications of our research findings from Chapter 5. Reflecting on the implications entails discussing the broader effects the findings might have on the development, implementation, and use of multi-sided data marketplace architectures in the logistics sector (6.1). Afterwards, we will explore the relationship between the architecture principles and the value proposition (6.2).

6.1 Information Marketplaces, not Data Marketplaces

Throughout this study, we referred to data marketplaces as: "a digital platform enabling and simplifying the transactions of machine-readable data by matching market participants, facilitating data sharing services and providing a pricing mechanism". However, the results of this study imply that logistics data providers as well as the respective data owners do not want to transact in machine-readable data at all. Instead, the results suggest arrangements of logistics data sharing should be made in terms of the specific information derived from the underlying data, rather than the data itself. In this section, we build on the results discussed in the general outlook and platform design of the results section. We relate to what the results imply for the business model of the MSDM as well as reflect on their implications for the architecture design (6.2).

On the topic of how large-scale data sharing could develop in the logistics industry, our study found that LSP's would not provide data without knowing what information is extracted from it. This might somewhat be surprising given the fact that other studies never made this distinction between data and information when referring to data sharing Bastiaansen et al. (2020). A possible explanation for this might be related to inherent characteristics of data. As we discussed in our literature review, data itself has no intrinsic value. It is only after the transformation and contextualization process that data buyers are able to extract valuable information, knowledge and ultimately value from the data. The Data-to-Value framework is a conceptual model that illustrates this process of extracting valuable insights and actionable information from raw data (Figure 14). According to the model, data-based value creation is divided into four areas: data collection (1), information creation (2), value creation (3), and distribution through the provider network (4) [9]. What this model makes clear is that data itself is a building stone. Depending on how it is contextualized, it can be used for any purpose.

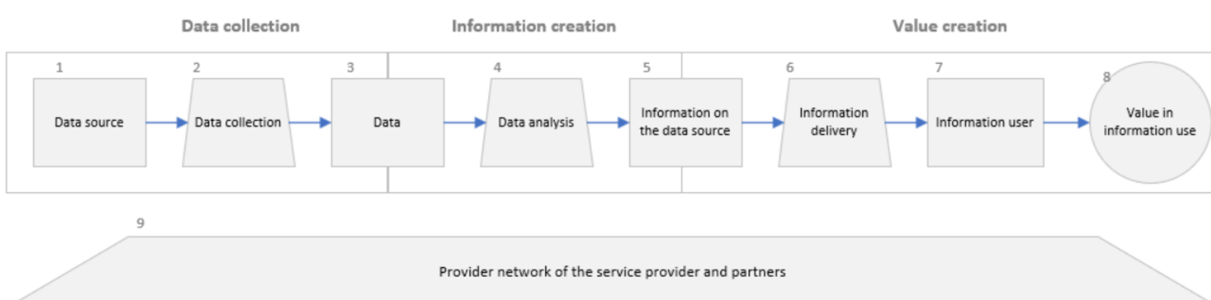

Figure 14: Data-2-Value Framework (adapted from Lim et al. (2018)

In the absence of a well-defined framework that articulates the benefits of sharing and utilizing data, data providers may struggle to grasp the value proposition of participating in these marketplaces. This explains the reluctance to share data, limited data availability, and reduced marketplace effectiveness we found in our results and which earlier studies have pointed out (Koutroumpis et al., 2020b; Sharma et al., 2020). Moreover, without a comprehensive understanding of the value that can be derived from data trading, organizations within the logistics industry may be hesitant to invest in the necessary infrastructure, technologies, and processes required for effective data sharing. The absence of a clear value framework

[9]The provider network in this context refers to the (third party) service provider and its partners, i.e.

implies that the logistics industry may miss out on opportunities for improved efficiency, cost reduction, and enhanced decision-making offered by data-driven insights. By not fully understanding the potential value of shared data, organizations may fail to capitalize on the competitive advantages that can be gained from leveraging data analytics and advanced technologies, such as artificial intelligence and machine learning (Reinsel et al., 2020).

Trust is also crucial for fostering collaboration among participants in a multi-sided data marketplace. However, as there seems to be little trust, logistics service providers (LSPs) will remain reluctant to participate in data-sharing platforms, resulting in limited availability and diversity of data offerings and slower adoption. We also learned from our interviews that the lack of trust also leads to increased transaction costs, as LSPs might implement additional measures, such as legal agreements or technical safeguards, to protect their data and mitigate potential risks. These extra measures can burden data buyers, making data acquisition more cumbersome and potentially discouraging participation in the marketplace. Since LSPs cannot fully anticipate what information can be extracted from their datasets, they face unmitigated risks by allowing other businesses unrestricted access. While we have proposed several architecture requirements, such as F-1 and F-2, to address concerns about losing control over shared data, they will not be sufficient to build adequate trust Koutroumpis et al. (2020b). Furthermore, as LSPs currently fulfill their data needs through their own supply chain in closed data-sharing platforms or bi-lateral exchanges, it remains uncertain how value will be created from offering data specifically through the marketplace. From our interviews, we found that third-party service providers may play a significant role in the future by offering value-added applications on top of data. They can fulfill any part of the Data-to-Value framework, from step 3 'Data' to step 7 'Information delivery'. It is possible, therefore, that third-party service can address this gap of unclear value proposition and trust between both sides of the marketplace. Overall, these findings suggest that developing a successful multi-sided data marketplace architecture for business-to-business data sharing in the logistics industry requires creating a clear and compelling value proposition, providing tools and support for LSPs through third-party service providers, and incorporating mechanisms and policies that promote trust and security.

6.2 Architecture Implications on Value Proposition

The relationship between architecture and the value proposition can be understood through the role that the architecture plays in enabling and supporting the platform's value proposition (Garlan, 2014). Given the business model implications we discussed in the previous section, moving beyond traditional data marketplaces and focusing on delivering actionable insights from data might stimulate higher adoption. In line with I3 from our interviews: "*I think this is the kind of data brokerage or, actually, information brokerage situation. Here, you are not really buying the data anymore but, instead, you are buying the information. So, somebody analyzes the data for you, gives you what you need to know and doesn't let you know anything that the other one does not want you to know. This is the data-to-value framework where I know what kind of value I am getting from the information. I don't know what data it is but on the other side is the data that is processed by someone to create information that has value for me. I think it is more than just selling data. In the context where data is sold, it is put on the platform and somebody will be getting the data after acquisition. If there is processing in between, data analytics such as descriptive, prescriptive, predictive analytics et cetera, then a lot of detail might be dropped while the algorithm is producing the information. So, I am using a neural network and nobody knows what data I have but you get the pattern from me, which you can use for data-driven products e.g.. I think this kind of thing is a big step from providing data and at this point it is not anymore just a data market or you could say that the data market is not there for final clients but only for those who process the data into something else*". The architecture can operate within the data-to-value framework, allowing users to access valuable information without needing to know the underlying data. By processing data through analytics, such as descriptive, prescriptive, and predictive techniques, the platform will extract patterns and insights, which can be used to create data-driven products. In essence, it would move beyond the traditional idea of a data marketplace, and in essence operate as an information factory with trusted third-party service providers to deliver value.

This approach offers several benefits:

1. Enhanced data ownership: By only sharing the resulting information rather than the raw data, businesses can maintain control over their data and protect sensitive information.

2. Improved value proposition: Users will receive actionable insights tailored to their needs, ensuring they gain value from participating in the platform.

3. Support for third party service providers (i.e. third-party service providers): The platform can serve as an intermediary between data providers and businesses that process data into valuable information. Data providers can be rewarded for contributing data to the platform, while trusted 'third-party vendors' can access and utilize (parts of) this data to generate insights for their clients.
4. Streamlined information acquisition: Users will no longer need to sift through raw data to extract valuable information. Instead, they can directly access the insights they need, saving time and resources.
5. Increased familiarization: the trust aspect is shifted from the data buyer to the data processor. In contrast to a data buyer, the relationship with the data processor can be built over multiple initiatives spanning different data buyers. This may help reduce the trust concerns outside the technical functionalities to enhance trust.

7 Conclusion

In this concluding chapter, we revisit the primary objective of this research, which was to develop architecture principles for the design of a multi-sided data marketplace in the logistics domain. Through a systematic exploration of the relevant literature, stakeholder interviews, and a rigorous analysis of our findings, we have contributed to a better understanding of the underlying factors and requirements that influence the design and development of such an architecture. This chapter synthesizes the main findings of our study by first answering the three sub-research questions (7.1), after which it answers the main research question (7.2). Afterwards, the scientific contributions of the study are discussed (7.3). Equally important, the chapter continues by highlighting the societal and managerial implications that emerge from our research (7.4. Moreover, we elaborate on the suitability of this research to the MSc. Management of Technology program at Delft University of Technology (7.5). We also acknowledge the limitations of our study (7.6 and provide directions for future research, emphasizing areas where further investigation could potentially enhance our understanding of the complexities and challenges inherent in the design and development of multi-sided data marketplaces for the logistics industry (7.7). The outcomes of this research answers the following main research question:

"What multi-sided data marketplace architecture principles can facilitate business-to-business data sharing in the logistics industry?"

7.1 Sub-Research Questions

The main research question was subdivided into three sub-research questions to guide the research process. In this section, we address each sub-research question individually leading up to answer the main research question in the following section. The first part answers sub-research question 1 (7.1.1), the second part answers sub-research question 2 (7.1.2), and the third part answers sub-research question 3 (7.1.3).

7.1.1 Sub-Research Question 1

The primary objective of the first sub-research question (SRQ1) was two-fold. First, it aimed to draw upon existing research by finding specific details about the building blocks and principal components necessary for developing a generic decentralized multi-sided data marketplace. Secondly, it sought to uncover the underlying domain-specific challenges within the logistics industry that are relevant for determining the functional and non-functional requirements of the architecture. By completing these objectives, we aimed to gain a deeper understanding of what a logistics multi-sided data marketplace should achieve and how it could add value for logistics service providers. SRQ1 was related to the second activity of the Design Science Research (DSR) framework, which involved establishing the objectives and desired outcomes of the artifact by understanding prior knowledge Peffers et al. (2007).

"What prior knowledge is available on the desired features and characteristics of a multi-sided data marketplace architecture for the logistics industry?"

To address SRQ1 comprehensively, we conducted a literature review. This process involved collecting and analyzing information related to decentralized data marketplace architectures, their applicability to the logistics domain, and any specific challenges or opportunities in facilitating business-to-business (B2B) data sharing. The literature review helped ground the research in existing knowledge and identify relevant gaps in the current understanding of MSDMs in the logistics industry. A primary advantage of conducting a literature review was that it enabled a comprehensive understanding of the current state of research on the topic (Greefhorst & Proper, 2011b). By performing a literature review, we were be able to identify knowledge gaps in the existing literature and establish the context for the remainder of the research. In conclusion, the analysis of the prior knowledge on the desired features and characteristics of a multi-sided data marketplace architecture for the logistics industry has revealed several key insights that were crucial for understanding the current state of research and areas that we further explored with SRQ2 and SRQ3:

1. First, we have identified that different types of data marketplaces exist, each with its own specific focus and purpose (Stahl et al., 2016). However, multi-sided data marketplaces stand out as suitable for large-scale data sharing within the logistics industry Koutroumpis et al. (2020b). These platforms enable multiple stakeholders to engage in data exchange, fostering collaboration and innovation across the sector.
2. Second, we found that there is limited empirical research available on domain-specific multi-sided data marketplace architecture requirements (Abbas et al., 2021; Koutroumpis et al., 2020a; **?**; Otto

et al., 2019; Otto, 2020), indicating a gap in the existing literature that should be addressed to develop the architecture principles in this study. This scarcity of research underscored the need for more in-depth investigation and analysis of the architecture principles and features that contribute to effective multi-sided data marketplace architectures.

3. Nevertheless, we discovered domain-agnostic requirements from international data spaces that can be applied as a starting point for designing the domain-specific architecture requirements (Nagel & Lycklama, 2021; Otto et al., 2019). These general requirements offer a valuable baseline for creating a robust and flexible architecture that can accommodate various data types and use cases.
4. However, to achieve a comprehensive understanding of the architecture and its suitability for the logistics industry, domain-specific requirements must be considered (Otto et al., 2019). These requirements will help ensure that the architecture principles reflect the unique challenges and opportunities within the logistics domain, such as specific data formats, privacy concerns, and regulatory compliance.

7.1.2 Sub-Research Question 2

As detailed by our answer on SRQ1, little was found in the literature on the question of what domain-specific architecture requirements are suitable for the design of a logistics-focused multi-sided data marketplace architecture. Instead, we found a comprehensive list of seven domain-agnostic requirements which we used as a starting point to get an understanding of the functioning of a generic multi-sided data marketplace. However, while these requirements provided a good starting point (Table 15), they were not sufficient to address the full domain-specific needs, concerns, challenges and characteristics of the logistics domain. In search for answers, we conducted interviews with domain experts to identify domain-specific concerns and translated a list of the initial architecture requirements (see Chapter 4.3.2). In this section, we present our a comparative analysis of the domain-specific architecture requirements we developed and the domain-agnostic architecture requirements found in the literature. The purpose of this comparison is to validate our requirements, identify gaps, and ensure that the system design satisfies the broader system-level requirements.

"What are the key architecture requirements necessary to design a multi-sided data marketplace for the logistics industry?"

The key requirements necessary to design a multi-sided data marketplace for the logistics industry are separated into functional and non-functional requirements. The following functional requirements are found to be as follows:

1. The architecture should incorporate encryption and access control mechanisms, enabling data providers to selectively share data and regulate the granularity of the disclosed information
2. The architecture should implement access control and authentication mechanisms, guaranteeing that only authorized users can access the shared data through secure user authentication, role-based access control, and monitoring for suspicious activity
3. The architecture should monitor and enforce usage agreements, mitigating unauthorized or non-compliant use of the data
4. The architecture should support integration with existing logistics systems, streamlining the exchange of data across diverse formats, protocols, and systems
5. The architecture should implement mechanisms to validate data accuracy, completeness, and timeliness
6. The architecture should be designed to enable access to a wide array of data sources, ensuring extensive data coverage to satisfy a variety of needs
7. The architecture should establish transparent and consistent pricing mechanisms for data sets
8. The architecture should enable real-time processing and analysis of data to accommodate data buyers' requirements for prompt insights and decision-making
9. The architecture should facilitate data standardization and contextualization processes, ensuring that shared data sets conform to unified formats, schemas, and contexts for simplified aggregation, comparison, and analysis
10. The architecture should provide tools and resources that assist in processing, analyzing, and deriving insights from extensive volumes of unstructured data

The following non-functional requirements are found to be as follows:

1. The architecture should provide data ownership and attribution policies within the marketplace that specify the rights and responsibilities of data providers concerning the data they share. This includes establishing guidelines for data usage, licensing, and intellectual property protection
2. The architecture should be designed to be cost-effective for data providers with limited resources while providing scalability as needed
3. The architecture should include a compliance management module to assist users in adhering to relevant laws and regulations governing data collection, use, and sharing in the logistics industry
4. The architecture should develop an intuitive user interface and comprehensive documentation, catering to users with varying levels of technical expertise
5. The architecture should ensure that the platform is governed by a neutral party or by a collaborative governance model that doesn't favor any specific stakeholder
6. The architecture should support interoperability and ease of switching between platforms or third-party service providers, minimizing dependency and vendor lock-in

7.1.3 Sub-Research Question 3

Building on the architecture requirements developed as an outcome of SRQ2, we aimed to translate the formulated architecture requirements into architecture principles. The aim is to address the sub-research question of how the architecture requirements of the proposed marketplace address the challenges and desired features identified during the problem and objectives stages. As we discussed in Chapter 1, we aim to understand how large-scale data sharing can be facilitated with a logistics-focused data marketplace.

This research aims to find the architecture principles that can guide the design of a logistics-focused multi-sided data marketplace. Architecture principles capture knowledge gained along with the process of building a solution and encompass knowledge about creating other instances that belong to this class (Arrow et al., 1996). By doing so, architecture principles connect the generalized outcomes to a class of solutions, and a class of problems. As a result, architecture principles function as recommendations on how a designer should design artefacts that solve comparable problems. Here, it is not about the generalization from a sample to the population, but rather about transportability: the usability of the architecture principles in problem contexts outside the study context (Degtiar & Rose, 2023; Lesko, 2017). Within this research, the class of problems refers to the challenge of designing a logistics-focused multi-sided data marketplace reference architecture that functions as a guiding tool for developers of large-scale data sharing platforms and researchers who aim to build similar architectures. The reference architecture forms a solution instance that represents a class of solutions. In this research, the class of open, multi-sided data marketplace reference architectures: architectures for large-scale data sharing that logistics service providers with any resource and technical availability can use to facilitate their data sharing needs or attract additional revenue streams.

"How does the proposed multi-sided data marketplace architecture requirements address the challenges and desired features identified in the problem and objectives stages?"

The formulated architecture requirements have paved the way to formulate overarching architecture principles which can be used as guidelines for the design and implementation of logistics-focused multi-sided data marketplaces aimed at facilitating B2B data sharing. The following seven principles therefore answer SQ3, after validating their relevancy, effectiveness and completeness in three interviews with logistics experts:

1. Embed comprehensive security measures and access control mechanisms to protect data and enable selective sharing while ensuring authorized access
2. Foster data quality and seamless integration by supporting diverse data sources, and promoting data standardization, contextualization, and validation
3. Enable efficient data processing, real-time analysis, and versatile tools to empower users with timely insights and informed decision-making
4. Maintain high transparency, fairness, and collaboration through predefined policy frameworks, consistent pricing mechanisms, and impartial governance structures
5. Optimize the architecture for user-centric design, cost-effectiveness, and scalability to cater to diverse needs, capability and resource availability

6. Adhere to relevant legal and regulatory requirements to maintain trust, protect user privacy, and ensure responsible data management
7. Promote flexibility and adaptability by embracing open standards and interoperability, allowing seamless integration and easy transitions between solution

7.2 Main Research Question

The answer to the main research question consists of seven architecture principles that guide the design of a domain-specific multi-sided data marketplace reference architecture. These seven architecture principles build on to the domain-agnostic architecture requirements the literature suggests for developing multi-sided data marketplaces (Nagel & Lycklama, 2021): tools and practises to define and monitor policies in data sharing as well as to govern the use (1), development of data sharing applications in accordance with security-, privacy- and assurance- by-design (2), tools to publish data (3), incentive and agreement mechanisms (4), interoperability between systems (5), ability to add customised features (6) and building on open solutions (7). Domain-agnostic architecture requirements, however, can be applied to any domain or industry (Otto et al., 2019). Therefore, while these requirements provide a good starting point, they are not sufficient to address the full domain-specific needs, concerns, challenges and characteristics of the logistics domain.

To address this limitation, it is essential to complement the domain-agnostic requirements found in the literature with domain-specific requirements that account for the unique needs, challenges and characteristics of the logistics domain. This research followed the Design Science Research framework to develop the domain-specific architecture requirements (Chapter 4) and subsequently the architecture principles (Chapter 5). The Design Science Framework is a research methodology that focuses on the creation and evaluation of artifacts to solve real-world problems (Peffers et al., 2007; Dresch et al., 2015). In the context of information systems, it involves the design and development of artifacts and their evaluation through rigorous empirical methods. The final set of architecture requirements outline an architecture that researchers and practitioners can use to build a multi-sided data marketplace architecture for the logistics data. The formalization of the domain-specific architecture requirements resulted in the seven architecture principles that can guide future developments of multi-sided data marketplaces in the logistics domain.

Defining the boundaries of domain-specific data types and deciding which data types should be included or excluded in a multi-sided data marketplace can be a complex task (Hagiu & Wright, 2015). It is important to consider that the concept of a "logistics industry" may be an oversimplification. Rather, the logistics sector can be more accurately described as a network of individual organizations, each fulfilling unique roles within the broader ecosystem. When examining the data sharing requirements of a single organization, it becomes evident that their interests extend beyond the logistics domain. These organizations must interact with various stakeholders, such as governmental bodies and insurance companies, sharing data with each of these entities. Consequently, the perspective of the logistics service providers place themselves at the center of data sharing activities, with a focus on their own needs and requirements. In light of these observations, organizations within the logistics sector may not be solely interested in a logistics-centric data marketplace. Instead, they may seek a data sharing platform that caters to their diverse needs and requirements, facilitating data exchange with multiple stakeholders across various domains.

The architecture principles for a multi-sided data marketplace in the logistics domain highlight several key aspects that are crucial for effective and secure data sharing in this specific context. These principles emphasize the need to:

1. Prioritize data security and access control: Ensuring that sensitive data is protected from unauthorized access and providing granular control over data sharing is vital for maintaining trust and encouraging participation among stakeholders.
2. Focus on data quality and integration: Supporting diverse data sources and promoting data standardization and contextualization helps ensure that the shared data is accurate, complete, and consistent, ultimately facilitating seamless data integration and analysis.
3. Enable real-time data processing and analysis: Providing tools and resources to process and analyze data efficiently and in real-time empowers users to make informed decisions and gain valuable insights from the shared logistics data.
4. Foster transparency, fairness, and collaboration: Establishing transparent policy frameworks, consistent pricing mechanisms, and impartial governance structures encourages trust, collaboration, and equal opportunities for all parties involved in the data sharing process.

5. Design user-centric, cost-effective, and scalable solutions: Creating intuitive, easy-to-use systems that can adapt to the diverse needs of stakeholders and varying levels of resources is essential for fostering widespread adoption and usage of the data marketplace.
6. Comply with legal and regulatory requirements: Adhering to relevant data privacy, protection laws, and industry-specific regulations ensures responsible data management practices and maintains trust among stakeholders.
7. Promote flexibility and adaptability through open standards and interoperability: Embracing open standards and interoperable technologies enables seamless integration with other systems and solutions, allowing users to easily switch between platforms and reducing vendor lock-in.

In conclusion, the architecture principles for a multi-sided data marketplace in the logistics domain emphasize the importance of creating a secure, reliable, user-friendly, and adaptable platform that facilitates seamless and responsible data sharing among stakeholders. By addressing key concerns such as data security, quality, integration, processing, transparency, compliance, and adaptability, these principles guide the design of a data marketplace that effectively meets the unique needs and challenges of the logistics domain.

7.3 Scientific Contributions

The seven architecture principles form a coherent set of guidelines for future designers. This section compares the architecture principles for designing a logistics-focused multi-sided data marketplace architecture with the guidelines that already exist in architecture design literature. By doing so, this section makes explicit the similarities and divergences between designing an architecture of a logistics-focused multi-sided data marketplace and a generic multi-sided data marketplace architecture. As a result, this section explicates the novelty of the architecture principles and proves the contribution of the architecture principles to the architecture design and logistics architecture design knowledge base. We compare our architecture principles with the four generic principles as proposed by the most comprehensive work on multi-sided data marketplaces Nagel & Lycklama (2021): Data sovereignty (1), Data level playing field (2), Decentralised soft infrastructure (3) and Public-private governance (4).

The seven architecture principles for a logistics multi-sided data marketplace and the four principles from the scientific literature share common objectives and address similar concerns within the context of data sharing and exchange. These principles converge in their emphasis on aspects such as data security, quality, transparency, collaboration, and adherence to legal and regulatory requirements. One commonality between the two sets of principles is the focus on data security and access control. Both Principle 1 (Prioritize data security and access control) and the principle of Data Sovereignty from the scientific literature highlight the importance of protecting sensitive data and ensuring authorized access. They both emphasize the need to establish trust among stakeholders and maintain the integrity and confidentiality of shared data. Data quality and integration are also key concerns shared by both sets of principles. Principle 2 (Focus on data quality and integration) and the principle of Level Playing Field for Data Sharing and Exchange from the scientific literature recognize the significance of accurate, complete, and consistent data. They advocate for the standardization and contextualization of data to enable effective integration and analysis across diverse sources. Real-time data processing and analysis is another shared principle. Principle 3 (Enable real-time data processing and analysis) and the principle of Decentralized Soft Infrastructure emphasize the value of efficient data processing and analysis. They both aim to empower users with tools and resources to make timely and informed decisions based on the shared (logistics) data. Transparency, fairness, and collaboration are principles that are mutually addressed. Principle 4 (Foster transparency, fairness, and collaboration) and the principle of Public-Private Governance underscore the importance of establishing transparent policies, impartial governance structures, and pricing mechanisms. These principles promote trust, collaboration, and equal opportunities for all stakeholders involved in the data sharing process.

While there are similarities, there are also notable differences between the two sets of principles. Principle 5 (Design user-centric, cost-effective, and scalable solutions) emphasizes creating user-centric and adaptable systems, while the principle of Level Playing Field for Data Sharing and Exchange and the principle of Decentralized Soft Infrastructure focus on fairness, equal opportunities, and decentralized infrastructure management, respectively. These principles address different aspects of user-centric design, fairness, and infrastructure management. Principle 6 (Comply with legal and regulatory requirements) emphasizes adhering to legal and regulatory obligations, while the principles of Data Sovereignty and Public-Private Governance delve into aspects of data control, ownership, and the involvement of public and private entities in governing data sharing. These principles tackle different aspects of legal compliance and data control. Principle 7 (Promote flexibility and adaptability through open standards and interoperability) highlights the significance of open standards and interoperable technologies. This is distinct from the principle of Decentralized Soft Infrastructure, which emphasizes a decentralized approach to infrastructure management. These principles provide different perspectives on achieving flexibility and adaptability within the data marketplace architecture.

7.4 Managerial and Societal Contributions

This study makes managerial contributions in several ways by providing guidance for designing a logistics multi-sided data marketplace. First, emphasizing data quality and integration enables managers to establish robust data governance frameworks. This involves defining data standards, implementing data validation processes, and ensuring consistent data across multiple sources. By prioritizing data quality, managers can enhance decision-making processes, improve operational efficiency, and foster trust among stakeholders who rely on accurate and integrated logistics data (Reinsel et al., 2020). Second, this study contributes by promoting transparency and collaboration within the marketplace. Managers can establish policies and governance structures that facilitate transparent data sharing, pricing mechanisms, and unbiased decision-making. By fostering an environment of transparency, managers promote trust among participants, encourage collaboration, and ensure fair and equitable opportunities for all stakeholders involved in the data exchange process. This contributes to the development of a collaborative ecosystem that drives innovation and value creation in the logistics industry.

The contributions of this study may also have societal benefits. First, the emphasis on collaboration and knowledge sharing within multi-sided data marketplace fosters an environment of collective learning and innovation. By facilitating the exchange of data, insights, and best practices, logistics service providers may better collaborate to address industry-wide challenges and develop innovative data-driven solutions (Bastiaansen et al., 2020). This collaboration contributes to the advancement of the logistics sector by fostering a culture of continuous improvement and shared learning, ultimately benefiting society as a whole. Second, the marketplace's role in informing policy and regulation may also have societal implications. Policymakers and industry regulators can access the shared logistics data to gain a comprehensive understanding of industry trends, challenges, and opportunities. This knowledge may empower them to develop evidence-based policies and regulations that promote efficiency, sustainability, and societal well-being. Informed policy decisions can foster a favorable business environment, encourage responsible practices, and drive positive social and economic outcomes in the logistics industry (Ceniga & Sukalova, 2015). Last, the accessibility and democratization of data within the marketplace contribute to empowering smaller logistics service providers. These entities, which may have limited resources and access to data, can leverage the shared logistics data to make informed decisions and compete with larger industry players. By providing equal access to valuable insights and opportunities, the marketplace promotes inclusivity, entrepreneurship, and economic growth. This empowerment of smaller businesses and stakeholders could drive innovation, diversity, and job creation within the logistics industry, benefiting society by fostering a dynamic and inclusive business ecosystem.

7.5 Suitability of this research to the MoT program

This research has been carried out in the context of a graduation thesis for the MSc. Management of Technology (MOT) program at the Technology, Policy and Management faculty of Delft University of Technology. MOT graduates learn to explore and understand how firms can use technology to design and develop products and services that contribute to improving outcomes such as customer satisfaction, corporate productivity, profitability and competitiveness. A typical MoT thesis fits one or more of following three criteria:

1. Research reports on a scientific study in a technological context;
2. Show understanding of technology as a corporate resource or from a corporate perspective;
3. Research includes scientific methods and tools as put forward in the curriculum.

We argue this master thesis satisfies the qualities of a typical MoT thesis as it addresses the critical intersection of technology, management, and innovation within the logistics industry. By focusing on the development of an effective data-sharing ecosystem, the research contributes to enhancing the overall performance and competitiveness of logistics service providers. The MOT program emphasizes the integration of technology, management, and innovation to create value in organizations, and this research on the design of a multi-sided data marketplace architecture principles for logistics data aligns with these objectives. By identifying the concerns of logistics service providers and formulating architecture requirements and principles, we provide valuable insights into the design considerations needed for a successful data marketplace. This enables organizations to make better-informed decisions on the implementation and management of such platforms, fostering innovation and collaboration in the logistics industry. Moreover, this study addresses aspects of technology management, such as data privacy, security, and compliance. Finally, this study contributes to the development of analytical and problem-solving skills that are central to the MOT program. By offering a comprehensive analysis of the concerns, requirements, and principles involved in the design of a multi-sided data marketplace architecture requirements and principles, we provide a foundation for further research and practical application in the field of technology management.

7.6 Limitations of the study

In this section, the limitations of this study are discussed, acknowledging the constraints and potential weaknesses inherent in the design of the study, methods, and analysis. By recognizing these limitations a balanced and transparent account of the findings can be provided, as well as to guide future research in addressing these challenges. By examining the boundaries of this study, we aim to enhance the validity and reliability of the conclusions.

One limitation of this study is the limited scope of the Systematic Literature Review, primarily due to the scarcity of data available on the research question. The emerging and rapidly evolving nature of the research domain, coupled with the novelty of the problem under investigation, may have contributed to this limited data availability. As a result, the understanding of the problem and the state of knowledge in the domain might not be entirely exhaustive, potentially affecting the ability to identify all relevant challenges, desired features, and existing solutions (Adams, 2015). Additionally, the extended duration of the research project further exacerbated this limitation, as it led to the exclusion of more recent publications and developments in the field.

Two other limitations of this study are related to the semi-structured interview methodology. First, the small sample size may have caused a bias and/or a lack of variability in data collection (Degtiar & Rose, 2023; Lesko, 2017). This means that the theoretical propositions provided in this study might not comprehensively address the research question.

Second, only one form of purposive sampling was employed, namely expert sampling. This study should have included stakeholder sampling in the interview process. By focusing solely on experts, valuable insights might have inadvertently been overlooked from other stakeholders who play a crucial role in the research domain, such as end-users, policymakers, or non-expert practitioners (Palinkas et al., 2015). This limitation may have led to a narrower understanding of the problem and its potential solutions, as findings are primarily based on expert opinions.

The final limitation of this study pertains to the validation of the architecture requirements and the derived architecture principles. While an effort was made to validate the requirements by comparing them with domain-agnostic requirements, this approach may not have been sufficiently rigorous to ensure the comprehensiveness and relevance of the final architecture requirements. As a result, there may be domain-specific nuances or considerations that were not adequately captured, leading to potential gaps or inaccuracies in the proposed architecture principles.

7.7 Directions for future research

In this section, we discuss potential directions for future research to expand upon the findings of this study, address its limitations, and contribute to a more comprehensive understanding of the multi-sided data marketplace architecture in the logistics industry. By exploring new avenues of investigation, refining methodologies, and incorporating emerging developments in the field, future research can build upon this study to advance knowledge and provide insights for both academics and practitioners. We outline several research directions that could shed light on aspects of the problem, validate and enhance the proposed architecture requirements, and ultimately contribute to the development of more effective and efficient data sharing solutions in the logistics domain.

Two directions for future research are to focus on enhancing the proposed architecture requirements and validating the architecture requirements. This can be achieved by refining and expanding the requirements based on new insights, emerging trends, and feedback from a diverse range of stakeholders. By conducting pilot tests, case studies, or real-world implementations of the proposed architecture principles, researchers can identify potential gaps, limitations, and opportunities for improvement. Iteratively refining the architecture requirements based on empirical evidence and practical experiences can contribute to the development of a more robust, flexible, and scalable solution that is better suited to the specific needs and challenges of the logistics industry. This, in turn, can enable more efficient and secure data sharing, fostering collaboration and innovation across the sector.

Another direction for future research is to investigate the business models and monetary value of multi-sided data marketplaces for the logistics industry. As many of the interviewees mentioned, the monetary aspect of data sharing is closely related to its perceived value in the logistics industry, particularly given the fact that margins in this sector are often small. A comprehensive understanding of the economic aspects of these marketplaces can play a role in driving their adoption, fostering sustainable growth, and maximizing the value for all stakeholders involved, while also considering the industry's tight margins and competitive landscape.

By identifying the key drivers and barriers to adoption, and incorporating the perspectives of industry professionals who emphasize the monetary value of data sharing within the context of small margins, researchers can inform the development of supportive policies, incentive structures, and industry standards that promote a thriving data-sharing ecosystem in the logistics sector. Ultimately, this line of inquiry can contribute to unlocking the full potential of multi-sided data marketplaces, enabling more informed decision-making, enhanced collaboration, and improved operational efficiency across the logistics industry.

References

Abbas, A., Agahari, W., van de Ven, M., Zuiderwijk, A., & de Reuver, M. (2021, 06). Business data sharing through data marketplaces: A systematic literature review.. doi: 10.18690/978-961-286-385-9.6

Adams, W. C. (2015, 10). Conducting Semi-Structured Interviews. *Handbook of Practical Program Evaluation: Fourth Edition*, 492–505. doi: 10.1002/9781119171386.CH19

ADVANEO. (2022). *Features of the ADVANEO Data Marketplace.* Retrieved from https://www.advaneo-datamarketplace.de/en/features-of-the-datamarketplace/#

Agahari, W., Ofe, H., & de Reuver, M. (2022, 7). It is not (only) about privacy: How multi-party computation redefines control, trust, and risk in data sharing. *Electronic Markets 2022 32:3*, *32*(3), 1577–1602. Retrieved from https://link.springer.com/article/10.1007/s12525-022-00572-w doi: 10.1007/S12525-022-00572-W

Arrow, K., Bolin, B., Costanza, R., Dasgupta, P., Folke, C., & Holling, C. S. (1996). Economic growth, carrying capacity, and the environment. *Environment and Development Economics*, *1*(1), 104–110. Retrieved from https://www.researchgate.net/publication/231747926_Economic_Growth_Carrying_Capacity_and_the_Environment doi: 10.1017/S1355770X00000413

Assarroudi, A., Heshmati Nabavi, F., Reza Armat, M., Ebadi, A., & Vaismoradi, M. (2018). Directed qualitative content analysis: the description and elaboration of its underpinning methods and data analysis process. *Journal of Research in Nursing*, *23*(1), 42–55. doi: 10.1177/1744987117741667

Baldwin, C. Y., Business, H., Woodard, S. C. J., & Woodard, C. J. (2009, 01). The Architecture of Platforms: A Unified View. *Platforms, Markets and Innovation*, 19. Retrieved from https://ink.library.smu.edu.sg/sis_research/2452 doi: 10.4337/9781849803311.00008

Banerjee, P., & Ruj, S. (2019). Blockchain Enabled Data Marketplace-Design and Challenges.

Bastiaansen, H., Nieuwenhuis, K., Zomer, G., Piest, J. P. S., van Sinderen, M., & Dalmolen, S. (2020, 08). The logistics data sharing infrastructure.

Bianco, V. D., Myllarniemi, V., Komssi, M., & Raatikainen, M. (2014). The role of platform boundary resources in software ecosystems: A case study. *Proceedings - Working IEEE/IFIP Conference on Software Architecture 2014, WICSA 2014*, 11–20. doi: 10.1109/WICSA.2014.41

CargoSmart. (n.d.). *CargoSmart | Global Shipment Management Solutions.* Retrieved from https://www.cargosmart.com/en-us/

Ceniga, P., & Sukalova, V. (2015, 01). Future of Logistics Management in the Process of Globalization. *Procedia Economics and Finance*, *26*, 160–166. doi: 10.1016/S2212-5671(15)00908-9

Chapman, A., Simperl, E., Koesten, L., Konstantinidis, G., Ibáñez, L. D., Kacprzak, E., & Groth, P. (2019, 08). Dataset search: a survey. *The VLDB Journal 2019 29:1*, *29*(1), 251–272. Retrieved from https://link-springer-com.tudelft.idm.oclc.org/article/10.1007/s00778-019-00564-x doi: 10.1007/S00778-019-00564-X

Commission, E. (2020). *Business-to-Business data sharing: An economic and legal analysis* (Tech. Rep.). Retrieved from https://ec.europa.eu/jrc/sites/default/files/jrc121336.pdf

Degtiar, I., & Rose, S. (2023, 3). A Review of Generalizability and Transportability. *https://doi.org/10.1146/annurev-statistics-042522-103837*, *10*, 501–524. Retrieved from https://www.annualreviews.org/doi/abs/10.1146/annurev-statistics-042522-103837 doi: 10.1146/ANNUREV-STATISTICS-042522-103837

De Reuver, M., Sørensen, C., & Basole, R. C. (2017). The digital platform: a research agenda. Retrieved from https://doi.org/10.1057/s41265- doi: 10.1057/s41265

Dresch, A., Lacerda, D. P., & Antunes, J. A. V. (2015). Design Science Research. , 67–102. doi: 10.1007/978-3-319-07374-3{_}4

Duparc, E., Otto, B., & Culotta, C. (n.d.). Evaluating Platform Openness in Logistics based on a Taxonomic Analysis. Retrieved from https://www.researchgate.net/publication/357016466 doi: 10.24251/HICSS.2022.598

Eisenmann, T., Parker, G., & Van Alstyne, M. W. (2006). Strategies for Two-Sided Markets. Retrieved from www.hbr.org

European Commission. (2017). *BUILDING A EUROPEAN DATA ECONOMY* (Tech. Rep.). Retrieved from https://digital-strategy.ec.europa.eu/en/library/communication-building-european-data-economy

European Commission. (2020). *Data governance act | Shaping Europe's digital future.* Retrieved from https://digital-strategy.ec.europa.eu/en/policies/data-governance-act

Evofenedex. (2021). Onderzoek data en digitalisering in de logistiek 2021. , 9.

Fruhwirth, M., Rachinger, M., & Prlja, E. (2020). Discovering business models of data marketplaces. *Proceedings of the Annual Hawaii International Conference on System Sciences*, *2020-Janua*, 5738–5747. Retrieved from https://click.endnote.com/viewer?doi=10.24251%2Fhicss.2020.704&token=WzMxNTU3NDksIjEwLjI0MjUxL2hpY3NzLjIwMjAuNzA0Il0.wpx6UVMh9wJUegpjzugwJop207I doi: 10.24251/HICSS.2020.704

Garlan, D. (2014, 05). Software architecture: A travelogue. *Future of Software Engineering, FOSE 2014 - Proceedings*, 29–39. Retrieved from https://dl.acm.org/doi/10.1145/2593882.2593886 doi: 10.1145/2593882.2593886

Gelhaar, J., Gürpinar, T., Henke, M., & Otto, B. (2021). Towards a taxonomy of incentive mechanisms for data sharing in data ecosystems. In *25th pacific asia conference on information systems.* Retrieved from https://www.researchgate.net/publication/352857138

Gelhaar, J., & Otto, B. (2020, 06). Challenges in the Emergence of Data Ecosystems. *PACIS 2020 Proceedings*. Retrieved from https://aisel.aisnet.org/pacis2020/175

Greefhorst, D., & Proper, H. A. (2011a). A practical approach to the formulation and use of architecture principles. *Proceedings - IEEE International Enterprise Distributed Object Computing Workshop, EDOC*, 330–339. Retrieved from https://click.endnote.com/viewer?doi=10.1109%2Fedocw.2011.18&token=WzMxNTU3NDksIjEwLjExMDkvZWRvY3cuMjAxMS4xOCJd.2PIfonF3ak6Say1tqGcJ88V0E0o doi: 10.1109/EDOCW.2011.18

Greefhorst, D., & Proper, H. A. (2011b). A practical approach to the formulation and use of architecture principles. *Proceedings - IEEE International Enterprise Distributed Object Computing Workshop, EDOC*, 330–339. doi: 10.1109/EDOCW.2011.18

Gupta, P., Dedeoglu, V., Kanhere, S. S., & Jurdak, R. (2021, 01). Towards a blockchain powered IoT data marketplace. *2021 International Conference on COMmunication Systems and NETworkS, COMSNETS 2021*, 366–368. doi: 10.1109/COMSNETS51098.2021.9352865

Hagiu, A., & Wright, J. (2015). Multi-sided platforms . Retrieved from http://dx.doi.org/10.1016/j.ijindorg.2015.03.003 doi: 10.1016/j.ijindorg.2015.03.003

Helo, P., & Hao, Y. (2022). Artificial intelligence in operations management and supply chain management: an exploratory case study. *Production Planning and Control*, *33*(16), 1573–1590. doi: 10.1080/09537287.2021.1882690

Hevner, A., & Chatterjee, S. (2010). Design Science Research in Information Systems. , 9–22. doi: 10.1007/978-1-4419-5653-8{_}2

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science In Information Systems. *Design Science in IS Research MIS Quarterly*, *28*(1), 75.

Iivari, J. (2015, 01). Distinguishing and contrasting two strategies for design science research. *European Journal of Information Systems*, *24*(1), 107–115. Retrieved from https://www.researchgate.net/publication/271195278_Distinguishing_and_contrasting_two_strategies_for_design_science_research doi: 10.1057/EJIS.2013.35

International Data Space Association. (2019). *Use Case Overview International Data Spaces* (Tech. Rep.).

Jesson, J., & Lacey, F. M. (2011). *Doing Your Literature Review: Traditional and Systematic Techniques* (Vol. 1). SAGE Publications Ltd. Retrieved from https://books.google.nl/books/about/Doing_Your_Literature_Review.html?id=NAYrLb8qsd4C

Karam, A., Reinau, K. H., & Østergaard, C. R. (2021, 12). Horizontal collaboration in the freight transport sector: barrier and decision-making frameworks. *European Transport Research Review*, *13*(1), 1–22. Retrieved from https://etrr.springeropen.com/articles/10.1186/s12544-021-00512-3 doi: 10.1186/S12544-021-00512-3/METRICS

Koutroumpis, P., Leiponen, A., & Thomas, L. D. (2020a). Markets for data. *Industrial and Corporate Change*, *29*(3), 645–660. doi: 10.1093/ICC/DTAA002

Koutroumpis, P., Leiponen, A., & Thomas, L. D. W. (2017). ETLA Working Papers The (Unfulfilled) Potential of Data Marketplaces. Retrieved from http://pub.etla.fi/ETLA-Working-Papers-53.pdf

Koutroumpis, P., Leiponen, A., & Thomas, L. D. W. (2020b, 06). Markets for data. *Industrial and Corporate Change*, *29*(3), 645–660. doi: 10.1093/icc/dtaa002

Kruchten, P. (1995). Architectural Blueprints-The "4+1" View Model of Software Architecture. *IEEE Software*, *12*(6), 42–50.

Lesko, B. A. L. W. D. E. J. K. H. M. G. . C., C. R. (2017). Generalizing Study Results: A Potential Outcomes Perspective. *Epidemiology (Cambridge, Mass.)*, 553-–561. doi: 10.1007/978-0-387-35610-5{_}9

Lindgren, B. M., Lundman, B., & Graneheim, U. H. (2020, 08). Abstraction and interpretation during the qualitative content analysis process. *International Journal of Nursing Studies*, *108*, 103632. doi: 10.1016/J.IJNURSTU.2020.103632

Lis, D., & Otto, B. (2021). Towards a taxonomy of ecosystem data governance. *Proceedings of the Annual Hawaii International Conference on System Sciences*, *2020-January*, 6067–6076. doi: 10.24251/HICSS.2021.733

Malekani, A., Bahame, A., & Aba, A. (n.d.). Assessing the efficacy of electronic document management Assessing the efficacy of electronic document management system in records management at Sokoine University of system in records management at Sokoine University of Agriculture Agriculture. Retrieved from https://digitalcommons.unl.edu/libphilprac

Mayring, P. A. (2023). Qualitative content analysis. *International Encyclopedia of Education(Fourth Edition)*, 314–322. doi: 10.1016/B978-0-12-818630-5.11031-0

McKinsey. (2017). *Fueling growth through data monetization* (Tech. Rep.). Retrieved from https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/fueling-growth-through-data-monetization

Midway, S. R., Sievert, N. A., Lynch, A. J., Whittier, J. B., & Pope, K. L. (2022, 9). Asking nicely: Best practices for requesting data. *Ecological Informatics*, *70*, 101729. doi: 10.1016/J.ECOINF.2022.101729

Mohanta, B. K., Jena, D., Panda, S. S., & Sobhanayak, S. (2019, 12). Blockchain technology: A survey on applications and security privacy Challenges. *Internet of Things*, *8*, 100107. doi: 10.1016/j.iot.2019.100107

Nagel, L., & Lycklama, D. (2021, 07). Design Principles for Data Spaces - Position Paper. Retrieved from https://zenodo.org/record/5105744 doi: 10.5281/ZENODO.5105744

Nagel, L., Lycklama, D., Ahle, U., Bastiaansen, H., Bengtsson, K., Caballero, M., . . . Valiño, J. (2021, 04). Design principles for data spaces – position paper..

Nakagawa, E., Oliveira Antonino, P., & Becker, M. (2011, 01). Reference architecture and product line architecture: A subtle but critical difference..

Navarrete, A. G. (n.d.). User Requirements Elicitation: A Comparison between Generative Techniques and Semi-Structured Interviews.

Noble, H., & Smith, J. (2015, 04). Issues of validity and reliability in qualitative research. *Evidence-Based Nursing*, *18*(2), 34–35. Retrieved from https://ebn.bmj.com/content/18/2/34https://ebn.bmj.com/content/18/2/34.abstract doi: 10.1136/EB-2015-102054

Ofoeda, J., Boateng, R., & Effah, J. (2019, 07). Application programming interface (API) research: A review of the past to inform the future. *International Journal of Enterprise Information Systems*, *15*(3), 76–95. Retrieved from https://www.researchgate.net/publication/334145268_Application_Programming_Interface_API_Research_A_Review_of_the_Past_to_Inform_the_Future doi: 10.4018/IJEIS.2019070105

Otto, B. (2020). *Data Governance in Data Ecosystems-Insights from Organizations* (Tech. Rep.). Retrieved from https://aisel.aisnet.org/amcis2020/strategic_uses_it/strategic_uses_it/12

Otto, B., Hompel, M. t., & Wrobel, S. (2019). International Data Spaces: Reference architecture for the digitization of industries. *Digital Transformation*, 109–128. Retrieved from https://www.researchgate.net/publication/333084851_International_Data_Spaces_Reference_architecture_for_the_digitization_of_industries doi: 10.1007/978-3-662-58134-6{_}8

Otto, B., & Jarke, M. (2019, 12). Designing a multi-sided data platform: findings from the International Data Spaces case. *Electronic Markets*, *29*(4), 561–580. doi: 10.1007/s12525-019-00362-x

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015, 09). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and policy in mental health*, *42*(5), 533. Retrieved from /pmc/articles/PMC4012002//pmc/articles/PMC4012002/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4012002/ doi: 10.1007/S10488-013-0528-Y

Park, J.-I., Jin, S.-g., & Jung, H.-K. (2012). XML-based EDI system for port logistics. *Proceedings of the Korean Institute of Information and Commucation Sciences Conference*, 518–520. Retrieved from http://www.unece.org

Parra-Moyano, J., Buhler, M., Avital, M., & Schmedders, K. (2021). Steps Toward Fostering Peer-to-Peer Blockchain-Based Data Markets. Pacific Asia Conference on Information Systems (PACIS). Retrieved from https://www.researchgate.net/publication/352372483_Steps_Toward_Fostering_Peer-to-Peer_Blockchain-Based_Data_Markets

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007, 12). A design science research methodology for information systems research. *Journal of Management Information Systems*, *24*(3), 45–77. doi: 10.2753/MIS0742-1222240302

PwC. (2019). *The future of the logistics industry.* Retrieved from https://www.pwc.nl/nl/assets/documents/pwc-shifting-patterns-the-future-of-the-logistics-industry.pdf

Rehof, J. (2020). Data Ecosystems: Sovereign Data Exchange among Organizations. Report from Dagstuhl Seminar 19391. Retrieved from https://www.iiconsortium.org/ doi: 10.4230/DagRep.9.9.66

Reinsel, D., et al. (2017). *Data Age 2025: The Evolution of Data to Life-Critical* (Tech. Rep.). Retrieved from www.idc.com

Reinsel, D., et al. (2020). *Worldwide Global DataSphere Forecast, 2020–2024: The COVID-19 Data Bump and the Future of Data Growth* (Tech. Rep.). Retrieved from https://www.idc.com/getdoc.jsp?containerId=prUS46286020

Rowley, J. (2012, 03). Conducting research interviews. *Management Research Review*, *35*(3-4), 260–271. Retrieved from https://click.endnote.com/viewer?doi=10.1108%2F01409171211210154&token=WzMxNTU3NDksIjEwLjExMDgvMDE0MDkxNzEyMTEyMTAxNTQiXQ.O6e2lyn0dobFsjgDzz0K67DAmbs doi: 10.1108/01409171211210154

Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., . . . Jinks, C. (2018, 07). Saturation in qualitative research: exploring its conceptualization and operationalization. *Quality and Quantity*, *52*(4), 1893–1907. Retrieved from https://link.springer.com/article/10.1007/s11135-017-0574-8 doi: 10.1007/S11135-017-0574-8/TABLES/1

Scerri, S., Tuikka, T., de Vallejo, I. L., & Curry, E. (2022). Common European Data Spaces: Challenges and Opportunities. *Data Spaces*, 337–357. Retrieved from https://link.springer.com/chapter/10.1007/978-3-030-98636-0_16 doi: 10.1007/978-3-030-98636-0{_}16

Sharma, P., Lawrenz, S., & Rausch, A. (2020, 03). Towards Trustworthy and Independent Data Marketplaces. *PervasiveHealth: Pervasive Computing Technologies for Healthcare*, 39–45. doi: 10.1145/3390566.3391687

Silva, R. M. d., Frederico, G. F., & Garza-Reyes, J. A. (2023, 02). Logistics Service Providers and Industry 4.0: A Systematic Literature Review. *Logistics 2023, Vol. 7, Page 11*, *7*(1), 11. Retrieved from https://www.mdpi.com/2305-6290/7/1/11/htmhttps://www.mdpi.com/2305-6290/7/1/11 doi: 10.3390/LOGISTICS7010011

Skjott Linneberg, M., & Korsgaard, S. (2019, 06). Coding qualitative data: a synthesis guiding the novice. *Qualitative Research Journal*, *19*(3), 259–270. doi: 10.1108/QRJ-12-2018-0012/FULL/HTML

Soto, A. C., Guerreiro Augusto, M., & Salomo, S. (2023). Building a Multi-sided Data-Driven Mobility Platform: Key Design Elements and Configurations. , 67–89. Retrieved from https://link-springer-com.tudelft.idm.oclc.org/chapter/10.1007/978-981-19-8361-0_6 doi: 10.1007/978-981-19-8361-0{_}6

Spiekermann, M. (2019, 07). Data Marketplaces: Trends and Monetisation of Data Goods. *Intereconomics*, *54*(4), 208–216. doi: 10.1007/s10272-019-0826-z

Stahl, F., Schomm, F., Vossen, G., & Vomfell, L. (2016, 08). A classification framework for data marketplaces. *Vietnam Journal of Computer Science*, *3*(3), 137–143. doi: 10.1007/s40595-016-0064-2

Tiwana, A. (2014). *Platform ecosystems: Aligning architecture, governance, and strategy*. Amsterdam: Morgan Kaufmann. Retrieved from http://www.sciencedirect.com/science/book/9780124080669

Tiwana, A., Konsynski, B., & Bush, A. A. (2010). Platform evolution: Coevolution of platform architecture, governance, and environmental dynamics. *Information Systems Research*, *21*(4), 675–687. doi: 10.1287/isre.1100.0323

TOGAF. (2023). *TOGAF Standard, Version 9.2 - Architecture Principles.* Retrieved from https://pubs.opengroup.org/architecture/togaf9-doc/arch/chap20.html

Topolšek, D., Čižiūnienė, K., & Cvahte Ojsteršek, T. (2018, 12). Defining transport logistics: A literature review and practitioner opinion based approach. *Transport*, *33*(5), 1196–1203. doi: 10.3846/TRANSPORT.2018.6965

Vaishnavi, V. K., Vaishnavi, V. K., & Kuechler, W. (2015, 05). Design Science Research Methods and Patterns : Innovating Information and Communication Technology, 2nd Edition. Retrieved from https://www-taylorfrancis-com.tudelft.idm.oclc.org/books/mono/10.1201/b18448/design-science-research-methods-patterns-vijay-vaishnavi-vijay-vaishnavi-william-kuechler doi: 10.1201/B18448

van de Ven, M., Abbas, A. E., Kwee, Z., & de Reuver, M. (2021, 07). Creating a Taxonomy of Business Models for Data Marketplaces. , 309–321. doi: 10.18690/978-961-286-485-9.23

vom Brocke, J., Hevner, A., & Maedche, A. (2020). Introduction to Design Science Research. , 1–13. Retrieved from https://click.endnote.com/viewer?doi=10.1007%2F978-3-030-46781-4_1&token=WzMxNTU3NDksIjEwLjEwMDcvOTc4LTMtMDMwLTQ2NzgxLTRfMSJd.1zOyvED1_jzBSYa1Q_vJ6Ogridk doi: 10.1007/978-3-030-46781-4{_}1

Wulfert, T., & Schütte, R. (2022, 05). Retailer's Dual Role in Digital Marketplaces: Reference Architectures for Retail Information Systems. *SN Computer Science*, *3*(3). Retrieved from https://click.endnote.com/viewer?doi=10.1007%2Fs42979-022-01098-w&token=WzMxNTU3NDksIjEwLjEwMDcvczQyOTc5LTAyMi0wMTA5OC13Il0.Hl3VxAC1wVa1tXLk95vl8jTL8Vg doi: 10.1007/S42979-022-01098-W

Yontar, E. (2022, 09). Assessment of the logistics activities with a structural model on the basis of improvement of sustainability performance. *Environmental Science and Pollution Research*, *29*(45), 68904–68922. doi: 10.1007/S11356-022-20562-X

A Overview of the sources included in the literature review

	Authors	Title	Explanation	Type of Study	Study Context
1.	Stahl et al. (2016)	A classification framework for data marketplaces	The research paper presents a framework for classifying different types of data marketplaces.	Conceptual	Generic
2.	Markus Spiekermann (2019)	Data Marketplaces: Trends and Monetisation of Data Goods	This paper explores the trends and monetization of data goods in data marketplaces.	Empirical	Multi-sided data marketplaces
3.	van de Ven, M.R. (2020)	Creating a Taxonomy of Business Models for Data Marketplaces	The paper proposes a taxonomy of business models for data marketplaces.	Empirical	Generic
4.	Bergman, R. (2020)	A Business Model Taxonomy for Data Marketplaces	This paper presents a taxonomy of business models for data marketplaces.	Empirical	Automotive Industry
5.	Koutroumpis et al. (2017)	The (unfulfilled) potential of data marketplaces	This paper examines the potential benefits of data marketplaces and identifies the factors that have hindered their growth.	Empirical	Generic
6.	Lawrenz et al. (2019)	The Significant Role of Metadata for Data Marketplaces	This research paper highlights the importance of metadata in data marketplaces and discusses the challenges associated with its management.	Empirical	Generic
7.	Koutroumpis et al. (2020a)	Markets for Data	This paper provides an overview of the characteristics, challenges, and opportunities of markets for data.	Empirical	Generic
8.	Chakrabart et al. (2017)	Business Process Modelling for a Data Exchange Platform	This paper presents a business process model for a data exchange platform.	Empirical	Generic
9.	Chakrabarti et al. (2018)	Goal-Oriented Modelling of Relations and Dependencies in Data Marketplaces	This paper proposes a goal-oriented model for data marketplaces to identify and manage relationships and dependencies.	Empirical	Generic
10	Dolci R (TUDelft)	Realising platform control in data marketplaces through Secure Multi-Party Computation	This paper proposes a solution to realize platform control in data marketplaces through Secure Multi-Party Computation.	Empirical	Generic
11.	Agahari, Wirawan; de Reuver, Mark; Fiebig, Tobias	Understanding how privacy-preserving technologies transform data marketplace platforms and ecosystems	This paper discusses how privacy-preserving technologies can transform data marketplace platforms and ecosystems.	Empirical	Generic
12.	Roman et al. (2016)	Towards a Reference Architecture for Trusted Data Marketplaces: The Credit Scoring Perspective	This paper proposes a reference architecture for trusted data marketplaces.	Conceptual	Credit Scoring Perspective

	Authors	Title	Explanation	Type of Study	Study Context
13.	Roman, D., and Vu, K. (2018)	Enabling Data Markets Using Smart Contracts and Multi-party Computation	This paper explores how smart contracts and multi-party computation can be used to enable data markets.	Conceptual	Generic
14.	Otto et al. (2019)	International Data Spaces: Reference architecture for the digitization of industries	This paper presents a reference architecture for data spaces in order to support the digitization of industries.	Conceptual	Digitization of Industries
15.	Ranganthan et al. (2019)	A Decentralized Marketplace Application on the Ethereum Blockchain	The paper presents a decentralized marketplace application on the Ethereum blockchain.	Empirical	Ethereum Blockchain
16.	Hynes et al. (2018)	A demonstration of sterling: a privacy-preserving data marketplace	This paper presents a demonstration of Sterling, a privacy-preserving data marketplace.	Empirical	Privacy-Preserving Data Marketplace
17.	Sharma, P. (2019)	Towards Decentralized Data Marketplaces	The paper explores the challenges and opportunities of decentralized data marketplaces.	Conceptual	Decentralized Data Marketplaces
18.	Oliveira et al. (2019)	Investigations into Data Ecosystems: a systematic mapping study	A research paper that presents a systematic mapping study investigating data ecosystems.	Empirical	Data Ecosystems
19.	Otto and Jarke (2019)	Designing a multi-sided data platform: findings from the International Data Spaces case	This paper presents findings from a case study on designing a multi-sided data platform, the International Data Spaces.	Empirical (single case-study)	Alliance-driven generic multi-sided data platform
20.	Gelhaar and Otto (2020)	Challenges in the Emergence of Data Ecosystems	A research paper that identifies and discusses challenges that arise during the emergence of data ecosystems.	Empirical	Generic
21.	Lis and Otto (2020)	Data Governance in Data Ecosystems – Insights from Organizations	A research paper that presents insights from organizations on data governance in data ecosystems.	Empirical	Generic
22.	Möller et al. (2020)	Towards a Method for Design Principle Development in Information Systems	The paper presents a method for developing design principles in information systems.	Conceptual	Generic
23.	Jagjit Sign Srai (2019)	Developing design principles for the digitalisation of purchasing and supply management	The paper proposes design principles for the digitalization of purchasing and supply management.	Empirical	Purchasing and supply management
24.	Jannik Lockl, Vincent Shlatt (2019)	Toward Trust in Internet of Things Ecosystems:Design Principles for Blockchain-Based IoT Applications	The paper presents design principles for building trust in IoT ecosystems using blockchain-based applications.	Conceptual	IoT ecosystems

B Invitation E-mail Interview Participation

Dear Mr./Mrs,

I hope you are doing well.

My name is Saga Kassa, an MSc. Management of Technology student at Delft University of Technology. I am currently carrying out my Master Thesis with the research objective of designing an open, decentralized multi-sided data marketplace for logistics data. A multi-sided data marketplace is essentially a digital platform that connects different types of market participants (e.g., data buyers with sellers) and facilitates their transactions (please see attached figure).

The data infrastructure of the Dutch logistics industry is characterized by large data silos, meaning data sets of individual organizations are often inaccessible to external organizations. Whereas larger logistic service providers increasingly share their data with their supply chain partners, little data is shared with industry competitors or across industries for fear of harming their interests, among others. Data transactions at present thus occur in low volumes within closed data sharing platforms or bi-lateral transactions. Recent research in data marketplaces implies so-called open, multi-sided decentralized data platforms as a potential solution to enable large-scale data sharing. Although various publications have already discussed the requirements for a generic platform, the domain-specific requirements for logistics data are, however, not yet known.

I wish to conduct expert interviews to gather insights about industry-specific needs, concerns, and/or any other constraints that must be incorporated as requirements in the architecture design of such an open logistics data-sharing platform. Therefore, I will be deeply appreciative if you have time for an interview in the forthcoming weeks. This interview lasts approximately 60 minutes. If interested, please let me know your availability and I will follow up. Last but not least, please feel free to let me know if you happen to know anybody I should also ask for an interview.

As a way to express my sincere gratitude for your contribution of time and knowledge, I will make sure to send a copy of the final report and, if interested, an invite to the presentation of the research. Looking forward to hearing from you, enjoy your week!

Kind regards,

Saga Kassa

C Interview Questions

Design of An Open, Decentralized Multi-sided Data Marketplace for Logistics Data Interview Questions for Requirement Elicitation

Welcome

Thank you for making time to meet with me for this interview. My name is Saga Kassa, student Master Management of Technology at the TU Delft Faculty of Technology, Policy and Management. I conduct this research as part of my Master Thesis concerning the design of an open, decentralized multi-sided data marketplace for logistics data. I will elaborate this term further at the start of the interview. I conduct these interviews to gather insights about industry specific needs, related concerns, and/or any other constraints that must be incorporated in the architecture design of such a logistics data-sharing platform.

Informed Consent

Before we start the interview, however, I would also like to inform you about the way I will handle your personal information as well as the generated interview data. First, with regards to your personal information; I only collect your title, name, your e-mail and your signature as part of the informed consent form. The consent form is to verify that you have been adequately informed about your rights and the study you are participating in. You will always retain the right to receive access to your personal information I collected and/or have the right to make me rectify or destroy the data. During this study, your personal information is only accessible to the researcher (which is me) and the supervisors of this study. All personal information remain stored on my laptop and will be destroyed 1 month after the research. Second, the data generated during the interview may, anonymized, be made available for others by openly sharing this data through the TU Delft repository portal, and may therefore be used for (other) scientific publications and presentations. Finally, you always keep the right to withdraw from this research without any reason.

1. Can you briefly introduce yourself?
2. What are your thoughts about sharing and monetizing commercial logistics data through open, decentralized multi-sided data marketplaces?
3. What needs or concerns do you think a logistics data provider may have related to participating in an open, decentralized multi-sided data sharing platform?
4. What needs or concerns do you think a logistics data consumer may have related to participating in an open, decentralized multi-sided data sharing platform?
5. How would you prioritize these needs, concerns or constraints in terms of most important to least important when we consider their impact on the long-term growth of the platform?
6. Could you describe your ideal version of an open, decentralized multi-sided data sharing platform for logistics data?
7. What use-cases do you think will be interesting as a demonstration for the usability of the architecture?
8. Is there anything else you think I should know that we haven't covered during this interview or a problematic question you'd like to return to?

D Data Sharing Concerns

Table 26: All elicited concerns

User		Concern	Mentions	Description
Data Provider	1	Loss of competitive advantage	I5, I4, I2, I1, I7,	Concerned with losing competitive advantage over sharing data that may inadvertently disclose sensitive information to competitors.
	2	Shipper's ownership rights	I5, I6, I2, I1,	Data providers do not always have ownership of the data they provide, so there is a concern that data owners will feel they have no control or insight in what is happening with the data.
	3	Data confidentiality	I3, I2, I1, I6, I7,	Even if the data is shared, data providers are still concerned of unauthorized access by malicious actors.
	4	Usage control	I5, I6, I3, I7,	Concerned with the data being used in a way that was not agreed upon.
	5	Legal concerns	I6, I1,	Unaware of the relevant compliance procedures on laws and regulations that govern the collection, use, and sharing of logistics data.
	6	Large investment for providers	I2, I1, I6,	Concerned with the high cost aspect of the initial investment. As the logistics is a cost-sensitive industry, with tight margins, logistics service providers must minimize their costs to remain competitive.
Data Buyer	7	Low data quality	I6, I3, I2, I1,	Concerned with inaccurate, incomplete or outdated data resulting in misinformed decision making.
	8	Low data coverage	I6, I2, I1,	Concerned there will be an insufficient variety of logistics service providers offering data thereby limiting the application of the data assets.
	9	Pricing of data sets	I6, I1,	Concerned with overvaluation and inconsistent pricing of data assets.
	10	Real-time data processing	I2, I7,	As time sensitivity is a key aspect of logistics, the need for real-time processing and analysis of data was expressed.
	11	Insufficient data supply	I6,	Concerned there will be inadequate amounts of data in the same format and context to conduct comprehensive and rigorous analysis.
	12	Understanding data sets	I2,	Concerned that understanding and generating insights from the data sets is difficult due to the sheer volume and lack of context.
Shared	13	Heterogeneous IT systems	I6, I4, I1, I7,	Many different IT systems such as ERPs, CRMs, TMS, WMS, SCM amd GPS generate relevant data in logistics and therefore will require interoperability to fetch data and upload insights.
	14	Lack of in-house technical capability	I6, I2,	Concerned that there is inadequate technical capability to setup and maintain intensive data-sharing initiatives.
	15	Neutral platform ownership	I6, I5, I7,	Concerned of conflict of interest from the platform owner resulting in unfair market competition if the platform owner is competing in the market.
	16	Lock in effect	I6, I4, I7,	Concerned that using the platform causes dependency to the point it becomes difficult or costly to switch to an alternative platform or third-party service provider.

E Architecture Requirements Derived Objectives

	Requirement Title	Derived Objective	Priority
F-1	Data Control and Selective Sharing	Enabling data providers to selectively share data and regulate the granularity of the disclosed information	Emphasized
F-2	Data Security and Access Control	Guaranteeing that only authorized users can access the shared data.	Emphasized
F-3	Usage Agreement Enforcement	Mitigating non-compliant use of data	Emphasized
F-4	Interoperability and Data Integration	Integration with existing logistics systems, streamlining the exchange of data across diverse formats, protocols, and systems.	Emphasized
F-5	Data Validation and Quality Assurance	Validating data accuracy, completeness, and timeliness for reliable decision-making	Emphasized
F-6	Broad Data Provider Participation	Easy access to extensive and diverse range of logistics data	Emphasized
F-7	Transparent and Fair Data Pricing Mechanism	Transparency and consistency in the valuation of data sets	Standard
F-8	Real-Time Data Processing and Analysis	Ability to process and derive insights from real-time logistics data	Standard
F-9	Data Standardization and Contextualization	Ensuring shared data sets conform to unified formats, schemas, and contexts for simplified aggregation, comparison, and analysis.	Standard
F-10	Data Processing and Analytical Tools	Ability or support to process, analyse, and derive insights from extensive volumes of unstructured data sets.	Standard

Legend: Color schema: Blue = Data Provider, Yellow = Data Buyer, Green = Shared

	Requirement Title	Derived Objective	Priority
NF-1	Data Ownership and Attribution	Ensuring data provider retains control and oversight of their data assets as well as making clear what the roles and responsibilities are of the marketplace users.	Emphasized
NF-2	Low-Cost Entry and Scalability	Cost-effective architecture for data providers with limited resources with the possibility to scale as needed.	Emphasized
NF-3	Legal and Regulatory Compliance	Compliant with applicable legal and regulatory requirements related to data privacy and protection.	Standard
NF-4	User-Friendly Interface and Documentation	Minimizing financial and technical entry-barriers for logistics service providers with little technical expertise.	Standard
NF-5	Neutral Platform Ownership	Fair and impartial governance of the platform by the owner, avoiding any conflict of interest.	Standard
NF-6	Avoid Vendor Lock-in	Ability to easily switch to alternative platforms or third-party service providers	Standard

Legend: Color schema: Blue = Data Provider, Green = Shared