UNIVERSITY OF TWENTE.

ENGD THESIS

DESIGNING ESSENTIAL COMPONENTS FOR LOGISTICS DATA SPACES

CONNECTING LOGISTICS INTERFACES, CONVERTERS, KNOWLEDGE, AND STANDARDS

DANNIAR REZA FIRDAUSY

DESIGNING ESSENTIAL COMPONENTS FOR LOGISTICS DATA SPACES

Connecting Logistics interfaces, Converters,
Knowledge, and Standards

Danniar Reza Firdausy

DESIGNING ESSENTIAL COMPONENTS FOR LOGISTICS DATA SPACES

EngD Thesis

to obtain the degree of
Engineering Doctorate (EngD) at the University of Twente,
on the authority of the rector magnificus,
prof. dr. ir. A. Veldkamp
on account of the decision of the graduation committee,
to be defended
on Wednesday the 5th of July 2023 at 12.45 hours

by

Danniar Reza Firdausy

born on the 26th of February 1996
in Surabaya, Indonesia

“The best way to predict your future is to create it.”

Abraham Lincoln

ISBN (print) : 978-90-365-5716-0
ISBN (digital) : 978-90-365-5717-7
DOI : 10.3990/1.9789036557177
URL : https://doi.org/10.3990/1.9789036557177

Printed by Gildeprint
Cover design Viriega Fauzia R.

This EngD thesis was embedded under the collaborative research and development project called CLiCKS, which is financially supported by the Dutch Ministry of Economic Affairs and co-financed via TKI DINALOG and NWO (grant no. 439.19.633). The authors express their gratitude to the consortium partners involved in this project, whose support and contribution was instrumental in the success of this EngD thesis.

Supervision

This EngD Thesis has been approved by:

Thesis Supervisor	: Prof. Dr. Maria-Eugenia Iacob
Co-Supervisor	: Dr. Ir. Marten J. van Sinderen

Graduation Committee

Chairman:	
Thesis Supervisors:	
Prof. Dr. Maria-Eugenia Iacob	University of Twente
Dr. Ir. Marten J. van Sinderen	University of Twente
Daily Supervisors:	
Dr. J.L. Rebelo Moreira	University of Twente
Company Supervisor:	
Samet Kaya	eMagiz Services B.V.
EngD BIT Program Director:	
Dr. L. Ferreira Pires	University of Twente
External Examiners:	
Dr. Ir. Mark de Reuver	Delft University of Technology
Dr. Patrício de Alencar Silva	Universidade Federal Rural do Semi-Árido
Jean Paul Sebastian Piest	Port of Twente

Acknowledgment

A wise person once said, one may fail several times along the way, but one should not stop pursuing what they are after. This EngD journey has taught me that success is attained through diligent effort, continuous learning, personal sacrifices, unwavering perseverance, and a genuine passion for our chosen endeavors, all accompanied by the ongoing support of those around us. Despite the period of this EngD project being only 2 years, it has afforded me invaluable exposure to a wide range of dynamics in academia, from the academic coursework and professional developments, apart from the design project itself, to academic paper publications and master thesis supervisions. These achievements would not have been possible without the invaluable support of those surrounding me.

I consider myself extremely fortunate to have Prof. Dr. Maria E. Iacob and Dr. Ir. Marten J. van Sinderen as my thesis supervisors, who have been guiding me since my master's thesis at the University of Twente. Thanks to their supervision, I have been granted the opportunity to apply my knowledge and contribute to the success of the CLiCKS project under their guidance. I am also deeply grateful to them for consistently making the time in their busy schedules to provide guidance and feedback on my work and progress. I consider Maria as my second mother at UT, who has always given me life and career advice that I know I will need in the future. Her role as a mother figure in my academic life is further exemplified by her continuous monitoring of my progress in completing my EngD journey. Likewise, I regard Marten as my second father at the University of Twente who has consistently pushed me and provided unwavering support to level up the quality of my deliverables to complete my EngD thesis. I consistently hold the impression that both of you form an exceptional team in guiding my progress toward achieving remarkable results. Therefore, I express my deepest gratitude for all the support you have given me.

My deepest gratitude also extends to those who have provided their tangible contribution to the quality of my work and the success of the CLiCKS project. I sincerely thank Dr. Patrício de Alencar Silva for his dedicated contribution in enhancing this project's deliverables and co-authoring all associated paper publications associated to this EngD thesis. My appreciation extends to Dr. João L. Rebelo Moreira, for his invaluable knowledge to support my deliverables and supervise master theses derived from this EngD thesis. My gratitude also extends to an inspiring colleague, J. P. Sebastian Piest, who has been motivating me to become a proficient industry practitioner capable of bridging the gap between academia and industry. I'm grateful to Dr. Harrie Bastiaansen and Wout van den Heuvel from TNO and SUTC for their generous time, expertise, and valuable insights into the industrial implications and relevance of the deliverables for this EngD thesis. I also extend my appreciation to Samet Kaya, Erik Bakker, and Geert-Jan Waanders from eMagiz Services B.V. for their invaluable support and training in utilizing the eMagiz iPaaS platform as an alternative to the proposed design. Additionally, my gratitude also goes to Dr. Ir. Mark de Reuver from TU Delft for dedicating his time

and interest to review and participate in the defense committee of this EngD thesis. I am also grateful to Antragama Ewa Abbas from TU Delft for his enthusiastic knowledge sharing and contributions to improving the results of this EngD thesis.

I would like to extend my heartfelt appreciation to my family and friends, who have consistently offered their endless support and prayers throughout all the highs and lows of my journey. A special acknowledgment goes to my father, mother, and older brother, whose love and guidance have been instrumental in shaping the person I am today. They have nurtured me and provided me with the best education and life lessons, enabling me to forge my own path and build a successful career. I also thank colleagues and friends who have provided tremendous support for both my progress and well-being throughout my EngD journey. A special thanks goes to my IEBIS colleagues, including Martijn Mes, Elke, Gea, Iqbal, Yifei, Fabian, Robert, Antonio, Rogier, Martijn Koot, Yasir, Giovanni, and many others. Also, thanks to all of my EngD colleagues, in particular: Donika, Mohammad, Farideh, Thomas, Frank, Lexi, and many others, who have been stout comrades in facing the challenges and intricacies of our EngD program. My thanks extend to other friends, including Dea, Chris, Tifani, Michael, Amy, Jesús, Raef, Nikitha, Jeroen, and many others, who have been a passionate bunch of people to forget our academic life and have some fun. A special thanks goes to Yesaya and Aldi for their enthusiasm in expanding this work for their master theses, as well as to Tifani for providing the necessary deployment infrastructure for our prototypes. Last but not least, a great appreciation goes to Zaiimatul Ummah for being a remarkable partner and continuously inspiring each other to pursue personal growth and chase our goals. Her constant attention, admiration, and endless encouragement have been the driving force behind my continuous pursuit of achievements. Thus, I feel incredibly lucky to have her by my side for always pushing me to reach new heights in the future.

Finally, I would like to express my gratitude to the Almighty God, as I firmly believe that all the achievements I have accomplished are blessings from God. *"Lā ḥawla wa-lā quwwata ʾillā bi-llāh," which means "There is no power or strength except through God"*. May God bless you.

Summary

Data-sharing, in today's business environment, is perceived by many industrial sectors as the key to unlocking innovative and mutually beneficial business models. Past studies in several business domains reported that IT-supported data-sharing has enabled supply chain partners to improve their core business. Especially in the transport logistics sector, IT allows LSPs and their customers to share information such as purchase orders and shipment details to support smooth business process coordination and other operational purposes. Despite that, establishing a data-sharing ecosystem comes with its own set of impediments, such as conflicting data formats and schema. These challenges make the process of connecting companies' enterprise systems to data-sharing platforms challenging. The lack of technical enforcement in disclosing their data, particularly, how their data is going to be accessed, used, and proliferated, sets back companies' willingness to share data even more. These concerns in the data-sharing led to the adoption of the IDS initiative that puts forwards trust, security, interoperability, and data sovereignty in mind by distributing the responsibility of establishing such a data-sharing ecosystem into several trusted business roles and application components, such as the IDS Connector and IDS Data Apps. However, instantiating a complete IDS ecosystem calls for an elaborate effort, and participating and exchanging data in data space is not yet seen as economically attractive. Furthermore, prior to joining a data space, candidate data space's participants will also need to explore available IDS Connector(s) suitable for their use case and capability. This situation calls for the development of a Connector Store as a broker system that can facilitate the discovery and selection of IDS Connectors, data sources, as well as participants active in a data space.

Given these project motivations and problem statements, this EngD thesis aims to investigate a suitable design of a Connector Store and other related application components comprising a data space for the logistics sector essential for managing data interoperability, data sovereignty, and resource discovery. The design process starts with a Problem Investigation that identifies (1) the stakeholders involved in this design project and their expectations, (2) the challenges of data-sharing in the logistics industry and the latest solution to such problems, and (3) the interplay between the proposed Connector Store with other business roles and application components in managing data interoperability, data sovereignty, and resource discovery. Next, the work is followed with a Treatment Design, in which requirements elicitation and software and enterprise architecture design of the relevant application components for managing data interoperability, data sovereignty, and resource discovery take place. Then, this thesis is finalized with a Treatment Validation, in which we (1) instantiate a logistics data space demonstrator through the development of the essential application components based on the proposed design and (2) validate the contribution of the logistics data space demonstrator to achieving stakeholders' goals of managing data interoperability, data sovereignty, and resource discovery.

From the demonstration and interview with experts, we reached the conclusion that the proposed architecture of the logistics data space is suitable for participants in a real-world scenario to handle these three issues. The IDS Data Apps requested from a Connector Store and orchestrated by IDS Connectors can support data space participants solve data interoperability problems to a certain extent. The implementation of the data usage policy enforcement as accommodated by the proposed architecture also connects well with the need for a technically enforced trust for future ad-hoc data exchange. To further promote the value of realizing such a logistics data space, there is also a proposition resulting from the panel discussion to imbue the ecosystem with complementary customized brokering and consulting services. Based on these results, we argue that this EngD thesis has managed to (1) demonstrate the technical feasibility of developing the proposed IDS-compliant logistics data space brokered by a Connector Store to lower data interoperability, data sovereignty, and resource discovery issues, and (2) provide the lesson learned from the development and validation that can serve as a basis for the future related research endeavor.

Keywords: Data Interoperability, Data Sovereignty, Service Discovery, Data Exchange, Connector Store, International Data Spaces, Enterprise Architecture

Samenvatting

In het huidige bedrijfsklimaat wordt het delen van gegevens door veel industriële sectoren gezien als de sleutel tot het ontsluiten van innovatieve en wederzijds voordelige bedrijfsmodellen. Uit eerdere studies in verschillende bedrijfssectoren blijkt dat door IT ondersteunde gegevensuitwisseling de partners in de toeleveringsketen in staat heeft gesteld hun kernactiviteiten te verbeteren. Vooral in de sector vervoerslogistiek stelt IT-aanbieders en hun klanten in staat om informatie zoals aankooporders en verzendingsgegevens te delen ter ondersteuning van een soepele coördinatie van bedrijfsprocessen en andere operationele doeleinden. Desondanks brengt de totstandbrenging van een ecosysteem voor gegevensuitwisseling een aantal hindernissen met zich mee, zoals conflicterende gegevensformaten en -schema's, die de aansluiting van bedrijfssystemen van bedrijven op platforms voor gegevensuitwisseling bemoeilijken. Zelfs als een dergelijk probleem van gegevensinteroperabiliteit wordt aangepakt, vormt het gebrek aan technische handhaving bij het bekendmaken van hun gegevens, met name hoe hun gegevens zullen worden opgevraagd, gebruikt en verspreid, een nog grotere belemmering voor de bereidheid van bedrijven om gegevens te delen. Deze bezorgdheid over het delen van gegevens heeft geleid tot de aanneming van het IDS-initiatief dat vertrouwen, veiligheid, interoperabiliteit en gegevenssoevereiniteit naar voren schuift door de verantwoordelijkheid voor het opzetten van een dergelijk ecosysteem voor gegevensuitwisseling te verdelen over verschillende vertrouwde bedrijfsrollen en applicatiecomponenten, zoals de IDS Connector en IDS Data Apps. Het opzetten van een volledig IDS-ecosysteem vergt echter een grote inspanning, en het deelnemen aan en uitwisselen van gegevens in een dataruimte wordt economisch nog niet aantrekkelijk geacht. Voorts zullen de deelnemers aan een dataruimte, voordat zij zich bij een dataruimte aansluiten, ook moeten onderzoeken welke IDS Connector(en) beschikbaar zijn die geschikt zijn voor hun gebruik en mogelijkheden. Een dergelijke situatie vraagt om de ontwikkeling van een Connector Store als een makelaarssysteem dat de ontdekking en selectie van IDS-connectoren, gegevensbronnen en deelnemers die actief zijn in een gegevensruimte kan vergemakkelijken.

Gezien deze projectmotivaties en probleemstellingen beoogt deze EngD thesis een geschikt ontwerp te onderzoeken van een Connector Store en andere gerelateerde toepassingscomponenten die een gegevensruimte vormen voor de logistieke sector die essentieel is voor het beheer van gegevensinteroperabiliteit, gegevenssoevereiniteit en de ontdekking van bronnen. Het ontwerpproces begint met een Probleemonderzoek dat (1) de bij dit ontwerpproject betrokken belanghebbenden en hun verwachtingen identificeert, (2) de uitdagingen van gegevensuitwisseling in de logistieke sector en de nieuwste oplossing voor die problemen vaststelt, en (3) de wisselwerking tussen de voorgestelde Connector Store en andere bedrijfsrollen en toepassingscomponenten bij het beheer van gegevensinteroperabiliteit, gegevenssoevereiniteit en het ontdekken van bronnen, zoals voorgeschreven door de IDSA. Vervolgens wordt het werk gevolgd door een Behandelingsontwerp, waarin het opwekken van eisen en het ontwerpen van

software en bedrijfsarchitectuur van de relevante applicatiecomponenten voor het beheer van gegevensinteroperabiliteit, gegevenssoevereiniteit en het vinden van bronnen plaatsvindt. Vervolgens wordt het proefschrift afgesloten met een Behandelingsvalidatie, waarin we (1) een logistieke dataruimte demonstrator instantiëren door de ontwikkeling van de genoemde essentiële applicatiecomponenten op basis van het voorgestelde ontwerp en (2) de bijdrage van de logistieke dataruimte demonstrator valideren aan het bereiken van de doelstellingen van de belanghebbenden voor het beheer van gegevensinteroperabiliteit, gegevenssoevereiniteit en het ontdekken van hulpbronnen.

Uit de demonstratie en het gesprek met deskundigen hebben wij geconcludeerd dat de voorgestelde architectuur van de logistieke dataruimte de deelnemers aan een reëel scenario goed in staat stelt deze drie problemen aan te pakken. De IDS Data Apps die uit een Connector Store worden opgevraagd en door IDS Connectors worden georkestreerd, kunnen tot op zekere hoogte de deelnemers aan de gegevensruimte ondersteunen bij het oplossen van interoperabiliteitsproblemen. De uitvoering van het beleid inzake gegevensgebruik zoals dat in de voorgestelde architectuur is opgenomen, sluit ook goed aan bij de behoefte aan een technisch afgedwongen vertrouwen voor toekomstige ad hoc gegevensuitwisseling. Om de waardepropositie van het realiseren van een dergelijke logistieke dataruimte verder te bevorderen, is er ook een voorstel uit de paneldiscussie om het ecosysteem te voorzien van aanvullende op maat gesneden tussenhandel en adviesdiensten. Op basis van deze resultaten stellen wij dat deze doctoraalscriptie erin geslaagd is om (1) de technische haalbaarheid aan te tonen van de ontwikkeling van de voorgestelde IDS-conforme logistieke dataruimte, bemiddeld door een Connector Store, om de problemen inzake gegevensinteroperabiliteit, gegevenssoevereiniteit en ontdekking van bronnen te verminderen, en (2) de lessen te verstrekken die uit de ontwikkeling en validering zijn getrokken en die als basis kunnen dienen voor toekomstige verwante onderzoeksinspanningen.

Trefwoorden: Data Interoperability, Data Sovereignty, Service Discovery, Data Exchange, Connector Store, International Data Spaces, Enterprise Architecture

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

List of Acronyms

API	Application Programming Interface
AS4	Applicability Statement 4
AWS	Amazon Web Service
BSP	Broker Service Provider
B.V.	*Besloten Vennootschap* (Private Company)
CBS	*Centraal Bureau voor de Statistiek*
CDM	Common Data Model
CLiCKS	Connecting Logistics interfaces, Converters, Knowledge, and Standards
CQs	Competency Questions
CRM	Customer Relationship Management
CSV	Comma Separated Values
DBMS	Database Management System
DCAT	Data Catalog Vocabulary
DINALOG	Dutch Institute for National Advance LOGistics
DLT	Distributed Ledger Technology
DSRM	Design Science Research Methodology
EA	Enterprise Architecture
EDI	Electronic Data Interchange
EngD	Engineering Doctorate
EPCIS	Electronic Product Code Information System
ERD	Entity Relationship Diagram
ERP	Enterprise Resource Planning
ESB	Enterprise Service Bus
FAIR	Findability, Accessibility, Interoperability, Reusability
GS1	Global Standards 1
GUI	Graphical User Interface
HTTPs	Hypertext Transfer Protocol Secure
IDS	International Data Spaces
IDSA	International Data Spaces Association

IDS IM	International Data Spaces Information Model
IDS RAM	International Data Spaces Reference Architecture Model
IoT	Internet-of-Things
iPaaS	Integration Platform as a Service
ISP	Integration Service Provider
JSON	JavaScript Object Notation
JSON-LD	JavaScript Object Notation-Linked Data
MQTT	Message Queuing Telemetry Transport
NWO	*Nederlandse Organisatie voor Wetenschappelijk Onderzoek* (Dutch Organization for Scientific Research)
ORSD	Ontology Requirements Specification Document
OS	Operating System
OTM	Open Trip Model
OWL	W3C Web Ontology Language
P2P	Peer-to-Peer
RDF	Resource Description Framework
REST	REpresentational State Transfer
SCSN	Smart Connected Supplier Network
SLR	Systematic Literature Review
SMEs	Small and Medium-sized Enterprises
SOA	Service-Oriented Architecture
SOAP	Simple Object Access Protocol
SSH	Secure Socket Shell
SPARQL	SPARQL Protocol and RDF Query Language
SUTC	*Stichting Uniforme Transport Code* (Uniform Transport Code Foundation)
TKI	Topconsortia for Knowlegde and Innovation
TLN	*Transport en Logistiek Nederland* (Transport and Logistics Netherlands)
TNO	*Nederlandse Organisatie voor toegepast-natuurwetenschappelijk Onderzoek* (Netherlands Organization for Applied Scientific Research)
TOGAF	The Open Group Architecture Framework

TOGAF ADM	The Open Group Architecture Framework Architecture Development Method
TPS	Transaction Processing System
UBL	Universal Business Language
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
UML	Unified Modelling Language
UUID	Universally Unique Identifier
VESDI	Vehicle Emission Shipment Data Interface
VPS	Virtual Private Server
WoS	Web of Science
XML	eXtensible Markup Language
XSD	XML Schema Definition

1 Introduction

This first chapter presents the general context and problem statement that will be addressed in this collaborative research and development project called CLiCKS, in which this EngD thesis was embedded. **Section 1.1** introduces the background and scope of this CLiCKS project. **Section 1.2** discusses the motivation that drives this EngD thesis. **Section 1.3** establishes the problems that arise and the goal that this thesis aims to achieve. **Section 1.4** lays out how these problems will be addressed by defining the research questions. Next, **Section 1.5** describes how this thesis is designed to answer these research questions. **Section 1.6** closes this chapter by presenting an overview of this thesis's structure.

1.1 CLiCKS Project

This design project is part of the work package defined within the CLiCKS Project, which is financially supported by the Dutch Ministry of Economic Affairs via TKI DINALOG and NWO under Accelerator 2019 (grant no. 439.19.633)[1]. CLiCKS is the acronym that stands for Connecting Logistics interfaces, Converters, Knowledge, and Standards. One of the focuses of the CLiCKS project is to design and propose an approach and solutions to make the exchange and sharing of real-time data more accessible and secure to logistics SMEs. To achieve this, one of the sought approaches is to adopt IDS-based Connectors and other relevant components. By supporting real-time data-sharing between logistics SMEs, the CLiCKS project aims to enhance end-to-end supply chain visibility, making logistics resources utilization and dynamic planning possible and more efficient in the process. For such a focus, another focal point of the project is to valorize previous research and development outputs. These outputs take the form of approaches, frameworks, standards, or tools that are relevant within transport and logistics companies. For this purpose, this project explores the adoption and testing of standards and new technologies, such as the Open Trip Model[2] (OTM) and the International Data Spaces[3] (IDS). This is to investigate and evaluate the implementation of suitable agreements and schemes for data-sharing scenarios in the logistics sector.

To achieve these goals, the CLiCKS project proposes the design of a demonstrator called the "logistics data space". This development of this demonstrator is divided into two EngD work packages planned for each other to be synchronized. This thesis, the design of a Connector Store, is the first one and the design of an Interoperability Simulator is the second one. The objectives of this thesis are (1) to facilitate the connection between logistics IDS components, government (open) data, and other communication platforms, and (2) to make available various collaborative

[1] https://www.nwo.nl/en/projects/43919633-0
[2] https://www.opentripmodel.org/
[3] https://internationaldataspaces.org/

services and data infrastructure relevant to the coordination and optimization of logistics processes by both SMEs and government organizations. For these objectives, this thesis focuses on the design of a Connector Store with predefined IDS Connectors to support logistics companies to participate in a collaboration network through the selection of suitable IDS Connectors and IDS Data Apps from the store.

1.2 Project Motivation

In today's business environment, data is seen as an essential asset for economic development. Sharing this data is perceived by many industrial sectors as the key to unlocking innovative and mutually beneficial business models. Past studies in several business domains reported that IT-supported data sharing among value partners enabled them to improve their planning processes as well as to stimulate innovative business models (Banek et al., 2008; Markus & Bui, 2012; Wang, X. et al., 2018). In the transport logistics sector, IT allows Logistics Service Providers (LSPs) and their customers to share information such as purchase orders and shipment details to support smooth business process coordination and other operational purposes (Pham et al., 2019). Improved end-to-end visibility in the supply chain can also be achieved by sharing and making use of real-time data. In the recent development, the sharing of real-time IoT sensor data enables logistics companies to mitigate disruptions and handle exceptions by embedding such data in the transportation planning process (Iacob et al., 2019). In another case, sharing data regarding truck parking spaces and occupancy rates from truck parking owners helps transport planners to optimize transport planning and dynamic rerouting (Slavova, 2021). These developments capture the growing industrial demand for business ecosystems where companies in the logistics sector can share data to stimulate innovation and business performance.

Establishing a data-sharing ecosystem comes with a set of challenges. First of all, as noted by previous studies, conflicting data formats and schema make the process of connecting companies' enterprise systems to data-sharing platforms challenging (Piest, Meertens, et al., 2020). In the Dutch logistics sector alone, a multitude of standards are currently present (e.g., OTM5, GS1, EDIFACT) (Bol Raap et al., 2016; GS1, 2021; OpenTripModel, 2021). Such data interoperability issue poses a barrier for companies to exchange data and fully exploit the potential of data sharing. Second of all, the lack of sovereignty over how their data is going to be accessed, used, and proliferated also hinders logistics companies to share data (Dalmolen, Bastiaansen, Somers, et al., 2019). The insufficiency of technical enforcement in disclosing their data sets back companies' willingness to share data even more. This is crucial since this data holds strategic value for logistics companies, and thus the ability to retain control over their valuable asset determines their willingness to collaborate (Piest, Iacob, et al., 2020). From there on, then, comes the third challenge that revolves around which solutions will become the chosen one. Even if a particular solution is selected, one technology adopter will then question if its partners will also adopt the same solution. Alternatively, business entities also question, which potential partners then have already adopted that solution who can

promise mutual benefits in future collaborations. As a result, companies find it difficult to make upfront investments if the contribution of establishing a data-sharing ecosystem to their profits is not apparent (McGuigan et al., 2022). These challenges, encompassing data interoperability, data sovereignty, and adoption uncertainties, on the development of such a data-sharing ecosystem, which eventually led to the adoption of the IDS initiative (Bastiaansen et al., 2020; Hofman, W., 2019).

IDS is an initiative of various international research institutes and industrial enterprises to establish a decentralized data-sharing platform in which partners from different sizes can exchange data while still granted the capability of being entirely self-determined with regard to their data (IDSA, 2019; Otto & Jarke, 2019). The IDS promotes trust, security, interoperability, and data sovereignty by distributing the responsibility in a data-sharing ecosystem into several trusted business roles and application components. IDS promises a solution to data interoperability and data sovereignty issues by prescribing the use of, among others, the IDS Connectors for companies to exchange data with each other while enforcing data usage policies at the same time. To fully benefit from a complete implementation of an IDS ecosystem, the IDSA has published the IDS RAM that provides a generalization of concepts, functionality, and overall processes involved in the creation of a secure data-sharing ecosystem (IDSA, 2019).

1.3 Problem Statement and Project Goal

Instantiating a complete IDS ecosystem with its organizational roles and technical mechanisms calls for an elaborate effort and is not yet seen as economically attractive (Firdausy et al., 2022c). Such a challenge is considered a significant barrier for small and medium enterprises (SMEs), which are constrained by limited resources and capabilities (Piest, Iacob, et al., 2020). To demonstrate the organizational and technological feasibility of building and adopting IDS-based data-sharing ecosystems, the chairman of the IDSA board has made a call for business cases and applications (Otto, 2019). This call connects well with the Dutch Topsector Logistiek[4] vision, which aims at the emergence of logistics data-sharing environments through a secure and interoperable data infrastructure (Dinalog, 2020). Therefore, this project answers such a call by developing a demonstrator of a "logistics data space" comprised of the essential application components for the core IDS business roles. This demonstrator will act as a testbed to evaluate the technical and organizational feasibility of an IDS-based data-sharing ecosystem implemented for business cases in the logistics sector.

Considering the central role of the IDS Connector in an IDS ecosystem, the development of this demonstrator needs to start with the investigation of how such an IDS Connector can be designed and configured to interact with the companies' enterprise systems and other supporting components within an IDS environment.

[4] https://topsectorlogistiek.nl/wp-content/uploads/2022/05/Topsector-Logistiek-Dutch-Industry.pdf

Although necessary, IDS Connectors alone are not sufficient to establish a fully functional IDS ecosystem. Prior to joining the data space, candidate IDS participants will first need to explore the available IDS Connector(s) suitable for their use case and capability. To satisfy the diverse sets of participants' needs and capabilities, more and more software and service providers will start to offer IDS Connectors in several configurations. At the same time, these candidates will also need to investigate the presence of their partners in the data space and the prospect of securing a strategic partnership with the existing participants. This exploration phase will signify the value they can gain from participating in the ecosystem, which in turn, influences their adoption level of the IDS vision. Moreover, after the data users acquire the suitable IDS Connector, they will need to discover the available data sources before they can initiate a data exchange with the data owners. Similarly, data owners will also need to obtain and use an IDS Connector to describe and share their data with other participants in the data space. As a result, there is a need for a broker system that can facilitate the discovery of the IDS Connectors, data sources, as well as participants that are active in a data space. The IDSA prescribed that the presence of a metadata broker system is necessary to facilitate the discovery of IDS-related resources and components (IDSA, 2019). The implementation of such a broker system answers the call for an IDS-based infrastructure to facilitate the connection between logistics partners with other than IDS-based coordination platforms (Piest, Iacob, et al., 2020). Considering the open opportunity for a parallel implementation of such a metadata broker for an IDS-based ecosystem (Bader, Bruckner, et al., 2020), the CLiCKS project proposes the development of a Connector Store, which is a repository of metadata that aims to support the discovery and selection of IDS Connectors and data sources in an IDS ecosystem.

This EngD thesis, at the end, contributes to both the industry and academia with a reference architecture of a Connector Store to support logistics companies' onboarding to a logistics data space. Such reference architecture is contextualized to the logistics sector in The Netherlands to answer the Dutch Topsector Logistiek's call. The research contribution (RC) provided by the reference architecture is threefold:

RC 1. The reference architecture provides a better understanding of the essential application components, business roles, and processes required for companies in the Dutch logistics sector to establish and participate in an IDS-based logistics data-sharing ecosystem. For this, a multi-viewpoint design approach to separate concerns that cover at least business, information, and process levels will be adopted (Cicchetti et al., 2019).

RC 2. The reference architecture provides viewpoints from the perspective of enterprise and software architecture of a Connector Store, which is responsible for the discoverability of IDS Connectors, data sources, and participants that are active in a logistics data space. This discovery, in the process, covers the selection and provisioning of IDS Connectors provided by Software Providers for participants to participate in the data space.

RC 3. The reference architecture provides viewpoints from the perspective of enterprise and software architecture of relevant IDS components, which are responsible for facilitating data interoperability and enforcing data sovereignty in data transactions at runtime. Such architectural viewpoints will shed some light on how the essential IDS components for a data space can be developed.

Through these contributions, this thesis will evaluate how well the Connector Store can encourage companies to participate in the IDS-based logistics data-sharing environment through the discoverability of the IDS-related components and resources. Next to this, this work will also assess to what extent the relevant components guided by the IDS specifications can solve issues regarding data interoperability and data sovereignty.

1.4 Project Design Questions

Based on the background, context, and objective, we formulate the main research question (RQ) to be answered by this thesis as follows.

Main Design Question:

How to improve data sharing in the logistics sector by designing the essential components for a Dutch Logistics Data Space that satisfies requirements on data interoperability and data sovereignty, such that logistic processes across the supply chain can be more efficient?

By Dutch Logistics Data Space here we mean an IDS-based ecosystem where Dutch logistics companies can share data with a focus on lowered data interoperability barriers and enhanced data sovereignty. This main research question is then refined into the following sub-questions:

Sub Questions:

SQ 1. What is state-of-the-art on data-sharing in the logistics industry?

The goal of this sub-question is to gain an understanding of the current situation of data-sharing in the logistics industry and identify the problem within this area. For this purpose, in **Chapter 3**, this thesis will discuss a systematic literature review that focuses on (1) investigating the main challenges of companies and organizations in the logistics industry to participate in a data-sharing environment, as well as (2) exploring the proposed solutions to solve these challenges.

SQ 2. What is a suitable architecture to establish an IDS-compliant data-sharing ecosystem?

Subsequently, we intend to investigate a design that is prescribed by the IDS to treat the data-sharing challenges previously identified in **Chapter 3**. Answering this question will help us to identify what are the roles, application components, and processes involved within an IDS-based data-sharing ecosystem that is essential for the logistics data space to manage data interoperability and data sovereignty. The results of this investigation will be captured in a high-level architecture, which will be discussed in **Chapter 4**.

SQ 3. How can the underlying application components of an IDS-compliant data-sharing ecosystem be designed to manage data interoperability and data sovereignty at runtime?

Based on the high-level architecture obtained in **Chapter 4**, this sub-question aims to deliver a software requirements specifications and architectural viewpoints describing the involved application components in a more detailed way. Besides capturing the requirements of each essential application component, the architecture will also present the interaction between application components and processes of key functionalities that contribute to the management of data interoperability and data sovereignty in the data space's runtime. For this, the detailed architectural design will be discussed in **Chapter 5**.

SQ 4. How can a Connector Store be designed to support companies to discover and select the underlying application components?

Answering this question provides a set of software architecture viewpoints of the application component that supports the participants of the logistics data space to find IDS Connectors, data owners, and data resources. The aim of providing such an application component is to attract participants for adopting the IDS vision by using IDS Connectors to share data and enforce the data usage policies. **Chapter 6** and **7** will discuss how this application component can be designed and developed.

SQ 5. To what extent the IDS-compliant data-sharing ecosystem architecture can support logistics companies to manage data interoperability and data sovereignty at runtime?

This question concerns a validation step for investigating if the logistics data space demonstrator can serve the expected qualities. This question is answered by evaluating the extent to which the proposed architecture can facilitate logistics companies to tackle the challenge of conflicting data formats and technically enforce the policies to use their data at runtime. Therefore, **Chapter 8** will elaborate on how a workshop session with experts in the logistics industry is conducted to execute and evaluate the proposed architecture and its demonstrator.

Besides contributing to answering the main research question, these sub-questions also play a part to achieve the research contributions that are discussed in the previous section. Such mapping towards the contributions is then captured in **Figure 1**. Furthermore, to address these questions, the next section will discuss how this thesis is designed and adopts the design science methodology.

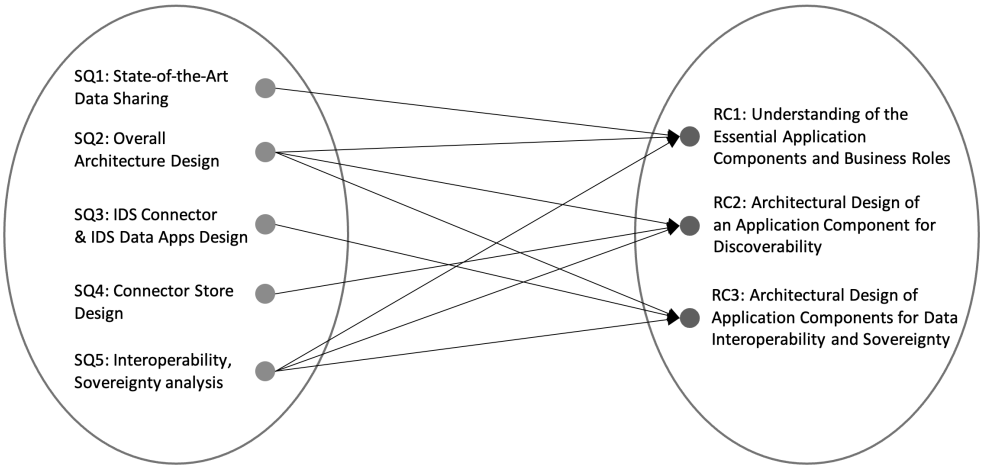


Figure 1 Mapping of the Research Sub-Questions to the Research Contributions

1.5 Project Methodology

This thesis necessitates the adoption of a design science methodology to investigate the architecture suitable to treat the problem context previously discussed. Peffers et al. (2007) introduced the design science research methodology (DSRM) as the procedural practice to create and evaluate artifacts that cover models, methods, and instantiations intended to solve identified organizational problems. The design science was then summarized by Wieringa (2014) as the study to investigate the design of an artifact to interact with a problem context for improvement in that context. This design project adopts the DSRM proposed by Wieringa, which prescribes four consecutive phases shown in **Figure 2** starting from the **Problem Investigation**, **Treatment Design**, and then **Treatment Validation** before the **Treatment Implementation**.

In the **Problem Investigation,** we start with the identification of stakeholders and their goals, which their goals will be elicited from initial input from the consortium partners involved in the project. In conjunction, a Systematic Literature Review (SLR) will be carried out. As mentioned earlier, this SLR is aimed to gain more understanding of the current situation and problem by extracting the state-of-the-art data-sharing settings in the logistics industry. Besides that, the SLR will also draw out pointers and base knowledge that complement the stakeholders' goals previously mentioned. Next to this, a supplementary literature review for extracting

IDS concepts and principles will be performed as well to support the design of the artifact in achieving the stakeholders' goals. This way, this first phase of the methodology is designed to provide answers for SQ 1 and SQ 2.

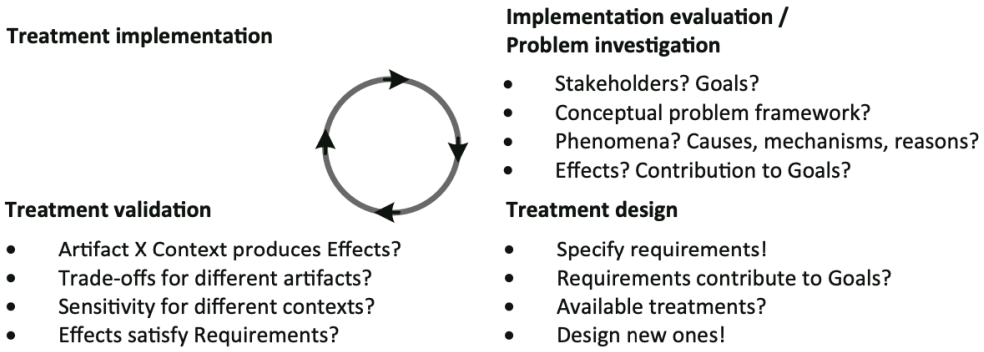


Figure 2 Design Science Methodology Engineering Cycle (Wieringa, 2014)

Next, the **Treatment Design** phase will be initiated with a requirement elicitation process. Requirements will be gathered from both the SLR and the literature review from IDS documents. The requirements related to the goal of discoverability are associated with the Connector Store, and requirements related to the goal of lowering interoperability and enforcing sovereignty are associated with the IDS components (e.g., IDS Connector, IDS Data App, and Clearing House). Subsequently, how these components' infrastructure, processes, and requirements can be aligned to contribute to the previously mentioned goals will be captured in an Enterprise Architecture (EA) model. Such a model will then be broken down into several viewpoints to separate the concerns, which in the process, will answer SQ 2, SQ 3, and SQ 4. This model will then represent the architecture for the demonstrator of a logistics data space previously discussed in **Section 1.3**. To support the validation of the produced architecture, at the end of this phase, a demonstrator comprising the prototypes of the relevant components will be developed.

Lastly, in the **Treatment Validation** phase, we will investigate the extent that the proposed architecture of the logistics data space, comprised of the Connector Store and the other IDS-related components, can support logistics companies to manage data interoperability and data sovereignty. The validation phase will be divided into three steps, in which, it adopts the single-case mechanism experiment and expert opinion methods. The first is to apply the instantiated logistics data space to solve a case study in the logistics sector. The second is to demonstrate this to a panel of experts in the logistics industry relevant to the case in a workshop session. Lastly, several questions by means of a questionnaire will be given to the experts, especially with regard to how the demonstrator is able to serve its intended purpose. The input from the experts will be captured in the form of quantitative scores using the Likert Scale as well as qualitative statements based on their opinions. Although important, **Treatment Implementation** is made out-of-scope, due to the limited timeframe and the focus of this EngD thesis to investigate on a suitable design as a proof-of-concept.

1.6 Thesis Structure

The remainder of this EngD thesis is structured as follows. **Chapter 2** identifies the stakeholders involved in this design project, their involvement, goals, and expectations from this project. **Chapter 3** conducts an SLR with the aim to (1) explore the current situation and challenges of data-sharing in the logistics industry, as well as (2) extract the latest approach to such problems. **Chapter 4** investigates the design principles prescribed by the IDSA, as well as the compliance of such a Connector Store and the other essential components with these principles to provide solutions for data interoperability, data sovereignty, and service discovery problems.

Table 1 Contribution of Related Chapters and (Sub)sections' in Answering Sub Questions

No.	Sub Questions	Chapters & (Sub)sections	Publications
1	What is state-of-the-art data-sharing in the logistics industry?	Chapter 3 - Section 3.4	- N/A
2	What is a suitable architecture to establish an IDS-compliant data-sharing ecosystem?	Chapter 4 - Section 4.3	- (Firdausy et al., 2022c)
3	How can the underlying application components of an IDS-compliant data-sharing ecosystem be designed to manage data interoperability and data sovereignty at runtime?	Chapter 5 - Section 5.1 - Section 5.2	- N/A
4	How can a Connector Store be designed to support companies to discover and select the underlying application components?	Chapter 6 - Section 6.2	- (Firdausy et al., 2022b) - (Firdausy et al., 2022a)
5	To what extent the IDS-compliant data-sharing ecosystem architecture can support logistics companies to manage data interoperability and data sovereignty at runtime?	Chapter 7 - Section 7.1 - Section 7.2 - Section 7.3 Chapter 8 - Section 8.1 - Section 8.3 - Section 8.4	- (Firdausy et al., 2022a)

Chapter 5 presents the requirements elicitation and software architectures of the relevant application components for managing data interoperability and data sovereignty to guide the development of the underlying components. **Chapter 6** investigates the suitable design of a Connector Store that supports the discoverability of the IDS Connectors, data sources, and active participants in a logistics data space. **Chapter 7** demonstrates the instantiation of the designs of the IDS Connector, IDS Data Apps, and Connector Store presented in the previous chapters into working prototypes to comprise a logistics data space demonstrator. **Chapter 8** discusses the contribution of the logistics data space demonstrator to achieving stakeholders' goals of managing data interoperability and sovereignty by means of single-case experiments and expert opinion. **Chapter 9** concludes this EngD thesis by summarizing the main results and findings, discussing the limitations and implications, explaining the significance for theory and practice, and giving pointers to future work. Based on this structure, **Table 1** is presented to provide a mapping of how each sub-questions previously listed in **Section 1.4** are addressed by this EngD thesis. In the following chapter, we will start addressing the problem framed by this thesis with the **Problem Investigation** phase.

PART 1
PROBLEM INVESTIGATION

2 Stakeholder Analysis

As part of the **Problem Investigation** within the adopted DSRM approach, the first step of this project is to identify the involved stakeholders and their goals. In this chapter, these stakeholders will be listed. **Section 2.1** describes the overview and the category of the stakeholders with respect to their contribution to the project. Whereas the following **Section 2.2**, **Section 2.3**, **Section 2.4**, and **Section 2.5** will elaborate on the details of each of the stakeholder categories and their influence on the system under design.

2.1. Stakeholder Identification

This design and development of the essential components for logistics data space project are being carried out under collaboration among consortium members from different backgrounds. The context diagram in **Figure 3** illustrates how these stakeholders are categorized into four main groups comprising (1) Academic Supervisors, (2) Design Adopter, (3) Knowledge Facilitator, and (4) End Users (Bonnema et al., 2016).

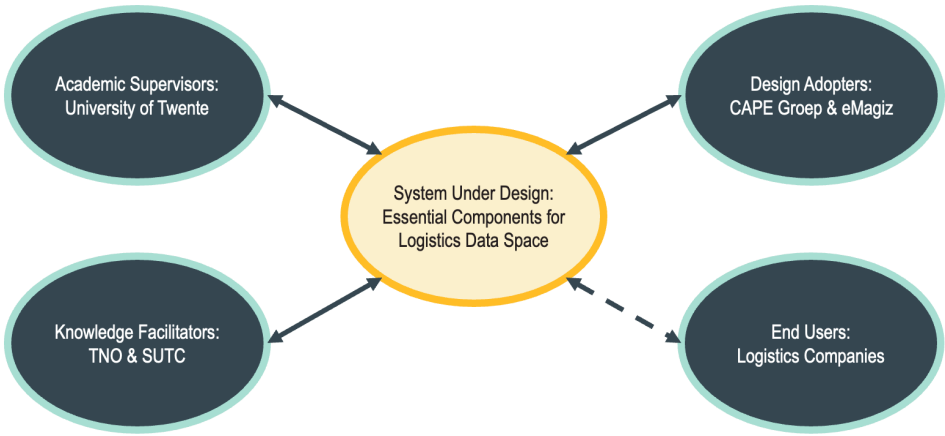


Figure 3 Stakeholder Context Diagram of the Essential Components for Logistics Data Space

The Academic Supervision category refers to the University of Twente (UT), which is responsible for supervising the two EngD candidates of this design project and coordinating the valorization of results from the academia to the industry. The Design Adopter role is assumed by CAPE Groep, which aims to incorporate the outcome of this project for its commercial iPaaS, the eMagiz platform, and make the results available for its customers and developers' community. The Knowledge Facilitator, which covers TNO and SUTC, enriches this project with relevant knowledge (i.e., expertise, frameworks, standards, etc.). Meanwhile, the End Users point to the logistics companies who are the logistics data space's to-be participants

utilizing a Connector Store and other essential components. In the following sections, the details of the individuals from each of these categories, and their influence on the system under design will be discussed.

2.2. University of Twente

From the side of the UT, two research groups are involved. The first research group is the Industrial Engineering and Business Information Systems (IEBIS) under the Faculty of Behavioural, Management and Social Sciences (BMS), and (2) the research group Services and Cybersecurity (SCS) under the Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS). These two research groups have a long history of conducting joint research. In this project, since the deliverables are designed for the logistics sector, the knowledge from IEBIS of business information systems in industrial engineering complements the knowledge of data interoperability and system integration from SCS. The individuals from the UT are mainly comprised of the thesis and daily supervisors, which their positions and goals are detailed in **Table 2**.

Table 2 Stakeholders Identification from the University of Twente

Stakeholder	Position	Involvement	Goals
Prof. Dr. Maria-Eugenia Iacob	Full Professor at IEBIS-BMS	Thesis Supervisor	- Instantiate the state-of-the-art data-sharing ecosystem (IDS) with the Connector Store prototype and operationalize it to the Dutch Logistics Sector. - Guide the EngD candidate throughout the project and ensure that the final thesis adheres to the quality standards. - Make sure that the candidate's Training and Supervision Plan (T&SP) aligns with the project and is achieved on time.
Dr. Ir. Marten J. van Sinderen	Associate Professor at SCS-EEMCS	Thesis Co-Supervisor & Project Owner	- Instantiate the state-of-the-art data-sharing ecosystem (IDS) with the Connector Store prototype and operationalize it to the Dutch Logistics Sector. - Guide the EngD candidate throughout the project and ensure that the final thesis

			adheres to the quality standards. - Organize research efforts and output among consortium members. - Make sure that the outcome of this design project is applicable to and can be benefitted by the industry.

2.3. CAPE Groep and eMagiz

CAPE Groep[5] is an IT consulting and business process digitization company with many customers in the transport logistics, supply chain, construction, and agri-food sectors. As a company that specializes in system integration, they have developed an iPaaS technology called the eMagiz platform, which is now managed by their subsidiary company eMagiz Services B.V.[6]. In this project, CAPE and eMagiz intend to participate in the development of the logistics data space demonstrator by providing the eMagiz platform to support data transformation and standards adoption. Additionally, part of their interest is also to investigate the additional values that the adoption of the IDS vision demonstrated in this project can bring to both the company and its customers, exploring a new form of business model in the process. In **Table 3**, the individuals from eMagiz are listed, which are mainly comprised of the company supervisor and the expert service of their platform.

Table 3 Stakeholders Identification from CAPE Groep and eMagiz

Stakeholder	**Position**	**Involvement**	**Goals**
Samet Kaya	Software Delivery Manager at eMagiz	Company Supervisor	- Develop and offer message converters based on eMagiz's portfolios with their clients to be reusable templates. - Promote the use of the OTM API Gateway based on the eMagiz platform to facilitate standards adoption and data transformation from and to OTM.

[5] https://capegroep.nl/over-cape/
[6] https://emagiz.com/en/ontstaan/

Erik Bakker	Expert Services at eMagiz	eMagiz Support	- Support the use of the eMagiz platform to facilitate data transformation and standards adoption for this project.

2.4. TNO and SUTC

As part of the Knowledge Facilitator category, the first contributor is TNO[7]. TNO stands for the Netherlands Organization for Applied Scientific Research (free translation). Within the context of the IDS, TNO is a member of the IDSA and plays its role as the "IDS Regional Hub" in the Netherlands due to the fact that they have co-authored the IDS RAM. Due to this, they support a project called the DASLOGIS that strives to leverage the Dutch Logistics Data Space (DLDS) into federated data space (Bastiaansen et al., 2020). In this project, TNO contributes to sharing its experiences in developing and implementing IDS Connectors and other IDS components for the logistics industry in the Netherlands.

There is also SUTC[8], which refers to the Uniform Transport Code Foundation (free translation). SUTC helps logistics companies to share data with partners securely and efficiently by managing and promoting the use of standards, with OTM being one of them, that are developed for and by the logistics sector. On behalf of branch organizations TLN and Evofenedex, SUTC provides its expertise on problems related to data sharing and ICT standards adoption currently experienced in the Dutch logistics industry. They also pose an interest in the valorization of the OTM standards to be used in the logistics data space demonstrator, enhancing the adoption of standards in the process. Thereupon, in **Table 4**, the contributing individuals from these two organizations are then listed.

Table 4 Stakeholders Identification from TNO and SUTC

Stakeholder	**Position**	**Involvement**	**Goals**
Dr. Ir. H.J.M. (Harrie) Bastiaansen	Senior Business Consultant at TNO	IDS Regional Knowledge Facilitator	- Investigate and valorize innovations in IT and data-sharing infrastructures, such as IDS, for the logistics industry.
Wout van den Heuvel	General Secretary at SUTC	Logistics Standards Knowledge Facilitator	- Investigate and valorize the adoption of standards, such as OTM, to be used in logistics data-sharing environments.

[7] https://www.tno.nl/en/
[8] https://www.sutc.nl/en_US/

3 Data Sharing in the Logistics Industry

Next to investigating the stakeholders, the **Problem Investigation** phase includes an SLR to explore the current situation of data-sharing in the logistics industry. First of all, **Section 3.1** establishes the methodology that this thesis adopts to guide the SLR. Next, **Section 3.2** elaborates on the planning of the SLR by defining the research questions and other related activities. **Section 3.3** covers the execution process by applying the inclusion and exclusion criteria and selecting the papers to be reviewed. **Section 3.4** concludes this chapter by analyzing and providing an overview of the knowledge extracted from the selected papers.

3.1. SLR Methodology

To perform the SLR, this thesis follows the methodology used by Rouhani et al. (2015), who adopted the guidelines from Kitchenham and Charters (2007) to perform SLR in software engineering. The guideline constitutes three successive stages, starting with the Planning, Execution, and Result Analysis. These stages define the structure of the remaining sections on how the SLR will be conducted. Activities related to each stage are listed in **Table 5**.

Table 5 Overview of SLR Activities

Planning	
1	Define knowledge questions
2	Select scientific databases
3	Formulate search query
4	Define inclusion and exclusion criteria
Execution	
5	Execute the formulated query for each scientific database
6	Apply inclusion criteria to select articles
7	Remove duplicate articles across scientific databases
8	Apply exclusion criteria to remove irrelevant articles
Result Analysis	
9	Extract data based on the research questions
10	Synthesize and conclude results to answer research questions

3.2. Planning

This section outlines how this SLR is designed as listed in **Table 5**. The first step is to define the research question to be answered. The second step is to select scientific databases. Thirdly is to formulate the search queries, which will be followed by defining the criteria used to include and exclude search results.

3.2.1. SLR Research Question

Sharing real-time data in the logistics sector has been a common practice for logistics operators and their customers to support operational and other innovational purposes (Pham et al., 2019). Multiple benefits have been perceived from it. Though, establishing such an ecosystem for sharing data also comes with some challenges. For that reason, this SLR is set in motion to (1) identify the challenges that hamper logistics companies to establish a data-sharing ecosystem and (2) explore the state-of-the-art data-sharing ecosystem to tackle them. By doing so, we aim to answer **SQ 1** defined in **Section 1.4** earlier and we set it to be the main research question for this SLR. Therefore, this SLR formulates the research questions as follows.

SLR Main Question:

What is state-of-the-art data-sharing in the logistics industry?

SLR Sub-Questions:

SLR SQ 1. What are the challenges that hamper companies in the logistics industry to establish a data-sharing ecosystem?

SLR SQ 2. What are the latest data-sharing ecosystems available in the literature to tackle data interoperability and data sovereignty in the logistics industry?

3.2.2. Scientific Databases

To provide good coverage of the academic literature on this topic, this SLR selects two scientific databases. These databases comprise Scopus[9] and Web of Science[10] (WoS). The reason for this is that these two belong to the top 5 most trusted academic resource databases, with the 3rd database containing intersecting results with Scopus, and the 4th and 5th databases being specialized in research fields irrelevant to the topic of this thesis. In addition, in the later stage, we also plan to broaden the search scope by scanning the selected articles' references to uncovering the obscured knowledge.

[9] www.scopus.com
[10] www.webofscience.com

3.2.3. Search Queries

This SLR formulates the search queries based on the keywords relevant to the main question of this thesis and the sub-questions of this literature study. These keywords are primarily focusing on the terms related to "Logistics", "Data Sharing", "Data Interoperability" and "Data Sovereignty". From these terms, **Table 6** lists the relevant keywords, along with synonymous (row) and contextual (column) keywords to cover relevant results. We group the contextual keywords into **Industry Context**, **Problem Context**, and **Requirements**. To standardize the terms within the industry, several keywords related to the **Industry Context** are taken from the GS1 Logistics Interoperability Model Application Standard document[11]. The **Problem Context** groups keywords related to "Data Sharing". Whereas the **Requirements** group involves keywords related to the "Data Interoperability" and "Data Sovereignty" aspects.

Table 6 SLR Synonymous Search Keywords

Industry Context	**Problem Context**	**Requirements**
Logistics	Data-Sharing	Interoperable
Logistics Sector	Data Sharing	Interoperability
Logistics Industry	Data-Exchange	Integration
Transport Logistics	Data Exchange	Interorganization
Freight Forwarding	Data-Sharing Ecosystem	Interorganizational
Warehousing	Data Sharing Ecosystem	Sovereign
Supply Chain	Data-Sharing Environment	Sovereignty
Logistics Service Provider	Data Sharing Environment	
Logistics Service Client	Data Sharing Architecture	
Freight Forwarder		
Transport Service Provider		
Warehouse Service Provider		
Distribution Center		

Next, we construct the search queries for the scientific databases. The search queries are assembled by combining the synonymous keywords with the "OR" logical operator, and contextual keywords with "AND". The resulting search queries for

[11] https://www.nweurope.eu/media/14879/gs1_logistics_interoperability_model_application_standard.pdf

the databases to be applied to the article's title, abstract, and keywords are formulated as follow.

Scopus Search Query:

TITLE-ABS-KEY (

("Logistics" OR "Logistics Sector" OR "Logistics Industry" OR "Transport Logistics" OR "Freight Forwarding" OR "Warehousing" OR "Supply Chain" OR "Logistics Service Provider" OR "Logistics Service Client" OR "Freight Forwarder" OR "Transport Service Provider" OR "Warehouse Service Provider" OR "Distribution Center")

AND

("Data-Sharing" OR "Data Sharing" OR "Data-Exchange" OR "Data Exchange" OR "Data-Sharing Ecosystem" OR "Data Sharing Ecosystem" OR "Data-Sharing Environment" OR "Data Sharing Environment" OR "Data Sharing Architecture")

AND

("Interoperable" OR "Interoperability" OR "Integration" OR "Interorganization" OR "Interorganizational" OR "Sovereign" OR "Sovereignty")

)

WoS Search Query:

TS=(("Logistics" OR "Logistics Sector" OR "Logistics Industry" OR "Transport Logistics" OR "Freight Forwarding" OR "Warehousing" OR "Supply Chain" OR "Logistics Service Provider" OR "Logistics Service Client" OR "Freight Forwarder" OR "Transport Service Provider" OR "Warehouse Service Provider" OR "Distribution Center") AND ("Data-Sharing" OR "Data Sharing" OR "Data-Exchange" OR "Data Exchange" OR "Data-Sharing Ecosystem" OR "Data Sharing Ecosystem" OR "Data-Sharing Environment" OR "Data Sharing Environment" OR "Data Sharing Architecture") AND ("Interoperable" OR "Interoperability" OR "Integration" OR "Interorganization" OR "Interorganizational" OR "Sovereign" OR "Sovereignty"))

OR

TI=(("Logistics" OR "Logistics Sector" OR "Logistics Industry" OR "Transport Logistics" OR "Freight Forwarding" OR "Warehousing" OR "Supply Chain" OR "Logistics Service Provider" OR "Logistics Service Client" OR "Freight Forwarder" OR "Transport Service Provider" OR "Warehouse Service Provider" OR "Distribution Center") AND ("Data-Sharing" OR "Data Sharing" OR "Data-Exchange" OR "Data Exchange" OR "Data-Sharing Ecosystem" OR "Data Sharing

Ecosystem" OR "Data-Sharing Environment" OR "Data Sharing Environment" OR "Data Sharing Architecture") AND ("Interoperable" OR "Interoperability" OR "Integration" OR "Interorganization" OR "Interorganizational" OR "Sovereign" OR "Sovereignty"))

OR

AB=(("Logistics" OR "Logistics Sector" OR "Logistics Industry" OR "Transport Logistics" OR "Freight Forwarding" OR "Warehousing" OR "Supply Chain" OR "Logistics Service Provider" OR "Logistics Service Client" OR "Freight Forwarder" OR "Transport Service Provider" OR "Warehouse Service Provider" OR "Distribution Center") AND ("Data-Sharing" OR "Data Sharing" OR "Data-Exchange" OR "Data Exchange" OR "Data-Sharing Ecosystem" OR "Data Sharing Ecosystem" OR "Data-Sharing Environment" OR "Data Sharing Environment" OR "Data Sharing Architecture") AND ("Interoperable" OR "Interoperability" OR "Integration" OR "Interorganization" OR "Interorganizational" OR "Sovereign" OR "Sovereignty"))

3.2.4. Inclusion and Exclusion Criteria

The next step is to define several criteria to further direct the search results' focus toward the primary topic of this study. This SLR defines and lists the inclusion (IC) and exclusion criteria (EC) in **Table 7**. Two main points need to be highlighted here. First, as shown in IC1, this study focuses on research articles that have been published in the last 10 years to ensure the state-of-the-art aspect of the knowledge. Additionally, this study selects articles that are focusing on the subject areas listed under IC3 due to their relevance to the logistics industry (e.g., Energy and Environmental Science investigating sustainability in logistics through supply chain partnership). Secondly, after removing duplicate articles from the two scientific databases, this study also excludes articles that do not directly contribute to answering the SLR sub-questions. This exclusion is done by assessing their titles, abstracts, and contents, which will be discussed in the next section.

Table 7 SLR Inclusion and Exclusion Criteria

ID	Inclusion Criteria	ID	Exclusion Criteria
IC1	English-based articles from the past 10 years (i.e., 2013-2023)	EC1	Duplicate articles based on their titles or contents.
IC2	Studies published in Conferences Proceedings, Journal Articles, and Book Chapters.	EC2	Studies not related to the main RQ are based on their titles, abstracts, and content.
IC3	Studies focusing on the subject areas of: Computer Science	EC3	Articles with incomplete or unavailable full texts.

	Engineering Decision Sciences Mathematics Business, Management, and Accounting Social Sciences Environmental Science Energy Economics, Econometrics, and Finance Multidisciplinary		

3.3. Execution

The next phase of the SLR is to execute the review based on the plan above. This phase comprises multiple steps. The first is to apply the formulated search queries along with the inclusion criteria to the academic databases. From executing the search queries to Scopus and WoS prior to applying the inclusion criteria, we have found 315 articles and 143 research articles respectively. Second, after we apply the inclusion criteria defined in **Table 7** as filters, the results narrow down to 161 and 98 articles from the respective databases as shown in **Figure 4**.

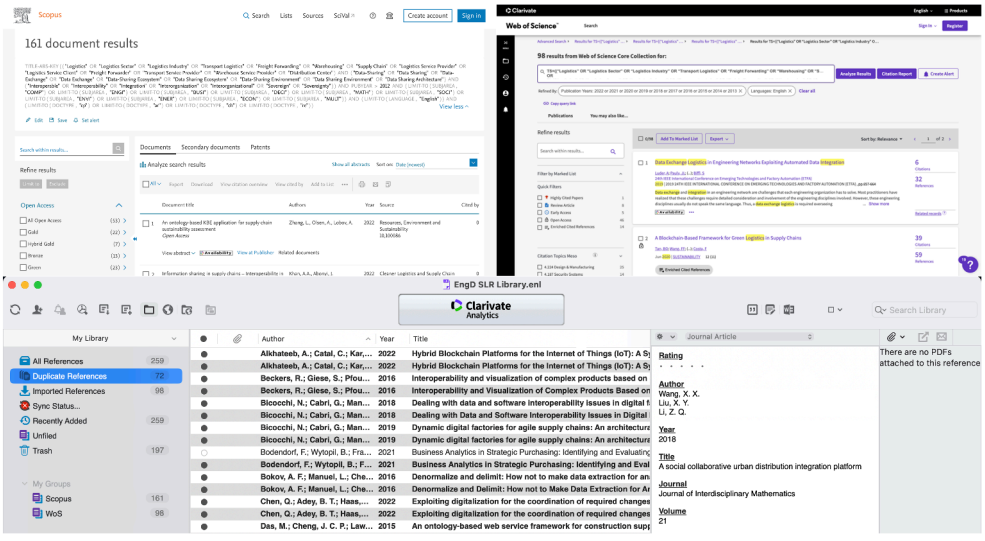


Figure 4 SLR Scopus and WoS Query Results

In the third step, we export the metadata of these filtered articles to EndNote[12] to further remove duplicates and exclude irrelevant articles based on their title and abstract. Using a feature provided by EndNote in combination with a manual check to remove missed duplicates, in this step, we have removed 72 entries from the 259 combined articles, resulting in 187 articles for reviewing. In the fourth step, we selected 66 relevant articles, which their titles and abstracts are aligned with the context of data-sharing and the requirements of data interoperability and data sovereignty. From these 66 entries, we investigated their full-text availability and at the end of this fifth step, we have only found 61 articles, which the full-paper of these articles are available online. Finally, we assess these full texts and select the articles that provide contributions toward answering the SLR questions. At the end of this process, we selected 35 articles, and we capture this overall procedure in **Figure 5**.

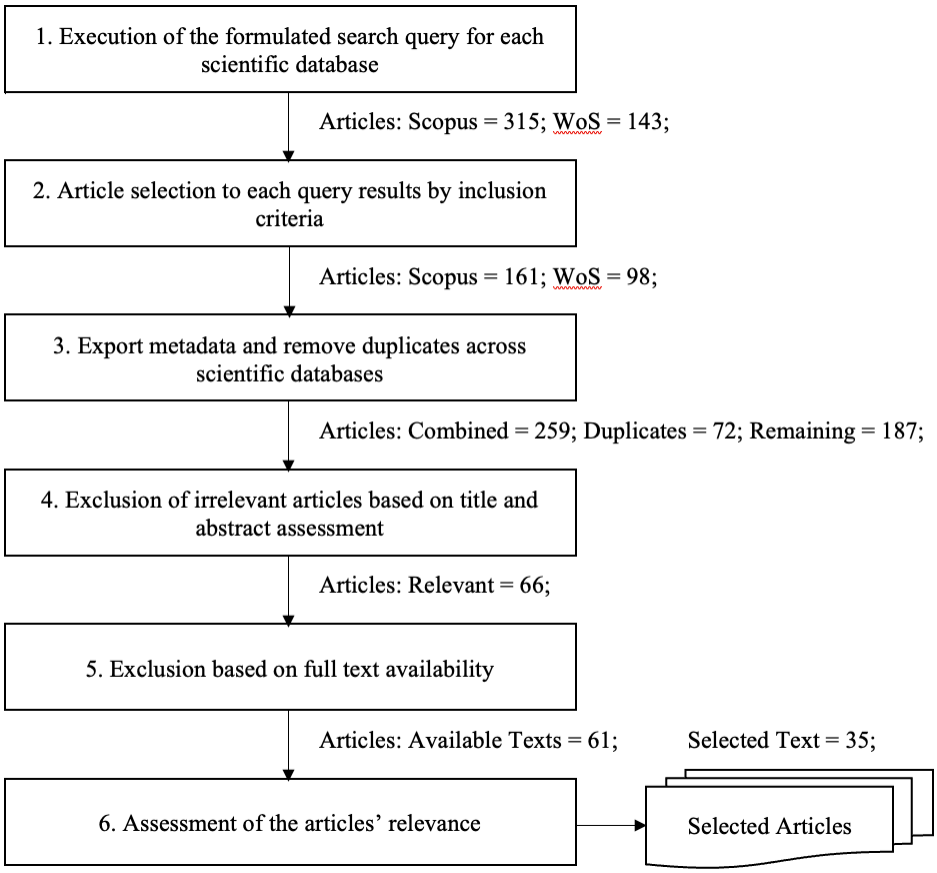


Figure 5 SLR Article Selection Flowchart

[12] https://endnote.com/

3.4. Result Analysis

Following the selection of the articles, we collected relevant information that is essential to address the SLR sub-questions. This information will contribute to capturing the current challenges perceived by the logistics industry to establish a data-sharing ecosystem and explore the latest approaches available in the literature to tackle these hurdles. **Appendix B** presents the selected 35 articles along with their research goals and contributions to addressing the sub-questions. To provide a preliminary overview, **Table 8** presents the assessment of their contribution by listing the challenges and treatments. From the literature, three major groups of challenges are identified, namely Data Interoperability (DI), Data Sovereignty (DS), and Centralized vs Decentralized Ecosystem (CDE). Several treatments are also extracted, revolving around the Standards Adoption (SA), Schema Mapping (SM), Integration Hub (IH), Semantic Web (SW) and Linked Data technologies, Data Space Ecosystem (DSE), and Blockchain Technology (BT). To investigate the CDE dilemma, we mark any articles that provide a decision to adopt either the centralized ecosystem (CE) or decentralized (DE). Lastly, we extract any architectural designs (AD) available that visualize the treatments they proposed.

Table 8 SLR Article Contribution Assessment Form

No	References	Challenges			Treatments								
		DI	DS	CDE	SA	SM	IH	SW	DSE	CE	DE	BT	AD
P1	(Ferreira et al., 2012)	✓	✓		✓	✓	✓			✓			✓
P2	(Bhatt & Zhang, 2013)		✓	✓	✓		✓		✓				
P3	(Främling et al., 2013)	✓			✓	✓	✓						
P4	(Abecker et al., 2014)	✓							✓	✓			✓
P5	(Das et al., 2015)	✓		✓		✓		✓			✓		✓
P6	(Gnimpieba Zanfack et al., 2015)	✓	✓				✓				✓		✓
P7	(Zhao & Liang, 2015)	✓			✓		✓						✓
P8	(Andreeva et al., 2016)	✓			✓			✓					
P9	(Campos et al., 2016)												✓
P10	(Hofman, 2016)					✓	✓						
P11	(Schöggl et al., 2016)		✓										
P12	(Tran et al., 2016)		✓										
P13	(Scholz et al., 2018)		✓			✓			✓				

P14	(Verhoosel et al., 2018)	✓						✓					✓
P15	(Wang, X. X. et al., 2018)								✓				✓
P16	(Abebe et al., 2019)		✓	✓							✓	✓	✓
P17	(Bicocchi et al., 2019)	✓		✓		✓		✓	✓		✓		✓
P18	(Dalmolen, Bastiaansen, Kollenstart, et al., 2019)		✓	✓			✓		✓		✓		✓
P19	(Debicki & Kolinski, 2019)	✓	✓			✓	✓						
P20	(Hofman, W., 2019)	✓						✓	✓				
P21	(Hofman, W. J., 2019)		✓						✓		✓	✓	
P22	(Wang et al., 2019)		✓								✓	✓	✓
P23	(Abu-elezz et al., 2020)											✓	
P24	(Bastiaansen et al., 2020)		✓	✓					✓		✓		✓
P25	(Carvalho et al., 2020)	✓		✓		✓	✓		✓		✓		✓
P26	(Piest, Iacob, et al., 2020)	✓	✓		✓				✓		✓		✓
P27	(Tan et al., 2020)		✓								✓	✓	✓
P28	(Voswinckel et al., 2020)	✓	✓	✓			✓		✓		✓	✓	✓
P29	(Cirullies & Schwede, 2021)	✓	✓					✓	✓		✓		✓
P30	(Karatas & Gultekin, 2021)	✓	✓		✓								
P31	(Bouter et al., 2022)	✓			✓			✓					
P32	(Frey et al., 2022)		✓						✓		✓		✓
P33	(Heinbach et al., 2022)		✓									✓	✓
P34	(Kazantsev et al., 2022)	✓	✓										
P35	(Top et al., 2022)	✓	✓					✓					
Total		20	21	8	8	8	10	8	14	2	14	7	20

From this preliminary assessment, several trends can be observed. Articles that discuss issues related to data interoperability, revolving around syntactic and semantic interoperability as well as hurdles in the adoption of standards, show a

relatively stable trend throughout the years. Interestingly, data sovereignty, which covers data security and confidentiality, starts to gain more attention from the year 2018. This indicates that data sovereignty has become a critical requirement for data-sharing ecosystems by the time this literature review is conducted. To gain more understanding of these findings, the detailed information extracted from these articles with regard to the challenges and the proposed treatments will be elaborated in the following **Sub-Section 3.4.1** and **Sub-Section 3.4.2**. Finally, this SLR will be concluded in **Sub-Section 3.4.3**.

3.4.1. Challenges of Establishing Data-Sharing Ecosystems for the Logistics Industry

Syntactic Interoperability, Semantic Interoperability, and Standardizations

Most of the selected studies have discussed the hurdles of setting up an ecosystem for companies and organizations to share data. Of these articles, Ferreira et al. (2012) noted that establishing and maintaining a data-sharing network with seamless interoperability can be demanding due to the heterogenous requirements, policies, information systems (ISs), and data formats (e.g., unique identifiers, data model and schema, etc.). One simple example of this issue is the usage of proprietary and vendor-specific formats to identify individual items (e.g., products, containers, etc.), as opposed to using a standardized or globally recognized format in the supply chain (e.g., EDI-based SMDG standard for sea modality vs XML-based GS1 standard for road modality) (Främling et al., 2013; Hofman, W., 2019). Another example takes the case of two data models that describe a person's name in two different ways: the first one has a field with the full name of the person, and the second one has the same name divided into first and last name. Such conflicts in the way data are represented and exchanged hamper, what is known as, the syntactic interoperability between the interacting systems. Another known issue related to this is semantic mismatches, which can occur from a simple declaration of an attribute of the same object. An example of this is the "selling price", which is being associated with different meanings and formulas for different stakeholders (e.g., manufacturer and distributor) in a supply chain (Andreeva et al., 2016). To illustrate them, the following **Table 9** lists how Das et al. (2015) have categorized the conflicts caused by the data heterogeneity.

Table 9 Categories of Data Heterogeneity Conflicts

Conflicts	**Description**	**Examples**	
		Entity 1	**Entity 2**
Naming Conflicts	Semantically similar but entities and attributes are represented with different names.	contractor (material_name)	supplier (item_name)

Data Representation Conflicts	Semantically similar but entities and attributes are represented with different data types, data structures, or measurement units.	material (ID, price) *price in USD	material (ID, price) *price in EUR
Aggregation Conflicts	Semantically similar entities, but one may be represented as the aggregate of the other.	getMaterial (material_name, company)	getMaterial (name_of_glass, company)
Context Conflicts	Semantically similar but different output requirements.	createOrder (ID, price) *ID here is the ID of the purchase order	createOrder (ID, product, name, price) *ID here is the ID of the product

One approach to solve such conflicts is through the standardization approach. Some standards have been proposed and adopted by the industry. These include, among others, the Universal Business Language (UBL)[13] to document procurement and other transportation transactions, commerce eXtensible Markup Language (cXML)[14] to communicate procurement documents, and e-business funStep Open Architecture (ebfSOA)[15] to integrate product transaction and management (Ferreira et al., 2012). Additionally, there exist, GS1 Electronic Product Code Information System (EPCIS)[16] to describe and share information on products, shipments, and events, UN/CEFACT[17] to describe products and services, OTM to describe real-time logistic trip data, etc. (Bouter et al., 2022; Piest, Iacob, et al., 2020). Most of these standards are represented in the XML format due to the possibility of defining and validating its format using the XSD. Though, the use of the JSON format has also been demonstrated in the case of the OTM standard.

Despite its advantages and promised benefits, the adoption of standards by organizations is reported to be suffering from several impediments. Främling et al. (2013) discussed that GS1 EPCIS can be too complex or costly to implement compared to the promised benefits. This is confirmed by other studies that implementing standards can impose a big investment burden for SMEs due to the required technical solutions for bridging the information model adopted by each of these companies with the global standard agreed by the network (Cirullies & Schwede, 2021; Ferreira et al., 2012; Kazantsev et al., 2022; Piest, Iacob, et al., 2020;

[13] https://www.oasis-open.org/
[14] https://cxml.org/
[15] http://www.funstep.org/
[16] https://www.gs1.org/standards/epcis
[17] https://unece.org/trade/uncefact

Scholz et al., 2018). Furthermore, Andreeva et al. (2016) and Debicki and Kolinski (2019) also argued that even though there exists a common transmission protocol and API message format like JSON, the effort to reconcile and foster semantic interoperability between these standards can be cumbersome. The reason for this is argued to be due to their incompatible semantic model. Researchers attempted to solve this problem in recent developments by introducing ontological models to the industry. More on this will be discussed in **Sub-Section 3.4.2**.

Data Sharing vs Data Confidentiality and Data Sovereignty

Despite the call for a solution to treat the issues of data interoperability, concerns related to data confidentiality and sovereignty have also gained traction when companies must work together and share information. Ferreira et al. (2012) mentioned that the concern over data confidentiality creates corporate resistance to engage in whole-chain collaboration more actively as some members interacting within the network may be each other's competitors on other services. Gnimpieba Zanfack et al. (2015) stated that the next question asked by companies, after the issues of data interoperability, is how supply chain partners will share data with all their partners, considering the confidentiality, security, access right, and service level agreement (SLA) of the shared data. If not properly governed, data-sharing can lead to the loss of protection over their formulation information and adverse effects on their competitive advantages. Such a loss of control over data justifies why some small companies still use paper-based systems for their tracking and tracing solutions (Bhatt & Zhang, 2013). Accordingly, Tran et al. (2016) made a point that the risks associated with inter-organizational information sharing escalate as the volume of exchanged information increases and the mechanisms in their information systems to protect the shared data decrease. Thus, companies' willingness to share sensitive data with trading partners will stay at the bare minimum unless they have sufficient confidence in the perceived security of supply chain information systems.

Among the selected articles, the term data sovereignty has not been discussed until 2019 when Dalmolen, Bastiaansen, Kollenstart, et al. (2019) introduced and defined it as the capability to be in control over the usage conditions of the shared data. They argued that the existing data-sharing architectures mainly lack the required data sovereignty capabilities. This is because security functionalities, which are mostly applied through encrypted data transactions and augmented with user authentication and authorization, are not enough to achieve data sovereignty. Together with the dynamics in today's business networks, they noted that trust and data sovereignty must be embedded by design in the digital data-sharing infrastructure itself and can no longer be based on long-term inter-organizational relationships. Without such a technically enforced trust in the supply chain, partners risk being unwilling to fully share and interact with information, impeding the efficiency of the entire supply chain as a result (Wang et al., 2019). Bastiaansen et al. (2020) highlighted that such trust and data sovereignty must be accompanied by additional security capabilities, which include (1) encrypted data transport and storage, (2) software certification and attestation, and (3) data-sharing agreements

and usage contracts. Although, such technical enforcement to manage data sovereignty also poses several challenges. One of them is the risk of vendor lock-in by their IT providers, which leads to another major integration effort to define, manage, and enforce data sovereignty and security solutions across multiple data-sharing relationships (Bastiaansen et al., 2020).

Centralized vs Decentralized Data-Sharing Ecosystem

Another concern that gained attention in the literature is the decision to adopt a decentralized approach against a centralized one when establishing a data-sharing ecosystem. Abecker et al. (2014) proposed a communication architecture based on the SOA paradigm, in which they highlighted a data warehouse that acts as a central data hub between multiple software systems in their design. Their aim was to have a data-exchange platform in which different actors in the supply chain can share data based on widespread standards and semantic technologies. This design decision seems rational due to the fact that their industrial applications were intended for supply chains in the water supply and water utility sector, where competition is relatively low, and thus, trust is established through mutual goals.

On the contrary, the rest of the literature expressed their take on such a design pattern differently. Bhatt and Zhang (2013) argued that a centralized database system approach tends to have serious scalability and performance issues when there is a need for processing large data sets that originate from complex systems. Das et al. (2015) also noted that a centralized collaboration system can be inefficient and impractical, especially for the construction industry with its multi-party nature, as the participants often hesitate to share and store their data in a third-party central database. Consequently, having a network with a centralized data repository can lead to the risk of deteriorating the underlying trust, which decentralized networks might provide (Abebe et al., 2019). Thus, a framework that allows network participants to exchange data securely in a distributed multi-party environment is called upon.

3.4.2. State-Of-The-Art Approaches for Tackling Data Interoperability and Sovereignty

From Data-Sharing Challenges towards Interoperable, Secure, and Sovereign Dataspaces

Besides discussing the challenges of establishing data-sharing ecosystems in their respective contexts, several solutions to treat them were also proposed by the selected articles. To facilitate the quest for an improved network's collaboration and interoperability, Bhatt and Zhang (2013) called for a design of an interoperability framework that can accommodate a wide variety of platforms, technologies, standards, and business practices. Such a framework has to be inclusive for anyone to follow, as well as, able to support the participants to protect data and maintain privacy while sharing data with partners and regulatory agencies. The SOA

paradigm is preferable for such a framework to leverage the interoperability of inter-organizational enterprise systems through web-service orchestrations (Abecker et al., 2014). Scholz et al. (2018) noted that to establish a data-sharing platform that is attractive for users to collaborate and integrate into, it should be based on a bottom-up approach of bringing together already existing solutions. This implies that several integration techniques, as well as different types of data contents, need to be supported. This brings to the discussion of whether a rigid specification of data standards is a preferred approach to foster network adoptions, as opposed to a flexible one that allows data format transformation on a case-by-case basis.

With respect to the implementation of standards to enhance interoperability between network participants, Ferreira et al. (2012) noted that network participants still need data mapping and transformation mechanisms to bridge the data model of their own systems to the one used in the network. For this purpose, they introduced the concept of model morphism, which describes the relations (e.g., mapping, merging, transformation, etc.) between the specifications of two or more ISs. Mappings can be defined on a P2P basis, but by adopting a standard or CDM, companies only need to consider the mapping from their internal model to the reference one and vice versa. However, the adoption of a standard, or even a CDM, can be a challenge for some SMEs when establishing a data-sharing ecosystem. Ferreira et al. (2012) and Främling et al. (2013) suggested the involvement of software and data integrator companies as the intermediary hub and mapping knowledge base to alleviate such a burden (e.g., by offering a standard compliant ready product, or, a platform that provides message routing and mapping capabilities). For this purpose, Gnimpieba Zanfack et al. (2015) and Hofman (2016) suggested the adoption of an ESB that provides protocol adapters for transforming the protocol (or format) of a message sent by a client to another one. Hofman, W. (2019) supports this by suggesting partnerships with commercial Integration Service Providers (ISPs) as intermediaries to hide the complexities imposed by the adoption of the amplitude of standards, especially for SMEs due to their business demands.

Alternatively, to foster inter-organizational interoperability, several studies suggested the adoption of ontologies and Semantic Web technologies. Das et al. (2015) proposed an ontology-based web service framework, which promotes the use of (1) an ontology as a common data model (created with Protégé[18]) for data mediation in the domain of application, and (2) Semantic Web technology (ontology query language, e.g., SPARQL[19]) for mapping and translating one data schema to another (i.e., XML to OWL and back). Zhao and Liang (2015) followed a similar path for the marine sector by establishing an ontology that describes terms (i.e., concepts, properties, relationships, and instances) in the sector and demonstrated information exchange based on XML between middleware and enterprises using data adapters. Andreeva et al. (2016) presented a higher-level ontology that oversees the concepts and relationships that exist within several ontologized logistics standards to facilitate the interoperability between them. Through the shared relationships of

[18] https://protege.stanford.edu/

[19] http://www.w3.org/TR/sparql11-query/

subclasses pointing to the same parent classes, they demonstrated the relationships and interoperability between standards. Verhoosel et al. (2018) proposed a platform that enables access to a variety of data sources via a linked data web-based mechanism, in which the same approach is followed by Bouter et al. (2022). In both of their work, they demonstrated the approach to interoperate several data sources (and data standards) via Linked Data and Semantic Web technologies, which involve representing the data sets in RDF[20]/OWL[21] format, storing them in a triple store (e.g., Apache Jena Fuseki[22]), and accessing them through SPARQL interface.

Subsequently, several other studies have contributed to the development of data-sharing platforms to lower the barrier of integrating and interoperating data(sets) from several sources. Bicocchi et al. (2019), for one, discussed several concepts of interoperability platforms to support agile and global multi-tier supply chains. One of them refers to a dataspace, which is understood as a digital environment that allows the coexistence and integration of heterogeneous data sources by providing basic functionalities (e.g., message mapping, transformation, etc.) over all data sources to reduce the effort required to exchange data in a *pay-as-you-go* fashion. Another related concept to this is the polystore systems, which pursue the idea of a flexible data sharing and interoperability architecture solution. This is achieved by enabling query processing over heterogeneous (data) stores while guaranteeing full-source autonomy, just-in-time transparent data transformation, and support for multiple query interfaces. They argued that by combining the properties of these two concepts, the integration of heterogeneous data in global multi-tier supply chains can be facilitated. On top of that, web services managing and providing the data should also be annotated with rich semantic descriptions that include keywords or synonyms (e.g., context, operations, and parameters) to support the discovery of these data sources.

Table 10 Support Processes for Data-Sharing Management (Dalmolen, Bastiaansen, Kollenstart, et al., 2019)

No.	Life-cycle Stage	Subprocesses
1	Defining and publishing a data set.	Definition of a data-sharing profile.
		Publication of a data-sharing profile.
2	Making a data-sharing agreement.	Definition of terms-of-use, including usage and access control policies.
		Definition of the commercial and juridical conditions.

[20] https://www.w3.org/RDF/
[21] https://www.w3.org/OWL/
[22] https://jena.apache.org/

		Negotiation, acceptance, and signing of the data-sharing agreement.
3	Performing a data-sharing transaction.	Clearing of the data-sharing transaction, including non-repudiation.
		Data transfer, including the binding of the transaction to an agreement.
		Settlement and discharging of the data-sharing transaction
4	Logging, provenance, and reporting.	Logging and binding of data transactions to data-sharing agreements.
		Tracking, monitoring, and reporting of data transactions to stakeholders.
		Auditing, billing, and conflict resolution.

One of the concerns of practitioners and organizations is the confidentiality and sovereignty aspects of the data exchange within and between data spaces. Dalmolen, Bastiaansen, Kollenstart, et al. (2019) noted that every data transaction should be complemented with metadata-enabled support processes that regulate its management and agreements around the exchange. Such support processes, as described in **Table 10**, cover the activities for data providers and consumers to comply with both internal and external policies on data-sharing. Several metadata artifacts should also be generated by these processes. One of them, besides the description of the to-be-shared data, is the specifications of the data-sharing agreements specifying the contractual terms of use. This metadata is associated with another metadata, namely (1) access control policy which individuals, roles, or systems are granted access to the shared data, and (2) usage control policy stating which individuals, roles, or systems are allowed certain operations on the data.

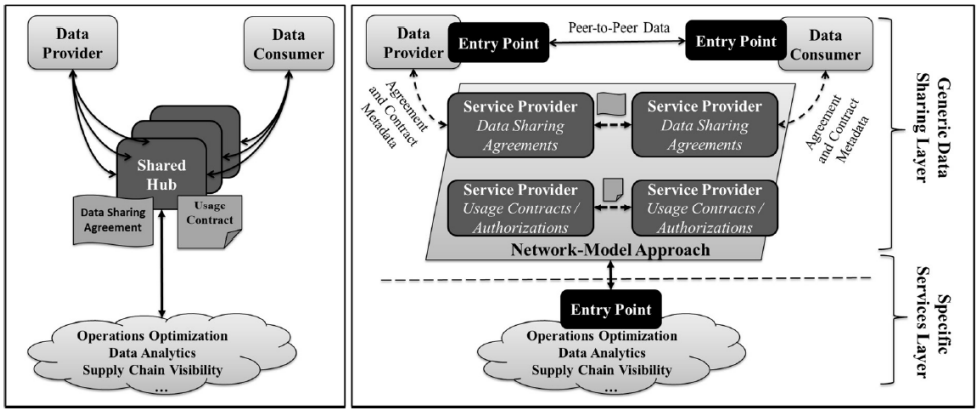


Figure 6 Data-Sharing Transition from Hub Model to Network Model (Bastiaansen et al., 2020)

As the metadata that regulates the data sovereignty aspect becomes more prominent, the management and the sovereignty of the metadata artifacts themselves gain more attention. The tasks to define and maintain these metadata of data transactions might stall data providers to focus on their core businesses. Additionally, these support processes' metadata artifacts may reveal data providers' sensitive information. As such, outsourcing these tasks to a central service provider threatens the sovereignty of the ecosystem participants. Therefore, Dalmolen, Bastiaansen, Kollenstart, et al. (2019) called for a transition from a solution-specific hub-model-based approach to an open network-model one for controlled data-sharing with a single-entry point. Such a switch, as illustrated in **Figure 6**, aims to improve end-user centricity in simultaneously managing multiple data-sharing interconnections (Bastiaansen et al., 2020). This approach also provides generic infrastructural data sovereignty capabilities, employing a single-entry point for data providers with common and agreed-upon protocols for defining and enforcing these data-sharing agreements. Such an approach lays the groundwork for an (open) network model for data-sharing and is currently being adopted as the underlying design principles employed by the IDS.

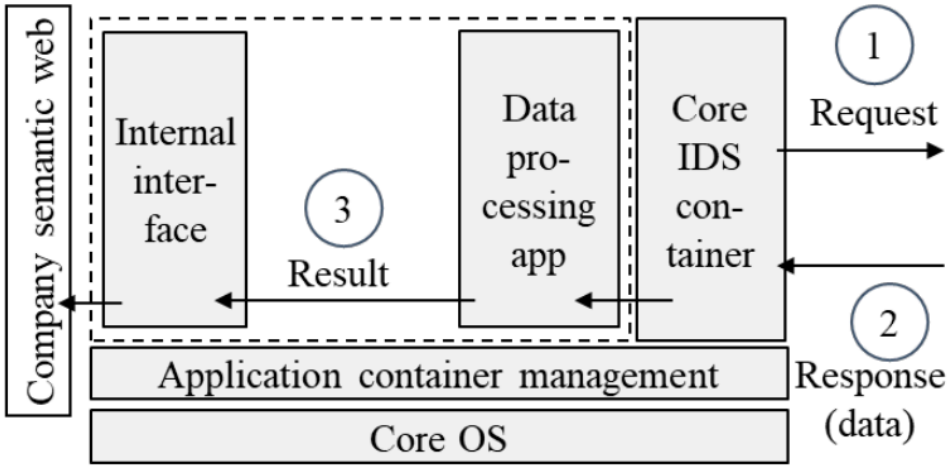


Figure 7 IDS Connector's Schematic of a Network Participant (Cirullies & Schwede, 2021)

Several other studies shared the same vision that is promoted by the IDS. Hofman, W. (2019) proposed the development of data-sharing in supply and logistics with key elements of the IDS considering that, first, the solution should be technology neutral and available to all organizations. Next to that, it should stimulate innovation and inclusion of SMEs through the connect-once-reach-all scheme. And more importantly, such an ecosystem for data-sharing should promote trust for members, facilitate data provenance, and control data access with agreed rules for data (re)use. Cirullies and Schwede (2021) connect to the IDS vision by prescribing data exchange between the data owner and user through a single-entry point called

the IDS Connector, which provides self-description metadata that describes accessibility information, prices, usage policies, etc.

As depicted in **Figure 7**, the architecture of this security gateway consists, among others, an application container management, a communication bus, and a configuration manager, which are required to manage and orchestrate container-based applications capable of performing data aggregation, anonymization, calculation, transformation, etc. Frey et al. (2022) also leverage the IDS architecture for their Bauhaus.MobilityLab platform with the IDS Connectors to isolate data exchange, describe data sources with formal metadata, and enforce data sources with usage policies. Cirullies and Schwede (2021) stated that IDS principles align well with their development of an on-demand shared digital twin for supply networks due to their requirements, stated below:

1. Digital twins for supply networks must be decentral since supply networks are decentral.
    a. The implementation of IDS Connector and its connection to local sources, retrievable from the data broker.
    b. The offering of data services by the data broker or directly with partners, executable by the IDS Connector.
2. Sovereignty about how the data is used in the supply networks has to be guaranteed.
    a. Require the definition of usage rules to consider data sovereignty.
3. Shared data for the digital twin should be independent of the local data formats.
    a. Connection of RDF-based requests to the local data source to locate data within the company and integrate received data.
    b. Mapping of RDF to local data sources to ensure independency from local data formats.
4. Data interchange should follow a global standard for easy usage in the whole network.
    a. The data is shared based on the agreed network ontology, stored in the ontology repository, and connected to metadata defined by the IDS architecture.

Data-Sharing Ecosystem Alternatives to IDS for Interoperable and Sovereign Data Spaces

Interestingly, the IDS is not the only initiative that adopted the network model for data sharing employing a single-entry point component and promoting security in the data exchange. As depicted in **Figure 8**, Carvalho et al. (2020) introduced the 4-corner topology model, in which two distinct (back-end) systems can securely and reliably exchange data with each other through the eDelivery access point nodes that are interfacing them. To communicate, a back-end system connects to its access point (using the connector) and submits a message outbound to the destination

access point. Similarly, to receive the inbound messages, a back-end system uses the connector to either pull messages from its access point or the messages are pushed by the access point. This way, organizations can securely communicate through the network once they have installed their own access point using an agreed protocol (AS4 messaging protocol in their case), irrespective of their proprietary IT systems. They noted that this approach will provide:

1. Trust establishment among participants in the message exchange network.
2. Secure data exchange with other members through standardized messaging protocols.
3. The use of reusable tools that are agnostic to the payload exchanged.
4. The possibility to increase the number of participants and exchanged messages.

Alternatively, the remaining literature promoted interoperability and security of data-sharing ecosystems by incorporating the Blockchain and DLT in their work. Abebe et al. (2019) mentioned that blockchain technology enables a decentralized data-sharing network through the *shared ledgers* (transaction records) maintained by a set of *peers*, instead of a central player. One of the main capabilities of blockchain implementations is the implementation of smart contracts, which are software codes designed to automatically facilitate, verify, and enforce the negotiation and implementation of digitally represented contracts and agreements without central authorities (Tan et al., 2020). In their work, they proposed a cross-network data transfer architecture for interoperability between permissioned blockchain networks through interfacing components called relays. These components are responsible for serializing and forwarding message requests to the destination relay as well as deserializing them and forwarding it to the recipient.

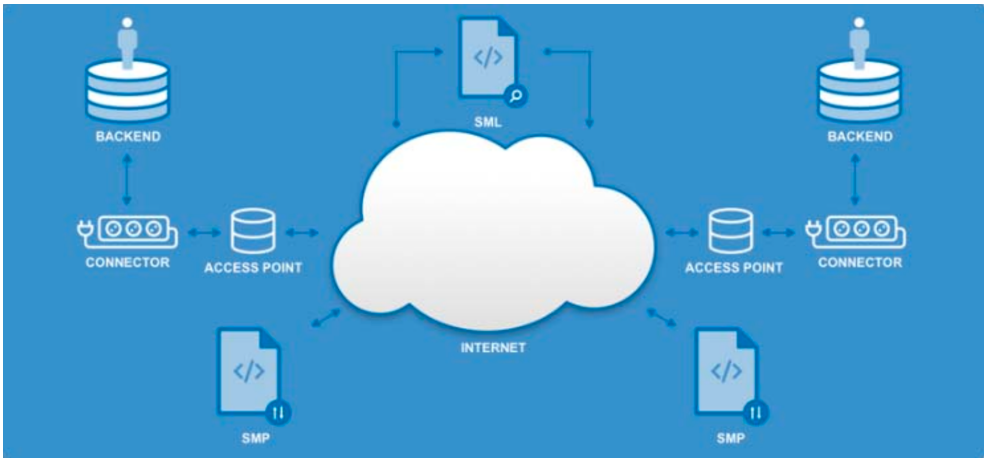


Figure 8 eDelivery 4-Corner Topology Model (Carvalho et al., 2020)

A similar architecture is adopted by Voswinckel et al. (2020) to realize a distributed storage network based on blockchain technology to share and distribute data to other companies. As presented in **Figure 9**, every company generates and keeps its data in its own environment and represents a node in the blockchain network to share data securely. Through this design approach, network participants could leverage smart contracts to automate specific business logic (e.g., to verify data correctness and completeness).

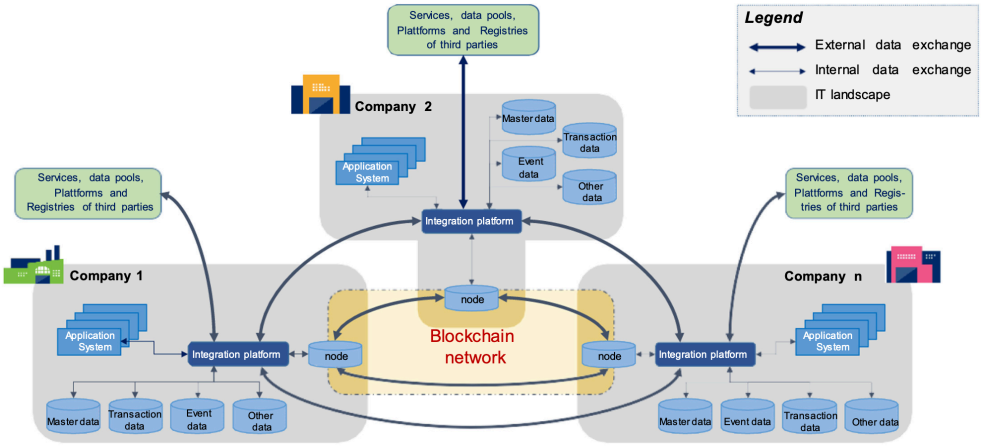


Figure 9 Decentralized Blockchain-based Data Sharing and Storage Network Architecture (Voswinckel et al., 2020)

Based on the literature so far, there are several drawbacks related to the implementation of a blockchain, becomes more acute with. Abu-elezz et al. (2020) described several key challenges associated with the implementation of blockchain technology in the healthcare sector, with high energy consumption and slow processing speed being the most prominent ones. This challenge becomes more acute with the increase of the number of network participants, leading to the question of the scalability of this solution once the demand grows. Tan et al. (2020) also share the same concern for their (green) logistics context that such technology requires each node in the blockchain network to store and validate every data collected and processed, which demands more electricity and large storage capacity. Consequently, real-time data collection can also lead to network congestion as more IoT devices are deployed, reducing the quality of service. Next to that, although blockchain can save costs in other aspects in the long run, the initial installation cost (along with other costs, e.g., training costs, operation costs, and maintenance costs) is perceived to be quite steep. The maintenance aspect of the blockchain-based platform can also become a problem in case of insufficient personnel and IT professionals who are experts in operating and troubleshooting the technology. Due to these arguments, rather than being willing to make a large investment, companies tend to defer their adoption of blockchain technology.

3.5. Summary and Conclusion

This literature review has provided an adequate understanding of the current data-sharing state of affairs in the logistics industry, along with the neighboring contexts, through a systematic approach. From the result analysis process, several barriers have been identified, covering the challenges of syntactic interoperability, semantic interoperability, standards adoption, and interoperability, data confidentiality and sovereignty, and (de)centralized data-sharing. Several solutions have been discussed to treat these challenges. There is a call for the development of an inclusive and sovereign data-sharing ecosystem that prioritizes bringing together already existing technologies and standards to foster network adoptions. Such an ecosystem may incorporate schema mappings and data transformation functionalities to alleviate syntactic interoperability problems, with several studies suggesting the adoption of the ESB that can be provided by an ISP. An alternative is the adoption of ontologies and Semantic Web technologies to lower semantic interoperability issues, which, through this approach, unlocks the possibility for Linked Data implementation. Although there was a dilemma between the centralized and the decentralized paradigm for the ecosystem, the trends from recent studies tend to promote a decentralized approach considering the demand for data sovereignty and a level playing field for every stakeholder in the network. Several initiatives have been proposed that satisfy these prerequisites, with the IDS initiative being the most discussed in recent developments. Blockchain technology came as an alternative, though, its cost and benefit justification are still questionable by the industry players, especially for SMEs.

To treat the previously stated interoperability and sovereignty issues in the literature, Piest, Iacob, et al. (2020) proposed a high-level architecture that comprises two main components. The first one is the Connector Store, which provides a repository of mappings for network participants to facilitate data exchange between heterogenous ICT environments. The second one is the Interoperability Simulator, which simulates collaboration opportunities between participants prior to implementation and might combine the Digital Twin and the IDS principles for simulating the interoperability within the network as demonstrated by Cirullies and Schwede (2021). This thesis aims to realize the former to support the connection between stakeholders in the Dutch logistics sector according to the IDS principles and realize an interoperable and sovereign data space for the participants. In combination with the latter component, the proposed ecosystem is expected to make the secure exchange and sharing of real-time data more accessible and affordable to SMEs. Due to its role as a guiding principle for this thesis, in the next section, an extensive description of the IDS ecosystem and its relationship with the Connector Store will be discussed.

4 International Data Spaces (IDS)

In the previous chapter, state-of-the-art data-sharing in the logistics industry has been discussed and it is discovered that the IDS has been adopted by most studies to establish an interoperable and sovereign data-sharing ecosystem. Therefore, this chapter describes what is currently known about the IDS, elaborates on the guidelines prescribed by the IDSA, and discusses how these design principles provide the solution for data interoperability and sovereignty. To elaborate on this, **Section 4.1** describes the background information of the IDS. Next, **Section 4.2** identifies the roles, components, and processes essential to manage data interoperability and sovereignty in an IDS ecosystem. In **Section 4.3**, we instantiate the IDS design principles into the problem context of this thesis, resulting in a set of enterprise architecture viewpoints that define the role of a Connector Store and other essential components in a logistics data space. **Section 4.4** concludes this chapter by discussing the implementation agenda for the following chapters.

4.1. Overview of the IDS

Initially referred to as the Industrial Data Space, the term was updated to the International Data Spaces in March 2018 to reflect the vision of building data-sharing ecosystems crossing national boundaries (Otto & Jarke, 2019). It is the initiative of various international research institutes and industrial enterprises to establish a decentralized data-sharing platform in which partners of different sizes can exchange data (regardless of the type of data), while still granted the capability of being entirely sovereign with regard to their data (IDSA, 2020c). Data interoperability and data sovereignty are, among others, the fundamental qualities in such a virtual space for data. The IDSA, the consortium of IDS members and contributors, aims to unlock the data economy of the future, where data remains with the data owner until it is needed by a trusted business partner. When the data is shared, terms of use are attached to the data, which will technically enforce how the data is allowed to be used on the data user's side (Otto et al., 2019). The IDS also provides the basis for the development of smart services and the adoption of existing standards and vocabularies. This allows seamless business processes orchestration across companies' borders in a semantically interoperable way (Bader, Pullmann, et al., 2020). The IDS puts forwards trust, security, interoperability, and sovereignty in mind by distributing the responsibility of establishing such a data-sharing ecosystem into several trusted business roles and application components.

As shown in **Figure 10**, IDS grants the participants access to participate in the ecosystem through an IDS Connector. This is an application component that facilitates secure data exchange between a data owner and data user through the enforcement of usage policies for the data consumer to use, process, and proliferate the shared data. An IDS Connector supports data interoperability between the enterprise systems of the data providers and data consumers through the execution of trusted software packages called the IDS Data Apps. Being an independent and

reusable software application retrievable from an IDS App Store, an IDS Data App provides additional data processing capabilities such as data transformation, cleaning, aggregation, analysis, anonymization, etc. (IDSA, 2019, 2021c).

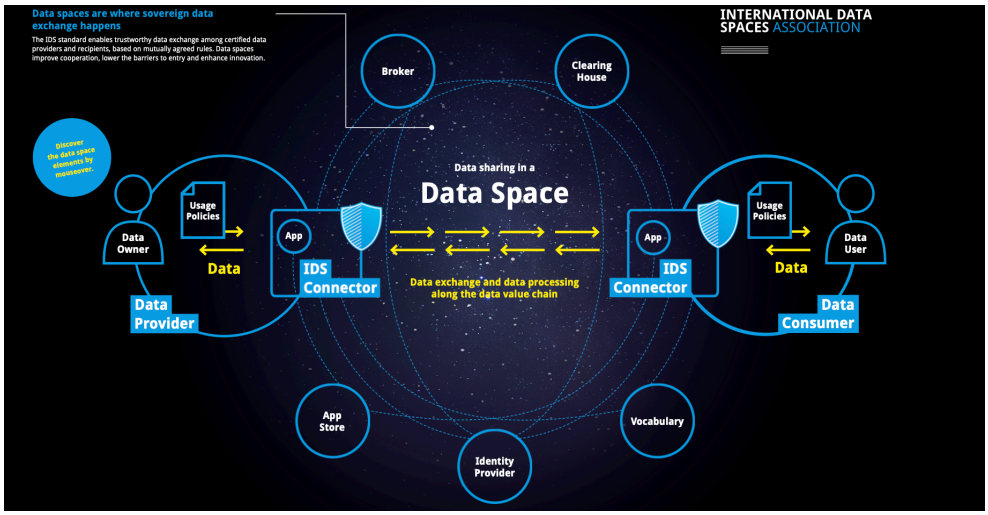


Figure 10 IDSA Infographic Data Sharing in a Data Space

Maintaining interoperability and sovereignty by the core participants alone can impose operational and cost-efficiency challenges. Thus, the IDS promotes the transfer of metadata management (for data interoperability, sovereignty, and provenance) to specialized intermediary organizations, e.g., trusted (meta)data brokers and clearing houses (Dalmolen, Bastiaansen, Kollenstart, et al., 2019). The IDSA has published several inter-related documents for companies to guide the development and adoption of the IDS principles. One of these documents is the IDS RAM, which serves as the general guideline on how to build up such an ecosystem for data-sharing, along with the distribution of these roles' responsibilities and the high-level architecture of the prescribed infrastructural components (IDSA, 2019). While the other documents describe further the details of the components or processes under discussion (e.g., the DIN SPEC 27070:2020-03 document that describes the reference architecture and requirements of the IDS Connector (IDSA, 2020a), etc.). In the following section, the key roles, components, and processes relevant to establishing an IDS ecosystem will be discussed.

4.2. IDS Guiding Principles and Architectural Layers

Proven trust takes time, however, that alone is not sufficient to ensure companies' trust and sovereignty to share data in today's dynamically changing supply network without adequate technical enforcement (Bastiaansen et al., 2020). Trust and data sovereignty are one of the main strategic requirements that the IDS aims to meet. To enhance trust with each other, each participant is evaluated and certified before being granted access to participate in the ecosystem. Evaluation and certification are also imposed on each technical component leveraged in the IDS. With respect to

data sovereignty, data owners are given the ability to attach (a machine-readable) usage policy to their data before it is transferred to the data user, implying that the data user has to fully accept this policy prior to consuming the data. The IDS promotes standardized interoperability for data exchange through the IDS Connector, which can be obtained from different vendors and execute IDS Data Apps for data format alignment. To manage these requirements, the IDS prescribes the distribution of responsibilities to several roles and components that are interacting with each other through standardized processes. The IDS RAM described this distribution in five architectural layers, starting from the Business Layer, Process Layer, System Layer, Functional Layer, and lastly the Information Layer (IDSA, 2019). In the following sub-sections, the former three layers will be discussed while also integrating descriptions from the other two layers if needed.

4.2.1. IDS Business Layer

The Business Layer defines the different roles that participants in a data space may assume, along with the responsibilities and relationships these roles have with each other. Next to that, it also lays the blueprint for the other layers as it also sheds some light on the interactions among roles as well as them with the underlying technical components and information entities. To ensure full functionality of an IDS ecosystem, the IDSA prescribes four core business role categories comprised of Core Participants, Intermediary Participants, Software / Service Providers, and Governance Bodies (IDSA, 2019).

The Core Participants refer to the participants who are directly involved in every data exchange activity in a data space. This category belongs to the Data Owners, Data Providers, Data Consumers, Data Users, and App Providers. A Data Owner is the entity that owns and creates the data, whereas a Data Provider, is the one who publishes the data. In some cases, a Data Provider, who could be an external IT service provider, may be needed by a Data Owner to manage and publish their data. Similarly, a Data User is the entity that has the legal right to use the data as specified by the usage policy, while a Data Consumer is the one who requests the data from either a Data Owner or a Data Provider after accepting the defined usage policy. A Data Owner can take the role of Data Provider if they publish their data themself. Likewise, a Data User can also assume the role of Data Consumer when they use the data that they requested themself. Meanwhile, the App Provider is the one who develops and provides Data Apps to be used and deployed into these participants' IDS Connectors. Therefore, these Data Apps should be described with metadata (that describes its functionality, compliance, etc.) by the App Provider.

Intermediary participants are the trusted entities responsible for establishing trust, providing and managing metadata, and supporting data exchange between core participants. This role comprises the Broker Service Provider, Clearing House, Identity Provider, App Store Provider, and Vocabulary Provider. In this category, one role is possible to assume the other intermediary roles at the same time. The Broker Service Provider facilitates the discovery of data sources and other

participants by managing and provisioning metadata to the participants in the ecosystem. The Vocabulary Provider is responsible to manage and offer vocabularies, reference data models, or metadata elements to describe datasets. The Clearing House provides clearing and settlement services for data exchange (and financial) transactions by logging all involved activities reported by Core Participants after a data exchange. The Identity Provider provides the service to manage and validate the identity information of other participants. Whereas the App Store focuses on managing and provisioning the Data Apps offered by App Providers.

Meanwhile, the last two categories refer to entities that provide software, services, evaluation, and certification processes for the participants and software components to operate in an IDS ecosystem. The Software Provider provides the software for implementing the functionalities required to participate in the ecosystem (e.g., IDS Connector, etc.). The Service Provider, on the other hand, provides deployment and hosting services of these software components and other technical infrastructure for the participants' cost and operational efficiencies. The Certification Body oversees the certification process of participants' organization and IDS-related software components. To investigate whether a participant who is requesting a certification is qualified to be certified, the Certification Body can appoint certified auditing companies (e.g., PwC, Deloitte, etc.) to play the role of Evaluation Facilities. Such a role evaluates the participants' organizational and technical capabilities compliance for operating in IDS. Thereupon, the Certification Body will issue the certificates based on their evaluation results.

4.2.2. IDS System Layer

The System Layer specifies the technical components of the IDS ecosystem and maps these components to the roles that are interacting with them. The IDS RAM, IDSA (2019), specified that the three core components comprise the IDS Connector, Metadata Broker, and App Store (along with the IDS Data Apps themselves). The IDS realizes a distributed data-sharing network that relies on the (peer-to-peer network) connection of multiple nodes where core components (e.g., IDS Connectors, IDS Data Apps, etc.) are hosted.

The DIN SPEC 27070:2020-03 document described the IDS Connector as a software component located at a company's logical border that defines its interfaces to enable data exchange with external entities, orchestrates smaller and more specific IDS Data Apps to improve data interoperability capabilities, and enforces data usage policies to maintain data sovereignty for data owners (IDSA, 2020a). Formerly called the security gateway, the IDS Connector is responsible for executing the complete data exchange process from the internal data sources and to the enterprise systems of other participants. It provides connector self-description metadata (e.g., technical interface description, authentication mechanism, exposed data sources, and associated data usage policies) to the Metadata Broker.

The IDS Data App is described by the IDSA as an independent and reusable software application that can be deployed on and executed by the IDS Connector to provide additional data processing capabilities (e.g., data transformation, cleaning, aggregation, analysis, anonymization, etc.) (IDSA, 2019, 2021c). Data Apps are data services bundled as container images for simple installation by a container management application (e.g., Docker[23]). Therefore, an IDS Connector should have a configuration manager that utilizes a middleware or ESB technology to facilitate the orchestration of these containerized Data Apps.

The Metadata Broker is a repository system, based on the Connector architecture, that manages the publication and maintenance of associated metadata of the data sources and other IDS Connectors available in an IDS ecosystem in a way that supports participants for lookup functionalities (IDSA, 2019). To carry out its metadata management responsibility, a Broker Service Provider exposes its repository's interface for data owners to publish their metadata. The metadata can be stored in the BSP's internal repository and made available for structured queries submitted by the data user. As the Metadata Broker only retrieves and distributes metadata of data and services, this repository application is not to be confused with brokers in common message-based systems retrieving and distributing the data offered by data owners and requested by data users. Therefore, the direct data exchange and usage negotiation processes by the data users and owners are not part of the responsibilities of this broker.

Another intermediary software component is the Clearing House application that logs all data transactions between participants to, if necessary, facilitate conflict resolution in data exchange scenarios. This application logs these transactions by providing an interface for Core Participants to notify when (1) a data user requests data from a data owner and (2) a data owner provides data to the data user, and afterward about the usage of the data by the data user. More details on the specifications, requirements, and architecture of these IDS application components (i.e., IDS Connector, IDS Data Apps, and Metadata Broker) will be discussed in the following **Section 4.3**.

4.2.3. IDS Process Layer

The Process Layer describes the sequential interactions that take place between different business roles and application components in the data space (IDSA, 2019). The IDS RAM specified the first essential process to be the Onboarding process. This process as depicted in **Figure 11**, prescribes the steps for a company to join the data space as a data owner or data user. Firstly, the company requests an identity from the Evaluation Facility to be used in the data space, which then will be issued by the Certification Body in the form of a certificate. Secondly, the company requests an IDS Connector from a Software Provider, which then will be installed in the environment of the company's preference. Thirdly, the IDS Connector receives a

[23] https://www.docker.com/

digital X.509 certificate that corresponds to the participant's and connector's certification. Fourthly, the participant may request and install Data Apps for the IDS Connector for additional data processing or transformation capabilities to facilitate data interoperability. Lastly, the IDS Connector may be made available for other participants to be discovered and interact with it by publishing its self-description to a selected Metadata Broker.

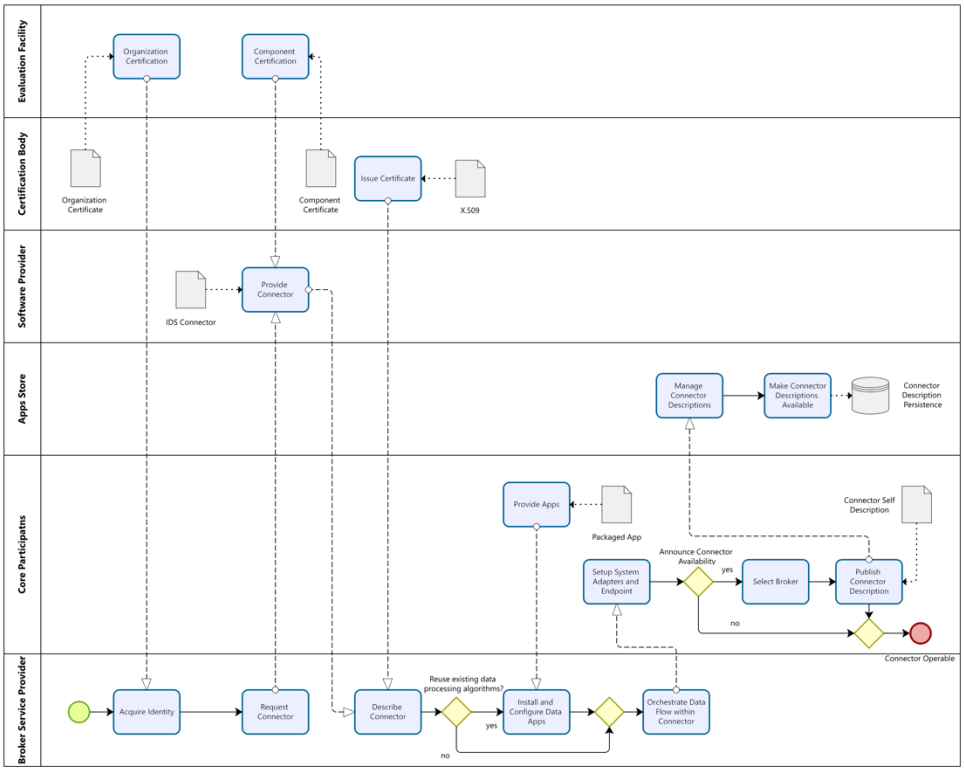


Figure 11 IDS Process Layer - Onboarding Overall Process

Upon acquiring and setting up the IDS Connector, the participant can initiate data exchanges with other participants. As shown in **Figure 12 (a)**, data users can request data from a data owner either by directly contacting the data owner's IDS Connector or if the data owner's IDS Connector address and endpoint are not known, then they may inquire about this information from the Metadata Broker. The broker will provide the requested information based on the self-description metadata the data owner submitted earlier. With this metadata, data users can contact the data owner's IDS Connector to select and request the offered data, which is further detailed as the Invoke Data Operation sub-process.

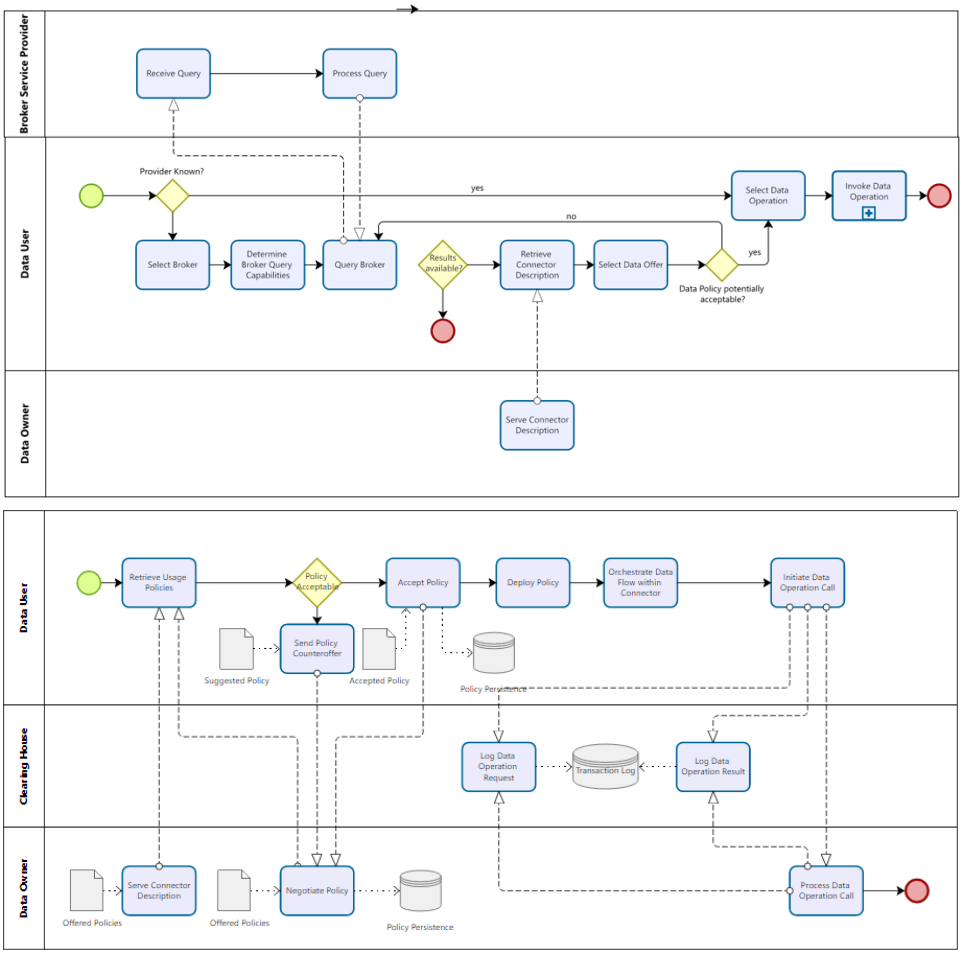


Figure 12 IDS Process Layer – (a) Exchanging Data and (b) Invoke Data Operation Processes

Data usage policy definition and enforcement are the foundations for data sovereignty management in an IDS ecosystem. This data usage policy enforcement implies that the data being shared can be used, processed, or further shared by the data user to another party only under the constraints that it was specified by the data owner and after the data user agreed to it. Therefore, as illustrated in **Figure 12 (b)**, upon selecting some offered data, data users will be presented with the usage policy that comes with it. In this stage, the policy negotiation will take place, meaning that the data user will either accept the policy or suggest a counteroffer. Afterwards, an agreement will be made, and the requested data will be transferred from the data owner's IDS Connector to the data user's IDS Connector. If data exchange logging for future conflict resolution is deemed necessary, then this data exchange's transaction log can be conveyed to a Clearing House.

4.3. IDS Adoption Reference Architecture Model for Dutch Logistics Sector

In the quest for the instantiation of a secure and interoperable "logistics data space", we apply these guidelines provided by the IDSA to the context of the Dutch Logistics Sector. To do this, this thesis leverages an EA approach. The TOGAF[24] is adopted to optimize an enterprise business and IT landscape into an integrated and aligned environment that contributes towards the reconciliation of its business strategies with the ones promoted by the IDSA. Adhering to TOGAF ADM, such an enterprise architecture should be modeled with formal notations and relationships that are mapped into multiple architectural layers (Lankhorst, 2009). ArchiMate[25] serves as a good fit, as it structures its architectural elements into seven architectural layers. This separation of concern promotes communication between business analysts and IT developers about the alignment of companies' high-level business requirements with the underlying software applications and communication infrastructure.

4.3.1. Motivation Viewpoint of IDS Adoption for a Logistics Data Space

The ArchiMate specification starts with a motivational viewpoint that associates business requirements with stakeholders, goals, assessments, drivers, and outcomes. **Figure 13** presents such a viewpoint. From the Stakeholder Analysis discussed in **Chapter 2**, the development of a Connector Store and other essential components for an IDS-based Logistics Data Space involves three groups of stakeholders. The first group comprises representatives of the Dutch Logistics Sector, which refer to the SUTC and TNO. SUTC belongs to this group, as it represents companies in the logistics sector to overcome interoperability problems between their enterprise systems for sharing data. This demand is met well by eMagiz Services B.V. as an Enterprise Integration Platform Provider, the second group, to support data transformation and standards adoption. Meanwhile, companies are also becoming more aware of the sensitivity of their data as an asset, and so is their demand for control over their shared data. Therefore, TNO, supporting the advancement of the Dutch Logistics Sector, is in the quest to investigate the suitable IT infrastructure for this purpose. Acting as the third stakeholder group of being a member of the IDSA, TNO aims to valorize the IDS infrastructure for improving data sovereignty in the sector.

Sharing operational and real-time data promises companies with enhanced supply chain coordination and optimized core competencies (Iacob et al., 2019; Pham et al., 2019; Slavova, 2021). To encourage logistics companies to share data, we need to lower the effort for (1) managing interoperability problems (e.g., syntactic, semantic, process, etc.) and (2) raising data sovereignty (e.g., data confidentiality, data access, usage policies, etc.) for data space participants. In the Motivation Viewpoint, we

[24] https://www.opengroup.org/togaf

[25] https://pubs.opengroup.org/architecture/archimate3-doc/toc.html

symbolize these targets as the outcomes (i.e., target dartboard) indicating the end results when the challenges are addressed. Associated with these outcomes are the data-sharing challenges, which are illustrated using the assessment notations (i.e., magnifying glass). We list these challenges according to the results extracted from the SLR in **Chapter 3**. Bouter et al. (2022); Ferreira et al. (2012), among other authors, reported the existence of diversified standards, message format, and data schema or structure within the logistics domain along with the hurdles of reconciling them. Therefore, setting up system integration and addressing data interoperability can be resource-intensive and time-consuming (Andreeva et al., 2016; Debicki & Kolinski, 2019). Veenstra (2018) also noted that much data related to logistics infrastructure (e.g., roads, bridges, maintenance schedules, water heights, etc.) were poorly published and most projects are directed to enhance the availability of these data (e.g., through APIs, etc.). From the data sovereignty perspective, Tran et al. (2016) pointed out that organizations perceive data sharing as a security risk for the leakage of sensitive information and competitive advantage. To address this problem, the IDSA promotes the IDS vision that leverages the IDS Connector and its functionalities to define and enforce data usage policies. However, this step requires a uniform organizational and technological maturity level for the network participants to comply with.

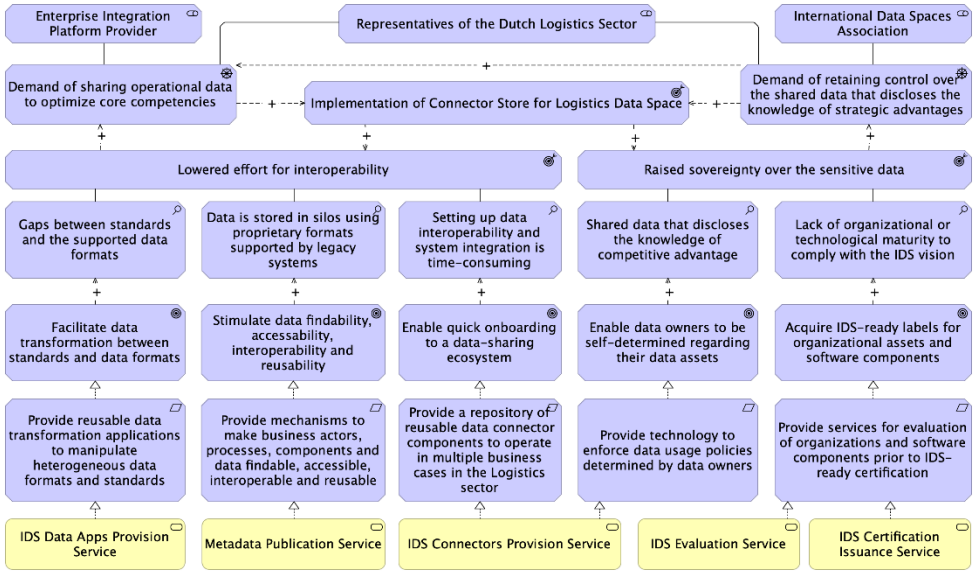


Figure 13 Motivation Viewpoint of IDS Adoption and Connector Store Implementation

To solve these data-sharing challenges, we list a set of goals and requirements mapped to the corresponding assessments. To treat the conflicting data schema and diversified standards, Ferreira et al. (2012) recommended data mapping and transformation mechanisms. Aligned with this purpose, the IDS recommends

participants to employ the reusable software packages (i.e., IDS Data Apps) provisioned by a Data Apps Store to be used in their IDS Connector environment. The value of data can be maximized if it can be discovered by, shared with, and reused by partners under formally defined descriptions of data provenance and terms of use. To stimulate the findability, accessibility, interoperability, and reusability of data assets, Top et al. (2022) advised the adoption of the FAIR principles in a data-sharing network. The instantiation of a Metadata Broker in the IDS environment connects well with this notion by providing their metadata publication service to the participants. As joining a data space can be a demanding investment for logistics companies, Debicki and Kolinski (2019) called for an IT infrastructure that can support all parties with a quick and easy integration to satisfy various business models. Therefore, network participants should be provided with a repository of IDS Connectors that are developed with and specialized for different sets of contexts and use cases. This notion is aimed to encourage the member of a logistics data space to use the IDS Connectors for defining and enforcing data usage policies when sharing their data assets with partners. With this goal in mind, we strive to enhance participants' control and sovereignty over their data within the logistics data space. Finally, as these goals require a balanced network organizational and technological maturity level, the IDSA advised the network participants to be evaluated and certificated with IDS-ready labels prior to joining the data space environment.

4.3.2. Service Realization Viewpoint of IDS Certification for a Logistics Data Space

The IDS RAM describes the candidate participants as being evaluated and certified before participating in data space (IDSA, 2019). **Figure 14** highlights two main business services: the IDS Evaluation Service provided by the Evaluation Facilities and the IDS Certification Issuance Service carried out by the IDSA's Certification Body. The reason for this is that, in a sovereign data space, data exchange is only allowed for certified participants using software components that are certified as well (IDSA, 2019, 2020c). These business services correspond with the onboarding process and fifth requirement in **Figure 11** and **Figure 13** respectively. In the following architectural viewpoints (i.e., **Figure 14**, **Figure 15**, and **Figure 16**) the notation for business actors and business roles appears in orange color to indicate the starting point for the readers in inspecting the model. Another reason is to visualize the business behaviors stakeholders will perform in an IDS-based data-sharing ecosystem.

Following the previous Onboarding Process, the certification process starts with a candidate participant requesting an evaluation to an Evaluation Facility (e.g., an independent audit company). The Evaluation Facility assesses the candidate's organizational maturity and software components' compliance to operate in an IDS ecosystem. After a successful evaluation, the facility sends a pre-evaluation report to the IDSA's Certification Body, granting the candidate company an X.509 digital certificate. The Evaluation Facilities and Certification Body are also responsible for evaluating and issuing certificates for the IDS Connectors before being published in

the Connector Store. This is to ensure that the IDS Connectors used in the data space are complying with the predefined IDS specifications (more details in **Chapter 5**).

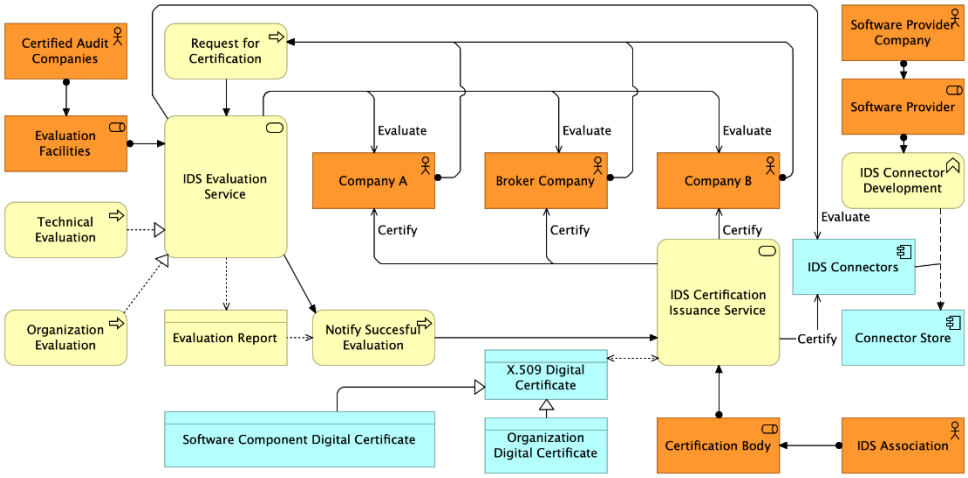


Figure 14 Service Realization Viewpoint of IDS Certification

4.3.3. Data and Metadata Exchange Viewpoint in a Logistics Data Space

Upon receiving certification to participate in the data space, the participants can request an IDS Connector and initiate the data-sharing processes afterwards. The viewpoint illustrated in **Figure 15** presents the roles, activities, and components essential to enact data sharing among participants in an IDS ecosystem. The business roles found essential to this viewpoint are the data owner, data user, and broker service provider (IDSA, 2019). After completing the certification process, data owners and data users can use one or more IDS Connectors requested from a Connector Store. In the viewpoint below, the Connector Store is located on the left side.

It is quite logical that companies interested to participate in a data space will have a diverse needs and requirements. Furthermore, each of the data spaces itself might also possess unique characteristics that require specific technical specifications of the IDS Connector. Therefore, multiple Software (and Service) Providers may start to develop, offer, and supply IDS Connectors for different purposes from now on (Firdausy et al., 2022b). These variants are revolving around different configurations (e.g., Base Connector, Mobile Connector, IoT Connector) and deployment configurations (e.g., on-premises vs in the cloud, Docker vs Kubernetes[26], etc.), among others. Such diversity also applies to the IDS Connector's security variant

[26] https://kubernetes.io/

(e.g., Base, Trust, Trust+), which governs different implementations of data usage policies and controls (e.g., Base Connector supports defining data usage rules but does not support its enforcement, as compared to the Trust and Trust+ variants) (IDSA, 2019, 2020a). In fact, in November 2022, the IDSA reported various IDS Connector implementations available and, so far, there are 16 variants have been provided and maintained by partners[27]. This may overwhelm potential participants and hinder their adoption of the IDS ecosystem in the future if the discovery of these connectors is left unattended.

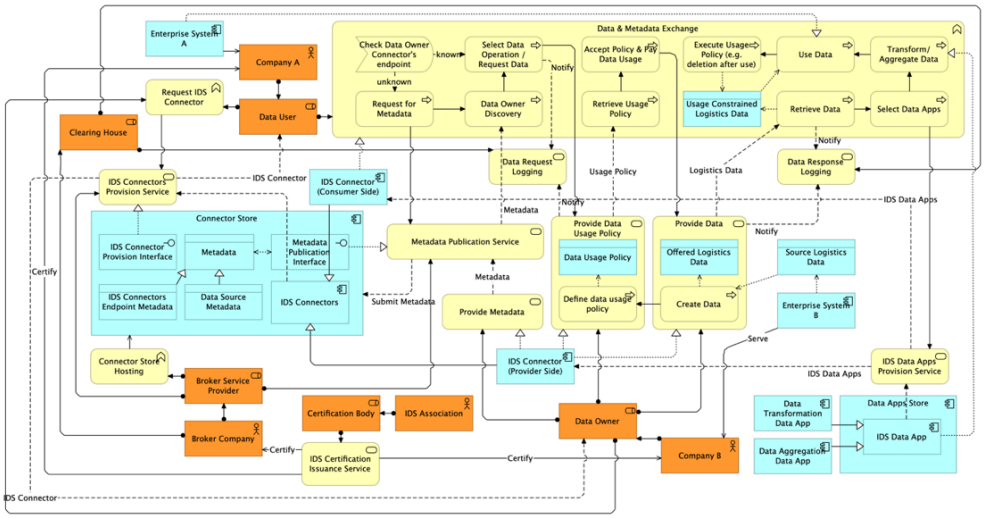


Figure 15 Data and Metadata Exchange Viewpoint brokered by a Connector Store

To cope with this issue, a Connector Store is proposed. This component is a repository of metadata that provides the participants with IDS Connectors that are suitable to their needs. The Connector store is responsible for the delivery of the IDS Connectors Provision Service, which is exposed to other participants in the environment. This meets the third and fourth requirements stated in **Figure 13**. The Connector Store offers a variety of IDS Connectors and facilitates their discovery by describing their specification, functionality, and other contextual information with metadata. In an IDS ecosystem, such a discovery process is normally facilitated by the Broker Service Provider role via the Metadata Publication Service of its Metadata Broker application. This broker typically only describes and publishes the information for the data space participants to find other members and their offered data sources or services. Despite this original function, we see that the responsibility of a broker service provider aligns well with the Connector Store's purpose. Thus, we propose the Connector Store to be managed by the Broker Service Provider role, as indicated in the bottom left corner. Besides, the IDSA acknowledged the possibility for the Broker Service Provider to also assume other intermediary roles

[27] https://internationaldataspaces.org/wp-content/uploads/dlm_uploads/IDSA-Data-Connector-Report-November-2022.pdf

at the same time (e.g., Clearing House, etc., as shown in the top left corner). This store extends the functionalities of a Metadata Broker by also supporting semantic discovery and selection of IDS Connectors from different Software Providers. This implies that it also provides participants with metadata about participating members and their data offering.

The other side of the viewpoint highlights a data owner and data user. As observed in the middle part of the viewpoint, data owners are the participants who generate, describe, and offer data to other participants in a data space. They describe the data by defining the metadata, which includes what the data is about and the policy to use it, in accordance with the Dataspace Connector's Data Model[28]. To improve the findability of their offered data, they can publish these metadata to a specific Metadata Broker for data users to query them through their own IDS Connectors. Meanwhile, data users are the entity that requests data from data owners, as well as use it in compliance with the previously agreed data usage policy. The top part of the viewpoint illustrates the process performed by a data user to find and request the offered data. This process relates well with the process previously defined in **Figure 12**, involving both a broker service provider and a clearing house for data discovery and logging respectively. Next to that, data owners or data users can select and request IDS Data Apps from an App Store to manipulate the data. This can be useful for each of them to, for example, transform the data from the as-is format into the format that is supported by their enterprise system or from their enterprise system's format into the standards commonly accepted in their community.

It is important to note that the IDS promotes data interoperability and data sovereignty as its two main value propositions through the distribution of responsibilities among business roles and software components. Despite that, establishing a complete IDS ecosystem with every role and software component instantiated can be too complex and costly (Firdausy et al., 2022c). Although important, not all the prescribed roles and components are directly contributing towards data interoperability and sovereignty. Taking the Vocabulary Provider as an example, this component manages and offers vocabularies and reference (meta)data models. However, this can add complexity to the overall landscape with regard to the process of exchanging data. Therefore, we omit the participation of this role and assume the usage of a specific data standard (e.g., OTM standard for this thesis) to be technically enforced by the IDS Data Apps. Another possible intermediary role in data exchange is the Identity Provider. Despite its vital contribution towards data sovereignty, in this thesis this role is omitted due to simplifying assumption that all participants have been authenticated, authorized, and are trusting each other. The Clearing House, however, is tracking the data usage by a data user, if this is specified by the data usage policy by the data owner.

[28] https://international-data-spaces-association.github.io/DataspaceConnector/Documentation/v6/DataModel

4.3.4. Infrastructure Functional and Deployment Viewpoint in a Logistics Data Space

The viewpoint below is presented to provide more clarity on the alignment between business elements and the IT infrastructure underlying the IDS-based Logistics Data Space. More specifically, the viewpoint is visualising the functional capabilities and deployment environment of the IDS Connectors and the IDS Data Apps in relation to the Connector Store. According to IDSA, a data space aims to facilitate data transfer to and from participants' systems, be it enterprise systems (e.g., CRM, ERP, etc.) or cyber-physical systems (i.e., IoT-enabled systems), by using a system adapter that supports necessary data format transformation and data usage policy enforcement. An IDS Connector constitutes such a system adapter that can be developed and delivered in different implementation types (e.g., web apps, mobile apps, or IoT apps) and deployment environments (e.g., on-premises or cloud environments) to serve multiple cases. Therefore, a software provider is needed in the data space to develop and provide the IDS Connectors to support the participants with several types of connectors (IDSA, 2019).

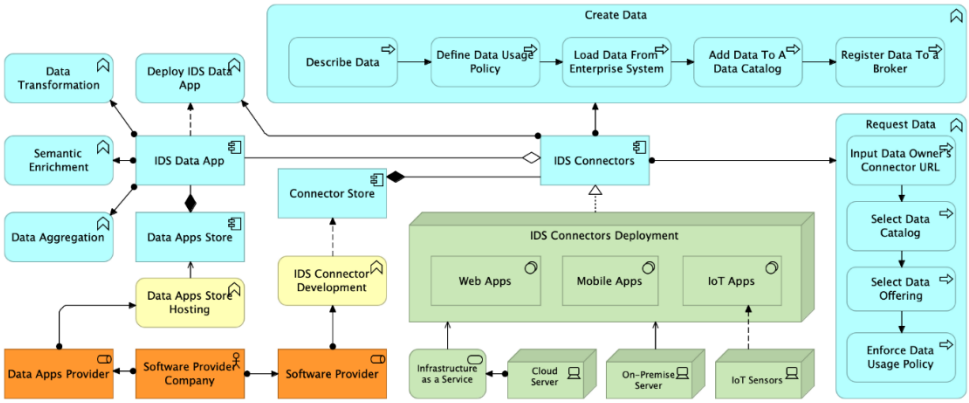


Figure 16 Infrastructure Functional and Deployment Viewpoint of IDS-related Components

As highlighted in **Figure 16**, a set of application processes support the IDS Connector in creating and offering data. Aligned with the support processes for data-sharing management explained by Dalmolen, Bastiaansen, Kollenstart, et al. (2019) in **Table 10**, the first step is to describe the data offered. In the second step, the IDS Connector provides the data owner with the option to define and attach the usage policy to the data. Such a usage policy is normally comprised of one or more rules. Thirdly, the data owner specifies the data sources for the IDS Connector to load the data. After the data is retrieved from (1) the company's internal enterprise systems, (2) REST APIs, or (3) direct database access, the fourth step is to specify which catalog this data is grouped under. The last step refers to registering the newly uploaded data to a specific Metadata Broker, which will make the data offering discoverable by the other participants of the data space.

Another set of application processes describes how the IDS Connectors are requesting data, initiated, and performed by data users. The first step covers the access of a data owner's IDS Connector through its endpoint URL. This endpoint URL is part of the metadata that the data users requested from the Metadata Broker when exploring and selecting the data offered in a data space. Next, the data user can select the catalog the offered data is grouped under. The data owner's IDS Connector will return the data usage policy of the selected data to the data user. After the data user approves the policy, its IDS Connector will make a contract agreement that bounds both parties to the data under discussion. It will then proceed to download the selected data from the owner and starts enforcing and monitoring data usage based on its corresponding access policy. An IDS Connector also supports the deployment and execution of IDS Data Apps. Such applications are developed and delivered by a Data Apps Provider, who is also responsible for describing them with metadata to make them discoverable and trusted.

4.4. Summary and Conclusion

This chapter concludes the **Problem Investigation** phase of the design cycle by providing an overview of the IDS and its design principles for enforcing data sovereignty and managing data interoperability. In addition, this chapter also elaborated on how the Connector Store can support the discoverability and provisioning of resources and infrastructures in an IDS ecosystem. Using an enterprise architecture modeling approach, we presented the contribution of each essential role and infrastructural component in a logistics data space toward achieving the strategic goals, which are defined by combining knowledge obtained from the SLR to satisfy the concerns of the stakeholders of this thesis. Based on the illustrated viewpoints, this thesis focuses on developing and demonstrating the operationalization of the IDS Connector and IDS Data Apps (i.e., to manage data interoperability and sovereignty on a logistics data space's runtime, as well as the Connector Store (i.e., to support the findability of the relevant components, enhancing connectivity between logistics partners). This calls for a deep-dive exploration of how such components can be designed and developed in accordance with the IDS technical specification, which will be discussed in **Chapter 5** and **Chapter 6**.

PART 2
TREATMENT DESIGN

5 The Design of an IDS Connector and IDS Data App

This chapter marks the beginning of the next phase of the engineering design cycle regarding the **Treatment Design** phase. The first step in this phase is to specify the requirements of the relevant application components for managing data interoperability and data sovereignty. Therefore, upon establishing the IDS Connector and IDS Data App as the essential components, the functional requirements of these components will be discussed in this chapter. For this purpose, **Section 5.1** elaborates on the software requirement specifications (SRS) of the essential components based on IDSA's specifications. Provided that the enterprise architecture specified in **Section 4.3** earlier contributes to achieving the stakeholder goals, **Section 5.2** investigates the software architectures to guide the development and the usage of the underlying components. Finally, **Section 5.3** discusses the conclusion of this chapter.

5.1. Software Requirement Specifications of an IDS Connector and IDS Data App

The first step in producing the SRS is to elicit the requirements of the IDS Connector and IDS Data App from several reference documents published by the IDSA. The first document being used is the IDSA RAM v3 which provides an overview definition and specifications of the individual components substantiating an IDS ecosystem, which has been discussed in the previous chapter (IDSA, 2019). The second referred document is the DIN SPEC 27070:2020-03 standards (IDSA, 2020a). It describes the requirements of an IDS Connector for the industrial data exchange, possible data exchange architectures supported by the IDS Connector, the communication infrastructure that the IDS Connector can be embedded in, and the minimum-security requirements to be ensured for participants in the network (e.g., data usage control). The third document, the IDSA Rule Book, presents a common governance framework that specifies the functional, technical, operational, and legal agreements that structure roles and interactions across the various parts of an IDS ecosystem (IDSA, 2020c). In addition, several GitHub documentation pages maintained by the IDSA are also referred to elicit more detailed specifications of the IDS software components (IDSA, 2021a, 2021b, 2021d, 2021e, 2021g).

5.1.1. IDS Connector Data Exchange & Communication Requirements

The IDS Connector is described as a gateway typically located at a company's logical border that defines its interfaces for external entities (IDSA, 2020a). This means that an IDS Connector can physically be implemented both at the company's premises and the premises of a service provider. The IDSA prescribes that an IDS Connector can be developed and offered to the market in three different schemes, namely, as open-source software, as an off-the-shelf product, or as an as-a-service offering

(IDSA, 2020c). As presented in **Figure 17**, the DIN-SPEC-27070:2020-03 standards document specifies the three different data exchange topologies that can be supported by an IDS Connector within an IDS data-sharing environment:

1. Peer-to-Peer

   Each client is on the same level as other clients and may also function as a server. A central infrastructure is not required for such architecture.
2. Client-Server

   The roles are clearly defined, i.e., each client communicates with several central servers.
3. Hybrid

   Each node in the network can function as a server and a client. For such infrastructure to work, a number of supporting components are required.

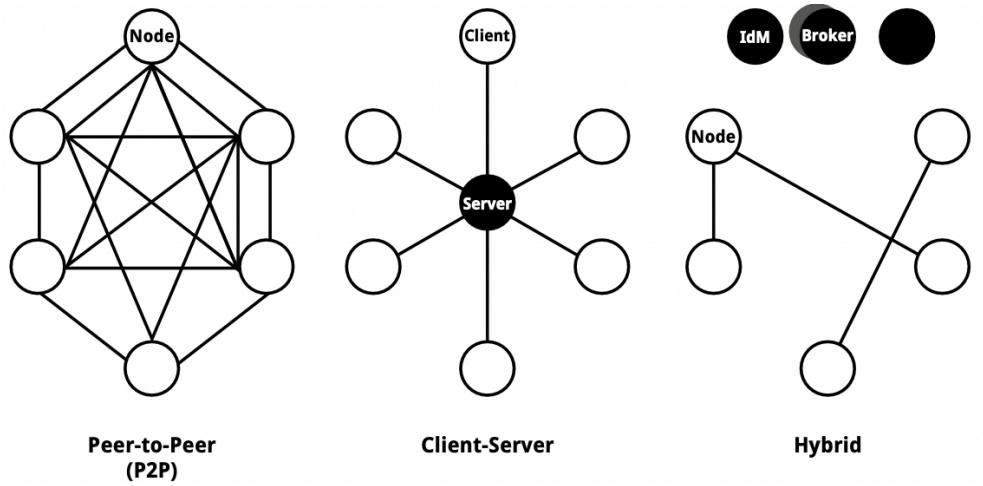


Figure 17 IDS Connector Supported Network Topology (IDSA, 2020a)

In principle, an IDS Connector is able to operate in a fully P2P or Client-Server topology. However, to ensure trust, security, and sovereignty while still maintaining discoverability and a level-playing field for all participants, the IDS prescribes that the IDS Connector should be embedded in an ecosystem in which the Hybrid topology is adopted. For this to work, connections with additional supporting components such as a Metadata Broker, Identity Provider, and Data Apps Store are required. The operations and management of these components themselves can be delegated to several different service providers. Nonetheless, the existence of these components is considered essential to ensure secure and flawless data exchange operation. This results in **Table 11** that lists the infrastructure components required for a network where the IDS Connector is embedded in.

Table 11 IDS Network Infrastructure Requirements

Infrastructure Requirement	Infrastructure Component	Description
NIR_1	IDS Connector	A Gateway that is typically located at a company's logical border and defines its interfaces with external entities
NIR_2	Identity Provider	Issuing identities to IDS Connectors and participants.
NIR_3	Metadata Broker	Acting as a service registry and registers IDS Connectors that are offering these services.
NIR_4	Clearing House	Logging all data transactions to facilitate conflict resolution in data exchange scenarios if necessary.
NIR_5	Data Apps Store	Providing data apps to be used by IDS Connectors for additional data processing services.
NIR_6	IDS Data Apps	Independent, functional, and reusable software application that can be deployed on and executed by an IDS Connector.

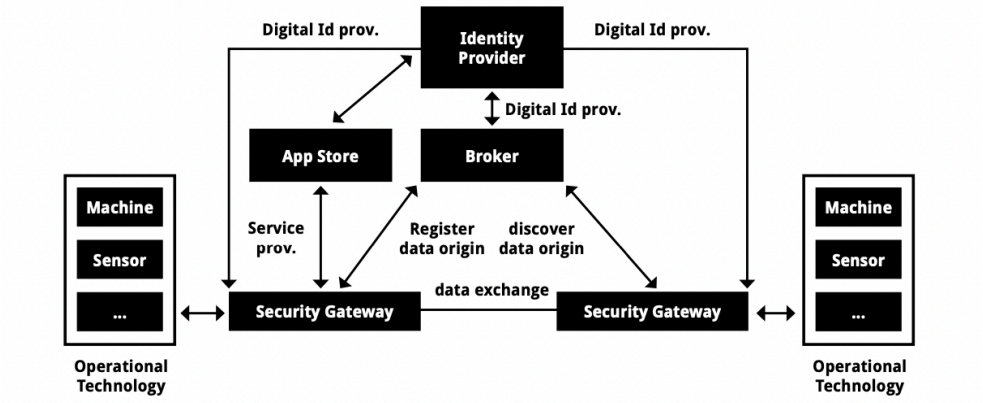


Figure 18 IDS Communication Infrastructure Architecture (IDSA, 2020a)

This table leads to the emergence of the architecture shown in **Figure 18**. The architecture depicts the IDS communication infrastructure that follows the previously mentioned Hybrid topology. For the IDS Connector to work in such a network, several requirements for each of the components listed in **Table 11** must be fulfilled. These requirements are then listed in **Table 12**, which groups them according to the related infrastructure components.

Table 12 IDS Communication Infrastructure Requirements

Related Infrastructure	Infrastructure Requirement	Description
IDS Connector	IR_1	The IDS Connector must connect the company's internal digital infrastructure with the external digital environment, controlling data flows and data access.
	IR_1.1	The IDS Connector must give proof of its own identity.
	IR_1.2	The IDS Connector must provide self-description metadata to inform other IDS Connectors about its data endpoints and other features or services it offers, e.g., supported security features, etc.
	IR_1.3	Both data and metadata to be exchanged must be available in a standardized format, e.g., following the Dataspace Connector's Data Model (IDSA, 2021b).
	IR_1.4	There must be a common vocabulary for describing and exchanging data, which can be represented by a domain-specific format agreed upon by participants, e.g., OTM or GS1 in the case of the Dutch Logistics Sector.
	IR_1.5	The IDS Connector must provide access control regarding data sources offered, i.e., define data usage policy (IDSA, 2021e).
	IR_1.6	The IDS Connector must be able to execute data services, i.e., IDS Data Apps, and orchestrate the interaction of such data services.
	IR_1.7	The IDS Connector must control access to internal networks and data sources.
Identity Provider	IR_2	The Identity Provider must issue identity attributes that allow identification, authentication, and authorization.
Metadata Broker	IR_3	The Metadata Broker allows participants to find other IDS Connectors by offering metadata to participants for searching and accessing other data sources or services.
Clearing House	IR_4	The Clearing House must provide transparency regarding all transactions.

Data Apps Store	IR_5	The Data Apps Store must offer IDS Data Apps to be used by IDS Connectors for them to perform additional functions or offer additional services.

It is important to note that, as discussed in **Section 4.3** earlier, not all the listed infrastructure components (e.g., Clearing House and Identity Provider) will be thoroughly explored. This is because this thesis only focuses on data interoperability management and data sovereignty enforcement. This decision applies to, for instance, the Clearing House component, which will be further investigated as a Master Thesis assignment. The Identity Provider is also beyond the scope of this thesis to reduce the complexity of the data space's initial development and implementation. Therefore **Table 12** is mostly focusing on the infrastructure requirements of the IDS Connector.

5.1.2. IDS Connector Operating System Architecture and Requirements

The IDSA introduces the DIN SPEC 27070:2020-03 standards as the reference architecture for the IDS Connector that specifies IDS Connector's architecture and requirements to be met when it is operational (IDSA, 2020a). The standards exhibit the OS stack that consists of a kernel and an Application Container Management Layer. As shown in **Figure 19**, the reference architecture comprises the following core functions:

1. Trust Anchor

   A component (can be implemented as software or hardware) that provides a manipulation-proof identity and allows integrity checks from outside the system.
2. Application Container Management Layer

   A container management layer (e.g., Docker) that is based on an operating system kernel that allows strict isolation of containers and restricts access to resources (e.g., memory or network interface).
3. Execution Core Container

   Controls the containers (IDS Data Apps) accommodating the services, which can be virtual machines or Linux containers with appropriate isolation. Furthermore, it controls communication among containers and the containers with the external environment.
4. Service Container (IDS Data Apps)

   Smaller container-based applications can be installed onto the IDS Connector, which can be obtained from the App Store or developed by the participants themselves. Each service provides an API over which it interacts with the external environment.

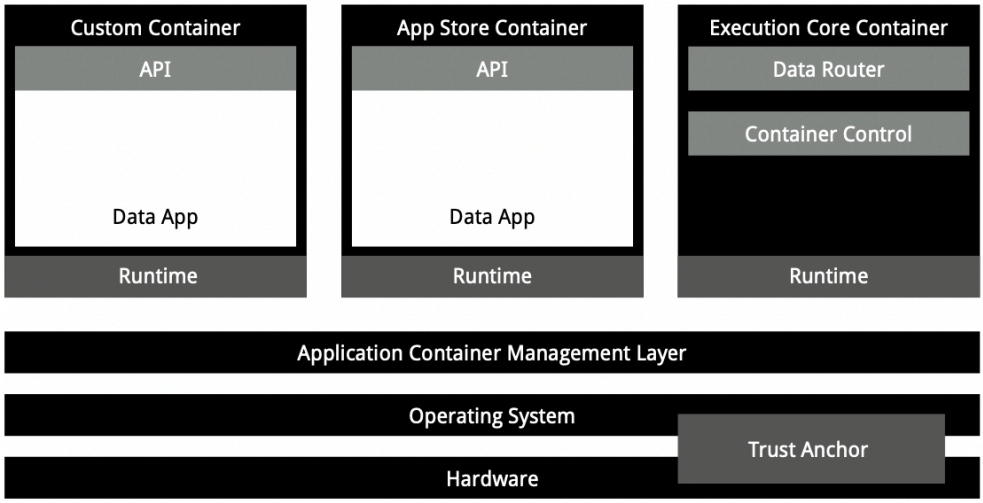


Figure 19 Functional Blocks of the IDS Connector Reference Architecture (IDSA, 2020a)

It is mentioned above that an IDS Connector should contain a container management layer such as Docker to support the installation and execution of strictly isolated containers as well as restricts access to resources. An IDS Connector should possess or should be deployed on an Execution Core Container that is capable of controlling other containers containing IDS Data Apps accommodating additional services. **Table 13** is presenting the list of detailed requirements related to the integrity of an IDS Connector's operating system.

Table 13 IDS Connector Operating System Requirements

Reference Identifier	Operating System Requirement	Description
NIR_1, IR_1, IR_1.6	OS_1	An IDS Connector supports the installation and execution of containers.
	OS_2	An IDS Connector enforces strict separation of data processing apps. Communication between apps takes place via approved channels only (i.e., whitelisting of data exchange channels).
	OS_3	An IDS Connector verifies the authenticity and integrity of IDS Data Apps prior to installation and execution.
	OS_4	An IDS Connector verifies the authenticity and integrity of all system components prior to execution.

	OS_5	Containers (of/for IDS Data Apps) are strictly separated from each other and from underlying operating system layers.
	OS_6	System data backups, as well as backups of data transferred between IDS Connectors, are always encrypted before being stored outside the system.

5.1.3. IDS Connector Data Apps and App Store Connection Requirements

The IDS Data App is described by the IDSA as an independent and reusable software application that can be deployed on and executed by the IDS Connector to provide additional data processing capabilities (e.g., data transformation, cleaning, aggregation, analysis, anonymization, etc.) (IDSA, 2019, 2021c). To enable communication between itself and the IDS Connector, the IDS Data App must expose a set of endpoints for data inputs and outputs. Through these endpoints, the IDS Connector manages the message routing and orchestrates the data flow from each other before the data is consumed by the underlying internal digital infrastructure (e.g., a company's enterprise system) or transferred to an external digital environment (e.g., other IDS Connector). Hence, an IDS Connector should have a configuration manager that utilizes a middleware technology (e.g., Apache Camel) to implement the routes. In conjunction with how an IDS Connector should be able to execute IDS Data Apps and orchestrate their interactions, **Table 14** lists the requirements prescribed by the DIN SPEC 27070:2020-03 standards that guide this design aspect.

Table 14 IDS Connector Data Apps and App Store Connection Requirements

Reference Identifier	Apps Connection Requirement	Description
IR_1.6	APS_1	An IDS Connector supports only apps possessing a valid signature. This signature is the signed checksum of the software artifact, which was created by means of a private key of the app publisher.
	APS_2	Before the self-description becomes available at the defined interface, the IDS Connector has to ensure that it is a valid instance of the information model for self-descriptions.
	APS_3	An IDS Connector supports IDS Data Apps carrying terms of use, allowing restriction of use and encapsulation of licensing information.

	APS_4	An IDS Connector checks the minimum requirements of IDS Data Apps regarding the runtime environment (e.g., with regard to memory capacity or the number of CPU cores) and ensures these requirements are fulfilled as long as an IDS Data App is active.
	APS_5	An IDS Connector supports IDS Data Apps to be delivered and installed as independent software containers (i.e., IDS Data Apps bring along possible dependencies of, e.g., software modules themselves and can be used irrespective of the IDS Connector's configuration).
	APS_6	An IDS Connector receives IDS Data Apps from a central App Store.

In terms of the extent of their control, the IDSA states that IDS Data Apps can be categorized into two kinds of app profiles, namely, Basic Profile and Supreme Profile (IDSA, 2021c). These distinctions are based on their administrative control of an IDS Connector and their capabilities to implement usage control policies on the data apps layer. **Table 15** summarizes the requirements that each profile of the IDS Data Apps has to fulfill.

Table 15 IDS Data Apps Profiles Requirements

Data Apps Requirement	Description	Basic	Supreme
IDA_1	The IDS Data App has clearly defined endpoints for the interfaces, at least for data input and/or data output.	√	√
IDA_2	The IDS Data App must be able to be integrated into the data flow of an IDS Connector and applied to it.	√	√
IDA_3	The IDS Data App must be signed by its developer to validate the origin.	√	√
IDA_4	The IDS Data App has no administrative control over an IDS Connector, and no direct interaction with an IDS Connector's API exists.	√	-
IDA_5	The IDS Data App has administrative control over an IDS Connector, and direct interaction with an IDS Connector's API exists.	-	√
IDA_6	The IDS Data App can execute usage control enforcement by itself, with	-	√

	direct, and indirect interaction with an IDS Connector. Direct interaction (e.g., requesting contract information) Indirect interaction (e.g., delivering usage control information via middleware)		

In terms of their functionalities and dependencies with other systems (e.g., data owner's enterprise systems, and other data sources), the IDS Data Apps can be categorized into four types of apps, with the fourth one is a combination of the other three app types (IDSA, 2021c). The first type is called the Data App, which is mainly responsible to perform small data processing tasks. This type of app should be designed as reusable and system independent as possible. The second type is referred to as the Adapter App, which connects data sources or backend (enterprise) systems to extend an IDS Connector. This app type extends the capabilities of the routing framework, or middleware, within an IDS Connector to connect to multiple endpoints and protocols.

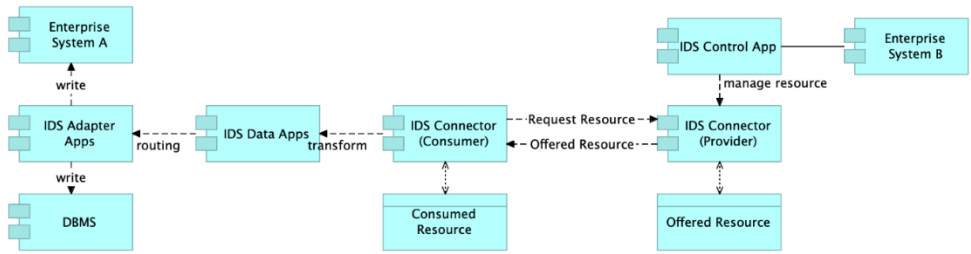


Figure 20 IDS Data App Types and Possible Interaction Viewpoint

Additionally, an Adapter App with the Supreme Profile will be capable to enforce usage control on its connected systems. Lastly, the Control App is the type of app that is directly coupled with a particular backend (enterprise) system to directly interact with the API of an IDS Connector (e.g., to manage resources). This means that this type of app only belongs to the Supreme Profile due to its administrative right on an IDS Connector. The architecture in **Figure 20** gives an illustration of how these different types of IDS Data Apps may interact with the other app types, IDS Connector, and the enterprise system of a company.

5.1.4. IDS Connector Data Usage Control Requirements

In IDS, data usage control is defined as the specification and enforcement of how data can be transferred, used, and processed within a data space (i.e., how different infrastructure components communicate with each other in exchanging and using the data) (IDSA, 2020a). The overall goal of this is to enforce data usage restrictions

on the Data User side after access to the data has been granted by the Data Owners (IDSA, 2019). In relation to the previously defined IR_1.5 regarding providing access control to the data sources offered (e.g., data usage policy), the IDS Connector Reference Architecture (IDSA, 2020a) specifies a list of data usage control requirements to be met by an IDS Connector. **Table 16** lists the requirements related to how an IDS Connector should support and manage data usage control.

Table 16 IDS Connector Data Usage Control Requirements

Reference Identifier	Data Usage Control Requirement	Description
IR_1.5	USC_1	An IDS Connector allows Data Owners to define usage policies with regard to the data being offered.
	USC_2	An IDS Connector offering data sends a usage policy to be applied to IDS Connector requesting data every time a connection is established.
	USC_3	An IDS Connector facilitates technical enforcement of data usage policy specified
	USC_4	Changes to the data usage policy can be made only by the Data Owner or the administrators of the IDS Connector. In case of changes made to the policy, the connection between two IDS Connectors is re-established.
	USC_5	Administrators of the IDS Connector cannot change rules regarding data flow without the Data Owner taking notice of the change and approving it.

Even though the DIN-SPEC-27070:2020-03 standards document provides specifications regarding how an IDS Connector should enforce the data usage policy, it does not elaborate on what kind of data usage policies an IDS Connector should be able to support (IDSA, 2020a). The IDSA RAM (IDSA, 2019) suggests some example characteristics for this purpose that describe the extent of the data usage policy enforcement:

- **Secrecy:** Classified data must not be forwarded to nodes that do not have the respective clearance.
- **Integrity:** Critical data must not be modified by untrusted nodes, as otherwise its integrity cannot be guaranteed any more.
- **Time to Live:** Data must be deleted from storage after a certain period of time.
- **Anonymization by Data Aggregation:** Personal data may be used only in an aggregated form by untrusted parties. To do so, a sufficient number of

distinct data records must be aggregated in order to prevent the deanonymization of individual records.

- **Anonymization by Data Substitution:** Data allowing personal identification (e.g., faces in video files) must be replaced by an adequate substitute (e.g., pixelized) in order to guarantee that individuals cannot be deanonymized.
- **Separation of Duty:** Two datasets from competitive entities must never be aggregated or processed by the same service.
- **Usage Scope:** Data may only serve as input for data pipes within the Connector; it must never leave the Connector and be sent to an external endpoint.

Based on these examples, the IDSA has further explained, on their GitHub repository, the policy patterns that their reference implementation of IDS Connector currently supports. The list of these patterns is shown in **Table 17**.

Table 17 IDS Connector Currently Supported Policy Pattern

Policy Pattern	Policy Pattern Name	Description
PP_1	Allow the Usage of the Data	Provides data usage without any restrictions
PP_2	Connector-restricted Data Usage	Allows data usage for a specific connector
PP_3	Interval-restricted Data Usage	Provides data usage within a specified time interval
PP_4	Duration-restricted Data Usage	Allows data usage for a specified time period
PP_5	Restricted Number of Usages	Allows data usage for n times
PP_6	Security Level Restricted Policy	Allows data access only for connectors with a specified security level
PP_7	Use Data and Delete it After	Allows data usage within a specified time interval with the restriction to delete it at a specified timestamp
PP_8	Local Logging	Allows data usage and sends logs to a specified Clearing House
PP_9	Remote Notifications	Allows data usage and sends notification messages

5.1.5. IDS Connector Information Model Requirements

In relation to the previously defined IR_1.1 and IR_1.2 regarding an IDS Connector capable of providing self-description metadata, the IDS Connector Reference Architecture specifies a list of information model requirements to be met by an IDS Connector (IDSA, 2020a). **Table 18** lists the requirements related to the mechanisms through which an IDS Connector should expose its self-description.

Table 18 IDS Connector Information Model Requirements

Reference Identifier	Information Model Requirement	Description
IR_1.1 & IR_1.2	INF_1	An IDS Connector provides self-description (i.e., metadata) via a defined interface.
	INF_2	An IDS Connector sends metadata to a Metadata Broker for being registered there.
	INF_3	The self-description contains at least the following information: a) A cryptographic hash of the IDS Connector certificate. b) The IDS Connector operator. c) Data endpoints offered by Connector. d) Log format of data endpoints offered. e) The security profile of Connector (i.e., security features supported). f) The IDS Connector ID.
	INF_4	IDS Connectors can evaluate the self-description of other IDS Connectors. This evaluation includes verifying that the self-description is a valid instance of the IDS IM.
	INF_5	Dynamic attribute tokens belonging to two communicating Connectors are transmitted every time a connection is established (see DIN SPEC 27070:2020-03 6.4.2) and can therefore be used for access control decisions.

5.1.6. IDS Connector Broker Service Connection Requirements

IR_1.2 and INF_2 state that an IDS Connector should be able to send metadata about its self-description and other features or services it offers to a Metadata Broker. This communication between them fulfills a part of the infrastructure architecture depicted in **Figure 18** earlier. **Table 19** lists the requirements that suggest the type of communication that should take place between an IDS Connector and a Broker.

Table 19 IDS Connector Broker Service Connection Requirements

Reference Identifier	Broker Service Connection Requirement	Description
INF_1, INF_2, INF_3	BRK_1	An IDS Connector supports broker service inquiries by means of browsing self-descriptions of Connectors registered in the Metadata Broker.
	BRK_2	An IDS Connector supports registration with a Metadata Broker by transmitting self-description.
	BRK_3	An IDS Connector supports updates of self-description stored at a Metadata Broker (e.g., when a new service is offered) and marking itself as available/unavailable.

5.2. Software Architecture Models of an IDS Connector and IDS Data App

This section provides more details on how an IDS Connector and IDS Data App can be assembled, can enforce data usage controls, and can support interoperable data exchanges as prescribed by the IDS specification. Architecture modeling in this stage further describes the application components under design using the graphical notations of a commonly used architecture modeling language. Sommerville (2015) indicates that different models can be developed to illustrate the system from different perspectives. In this section, we focus on presenting the IDS Connector and IDS Data App from their structural, behavioral, and interaction perspectives, while the environmental perspective of the components has been addressed in **Section 4.3** earlier.

For this, two modeling languages are utilized. The first one is ArchiMate. This modeling language is used for illustrating the structural perspective of the software components under design, along with the component's interactions with other elements from different layers (e.g., data owner or data user, and host hardware system). The second one is the UML. We use UML's Sequence Diagram to show the interaction between (1) a data owner or data user and their IDS Connectors and (2) their IDS Connectors with a Broker. To visualize the standardized object classes adopted by the IDS Connectors and the associations between them, the Class Diagram will be used.

5.2.1. IDS Connector and IDS Data App System Architecture

The design of the system architecture is essential for a software system to highlight the main structural components of a system and their relationships with each other (Sommerville, 2015). By the time this thesis is written, a wide range of IDS Connector implementations has been developed based on the software specifications discussed previously by several research organizations and software providers. At the same time, some of them have also been offered to the industry. Among them, two implementations are referred to in this thesis due to their high-level system architectures being publicly accessible. The idea for this is to extract the technical specifications of the already operating components in the field and translate them to the context of the logistics data space involving the Connector Store.

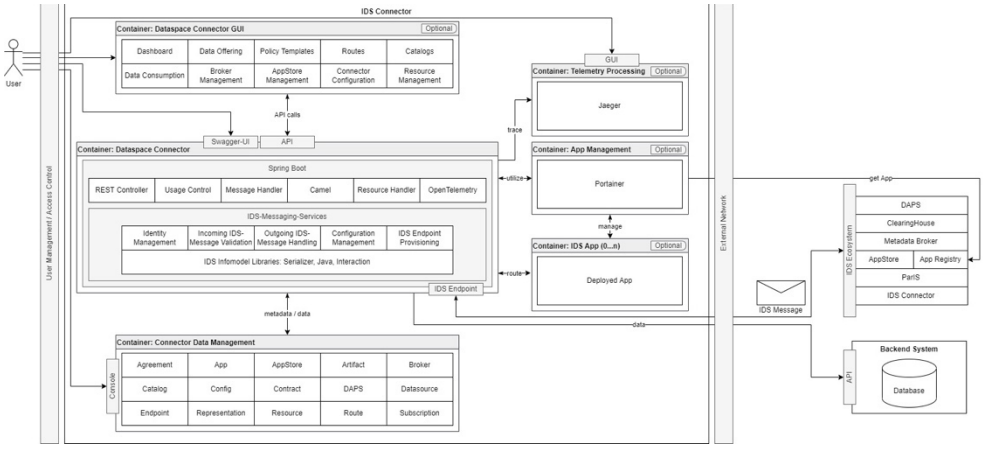


Figure 21 IDS Connector Implementation – Sovity Dataspace Connector

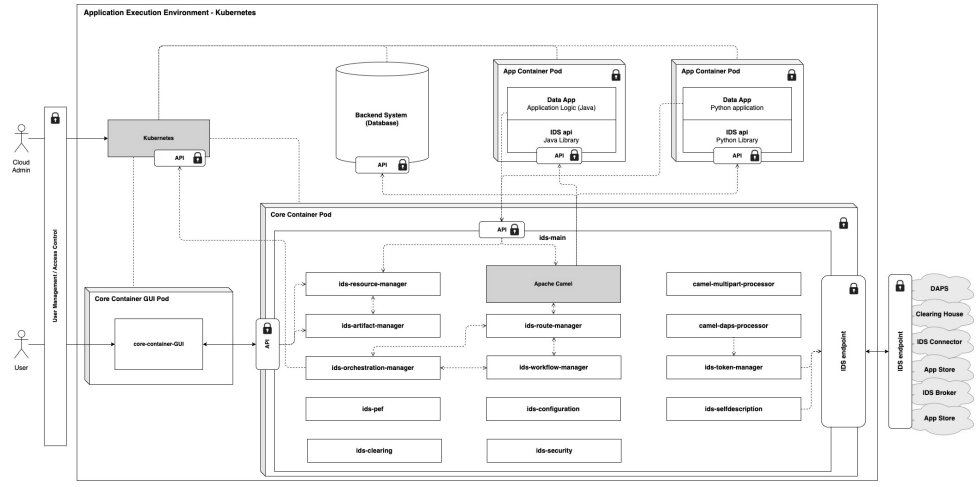


Figure 22 IDS Connector Implementation – TNO Security Gateway

The first one, represented in **Figure 21**, is the Dataspace Connector[29]. This IDS Connector was initially developed by the German research institute Fraunhofer ISST[30]. However, the newer version, this component is further developed and maintained by the institute's spin-off company called Sovity[31]. The second implementation, shown in **Figure 22**, refers to the TNO Security Gateway (TSG)[32], which is currently developed and maintained by the Dutch applied scientific research organization TNO[33]. From these two examples, we conclude that there are different implementation flavors of these two variants. However, several common elements and relationships that are essential for the IDS Connector environment can still be identified. To highlight this aspect, **Figure 23** is presented as a generalized viewpoint containing the common constructs extracted from these two examples.

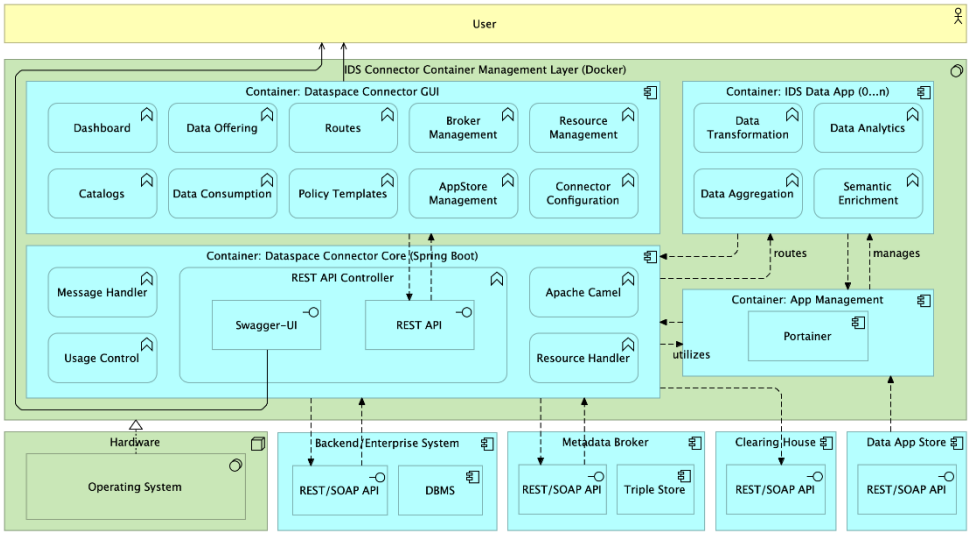


Figure 23 IDS Connector Internal System Architecture Viewpoint

In the two earlier mentioned system architectures, the implementation of the IDS Connectors is made up of several interconnected containerized applications deployed in one environment. The central unit, denoted as the "Container: Dataspace Connector" and "Core Container Pod", serves as the core back-end component. This central part is responsible for (1) the connector data management that complies with the IDS Connector Data Model, (2) the enforcement manager of the data usage controls, and (3) message routing from and to the other IDS components (e.g., IDS Data App, Metadata Broker, Clearing House, etc.) (IDSA,

[29] https://github.com/International-Data-Spaces-Association/DataspaceConnector
[30] https://www.isst.fraunhofer.de/en.html
[31] https://sovity.de/
[32] https://tno-tsg.gitlab.io/
[33] https://www.tno.nl/en/

2021b). Through this responsibility, the core connector satisfies the previously stated requirements IR_1.5 and IR_1.6. Correspondingly, a back-end database system should also be deployed in the environment to store the mentioned connector data. Although the IDSA does not specify a specific application development framework to be used, these two implementations used Java Spring Boot[34] as their backbone for the core component. Additionally, to facilitate the user's ease-of-use of this core connector, an optional graphical user interface (GUI), can be deployed. This user interface is denoted as the "Container: Dataspace Connector GUI" and "Core Container GUI Pod" in the figures. The Dataspace Connector variant used the JavaScript-based VueJS[35] framework for the GUI component. However, similar to the core connector itself, there is no prescribed application framework from the IDSA for this.

The earlier section discussed that an IDS Connector should be deployed in an Application Container Management Layer that runs on top of an Operating System Layer. Therefore, each of the IDS Connector variants may utilize different deployment platforms to manage the containerized applications. The Dataspace Connector variant uses Docker for this purpose. As an alternative, the TSG variant recommends Kubernetes as it offers more features and controls for cluster management and continuous integration and deployment. One of the critical considerations to adopting this containerized applications approach, apart from ensuring a standardized and seamless application deployment, is to facilitate the core connector's orchestration of the IDS Data Apps in an IDS Connector environment. To coordinate the data flow between applications and support the execution of the IDS Data Apps, the Dataspace Connector Core may use an enterprise integration technology such as Apache Camel[36]. This technology is used for its message routing capability which works by connecting different nodes (also called processors) in a route definition. An alternative for such functionalities is to utilize an ESB technology, for which the eMagiz[37] platform can be a good fit. Having this part of the architecture answers to the requirements previously specified, such as the IR_1.6, OS_5, APS_5, APS_6, and OS_1. The Dataspace Connector variant recommends the utilization of Portainer[38] to support this functionality, although optional. The IDS Data App is denoted as the "Container: IDS App (0…n)" in Dataspace Connector's system architecture. Meanwhile, in TSG's implementation, it is encapsulated under the "App Container Pod". Each of these elements includes an "IDS API" layer (which can be based on either Java, Python, or other libraries) that is responsible for receiving and returning message flows from and to the core connector. Within the Application Logic layer of the Data App, the message flow will then be processed according to the functionalities that this app is designed for the core connector to use (e.g., data transformation, aggregation, analytics, etc.). To

[34] https://spring.io/
[35] https://vuejs.org/
[36] https://camel.apache.org/
[37] https://emagiz.com/
[38] https://www.portainer.io/

support the discovery and retrieval process of these apps, data users and owners should be able to access and query the Apps Store from their core connectors.

By the time this thesis is written, there are limited sources of information available that describe a visual system architecture of an IDS Data App. Most of the visual representations available (i.e., provided by the IDSA or described by TNO) are presenting the app's interaction with each other or with the IDS Connector, as depicted in **Figure 20** earlier, but not the internal elements of it (IDSA, 2021c). Therefore, **Figure 24** is presented to illustrate the essential internal structure and behavior of an IDS Data App. An IDS Data App should extend an IDS Connector's functionalities with additional data processing capabilities (e.g., data transformation, cleaning, aggregation, analysis, anonymization, etc.) (IDSA, 2019, 2021c). In this viewpoint, these data processing capabilities are depicted as application functions that are grouped under the Application Logic function. Note that an IDS Data App does not have to support all these functionalities at once, but it should support at least one of them. On top of the supported function, it has to provide an interface for other application components to invoke. There are several options for this purpose. SOAP protocol might be used. However, TNO, through their OpenAPI Data App[39] implementation, has chosen the REST protocol as the preferred approach. This API will then expose the service endpoints for the IDS Connector to access the functionality provided by this IDS Data App, before forwarding the message flows to the target enterprise system.

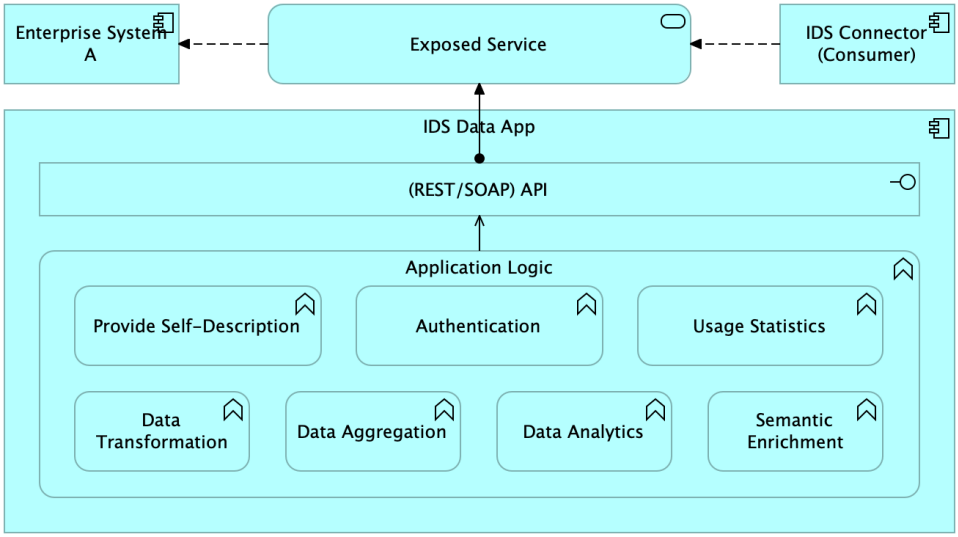


Figure 24 IDS Data App Internal System Architecture Viewpoint

[39] https://gitlab.com/tno-tsg/data-apps/openapi

Apart from the internal communication with the GUI and IDS Data Apps, the core connector should also provide API endpoints to enable external communication with other IDS components. This endpoint, in both variants' architecture, is denoted as the "IDS Endpoint", which interfaces this IDS Connector environment with the Metadata Broker, Clearing House, or IDS Connector of another stakeholder. Enabling such communication will, in turn, satisfy the requirements defined previously as the BRK_2 and BRK_3. Although the IDSA does not seem to specify a specific messaging framework for this purpose (e.g., REST API or SOAP), these two implementations preferred the former over the latter. Typically, this implementation of REST API includes the provisioning of the Swagger-UI for users to invoke several exposed functions through a simple graphical interface. In response to the requirements of IR_1.1, IR_1.2, and other related requirements (i.e., INF_1 to INF_5), one of the endpoints the core connector must expose is the function to provide its self-description. As a response to requirement IR_1, the core connector offers data to the other participants in the data space by connecting a company's internal digital infrastructure (which, in **Figure 23**, is depicted as the Backend/Enterprise System) with the external digital environment.

5.2.2. IDS Connector Data Model

Before the data can be offered to the data space and the actual data exchange can take place, a data owner needs to create the data and specify its metadata first. The IDSA prescribed in their GitHub repository a specific data model that an IDS Connector needs to comply with for creating, offering, requesting, and consuming data IDSA (2021b). **Figure 25** illustrates the mentioned Dataspace Connector's data model represented as ERD. This diagram complements the data model prescribed by the IDSA by providing more details on the underlying attributes of each entity and the multiplicity between entities. These attributes and multiplicities are analyzed and extracted from the reference implementation of the Dataspace Connector Core (IDSA, 2021d).

The data model begins with the Catalog entity, which can be used to categorize and group several Resources under it. A Resource itself serves as the root metadata that describes an offered data, or the requested one, by means of one or several Representations. A Representation annotates the offered or requested data with the media type (e.g., application/json for JSON, etc.), language (e.g., en_US, nl_NL, etc.), and standards (e.g., OTM, GS1 EPCIS, etc.) descriptions, in addition to the customary descriptive attributes such as title and creation date. Under a Representation, lies one or more Artifacts, which contain the metadata required for accessing the actual value of the offered or requested data. An Artifact object describes the accessURL (along with the username, password, and the possibility to be extended with an apiKey attribute if required) to retrieve or download the offered data. remoteID represents the ID of the offered data object used in its source system (i.e., enterprise system or database), distinguishing it from the ID that is generated by and used in the IDS Connector that offers this data. Other attributes in this entity are also provided for statistical and monitoring purposes, such as numAccessed and byteSize. Although, the numAccessed attribute can also be used as a counter to

enforce the data usage policy pattern of PP_5 (i.e., Restricted Number of Usages as described in **Table 17**).

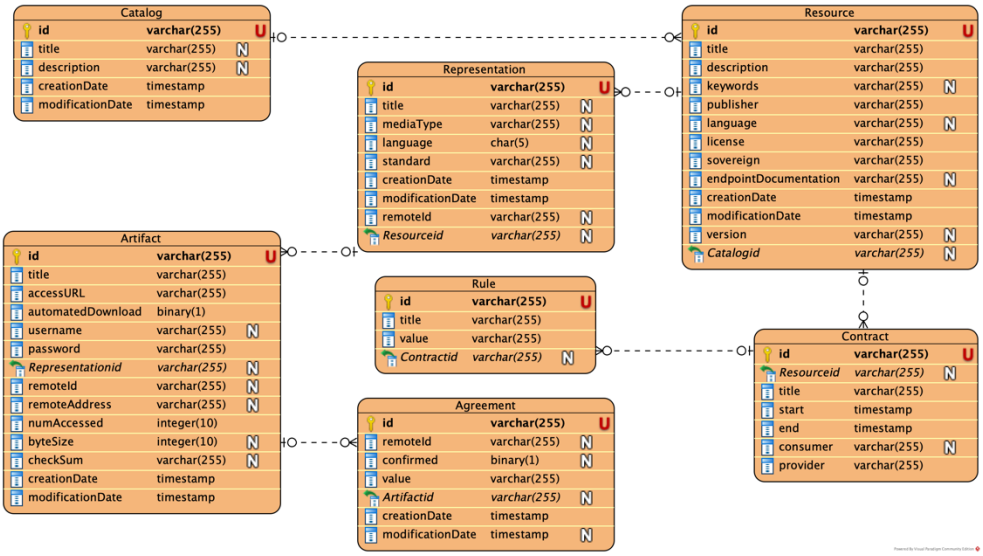


Figure 25 IDS Connector Data Model Represented in ERD (IDSA, 2021b)

On the other side of the model, there is the Contract entity. An object of it describes the start and the end date of an offered or requested resource, along with the ID of its consumer and provider. A Resource object can be associated with one or multiple Contracts, which describe several combinations of data usage Rules. The Rules here correspond to the policy patterns listed in **Table 17** that constrain the usage of the (offered and requested) data and enforce them in the IDS Connector of the data user. These objects describe the policy by containing the machine-readable pattern represented in a JSON-LD string under the attribute value. This JSON-LD string will be discussed more in the following sections. After data owners and users come to an agreement for using the requested data under the specified contract, an object of the Agreement will be instantiated containing the Contract object expressed in a JSON-LD format stored under the value attribute of the Agreement. Using this standardized (meta)data model of the IDS Connector, irrelevant to the format and structure of the data being transported, data owners and users can exchange data with each other in an IDS data space. In the next sections, with this data model in mind, the processes required to offer and request data, as well as to enforce the usage policy of the data will be discussed.

5.2.3. IDS Connector Resource Offering and Data Usage Policy Enforcement Process

One of the main value propositions of adopting the IDS principles for a data-sharing environment is the management of data sovereignty for all data owners and users through the enforcement of data usage policies. Therefore, the process required to manage this aspect needs to be investigated. **Table 16** indicates that the IDS Connectors must support the data owners when defining, attaching, and enforcing the data usage policy. As a means to address the requirements of USC_1, USC_2, and USC_3, the IDSA RAM explains how the data usage control can be defined by the data owner and enforced by the IDS Connector (IDSA, 2019). The process to define and enforce the data usage control is described as follows:

Step 1. At runtime, a Data Owner initiates the Data Offering activity by generating the data being offered and defining some metadata that describes it (e.g., title, description, owner, keywords, language, pricing, etc.).

Step 2. This Data Owner connects its IDS Connector with the endpoint of a data source (e.g., ERP, TPS, DBMS, etc.) and specifies the representations of the offered data (also known as media content type, e.g., application/json for JSON, application/xml for application/xml for XML, application/pdf for PDF, etc.).

Step 3. This Data Owner attaches data usage policy patterns to the offered data (e.g., Restricted Number of Usages, Use Data and Delete it After, etc., as listed in **Table 17**).

Step 4. This Data Owner specifies (i.e., creates a new one or chooses an existing one) the catalog that the data will be made available (**Note:** In the following **Section 3.3**, the details regarding these steps 1-4 will be further elaborated).

Step 5. Optionally, to facilitate resource and service discovery, this Data Owner may specify which Metadata Broker to publish the metadata of the data being offered.

Step 6. To discover the offered data in the data space, the Data User requests the URL of the Data Owner from the Metadata Broker.

Step 7. The Data User requests the data offered by the Data Owner by specifying on their IDS Connector the URL obtained from the Metadata Broker.

Step 8. Upon receiving a list of data offering artifacts/representations from the Data Owner, the Data User chooses one of the data representations to be requested.

Step 9. After receiving the policy from the Data Owner that binds the usage of the requested data, the Data User provides their acceptance of the Data Usage Policy, which then leads to the Data User receiving the requested data.

Step 10. Thereupon, the IDS Connector on both the Data Owner and Data User sides will continuously control the way data is processed, aggregated, or forwarded to other endpoints and prevent data to be handled in an undesired way (e.g., by forwarding personal data to the public endpoints).

To present this process of defining, attaching, accepting, and enforcing the data usage policy in a visualized manner, a UML's Sequence Diagram is illustrated in **Figure 26** and **Figure 27**. In most cases, a Data User will be the actor that starts a data request and data usage policy negotiation process as it has the intention to use the requested data as depicted in **Figure 26**. While this would be the common sequence, one could think of another scenario, such as where a Data Owner is in demand for a Service Provider to analyze its data set. For such purpose, the Owner needs to initiate the policy negotiation process itself and not the Data User. This scenario is exemplified by **Figure 27** which shows the role of the Data User, currently being represented as a Service Provider, who offers the service capability to process a particular dataset.

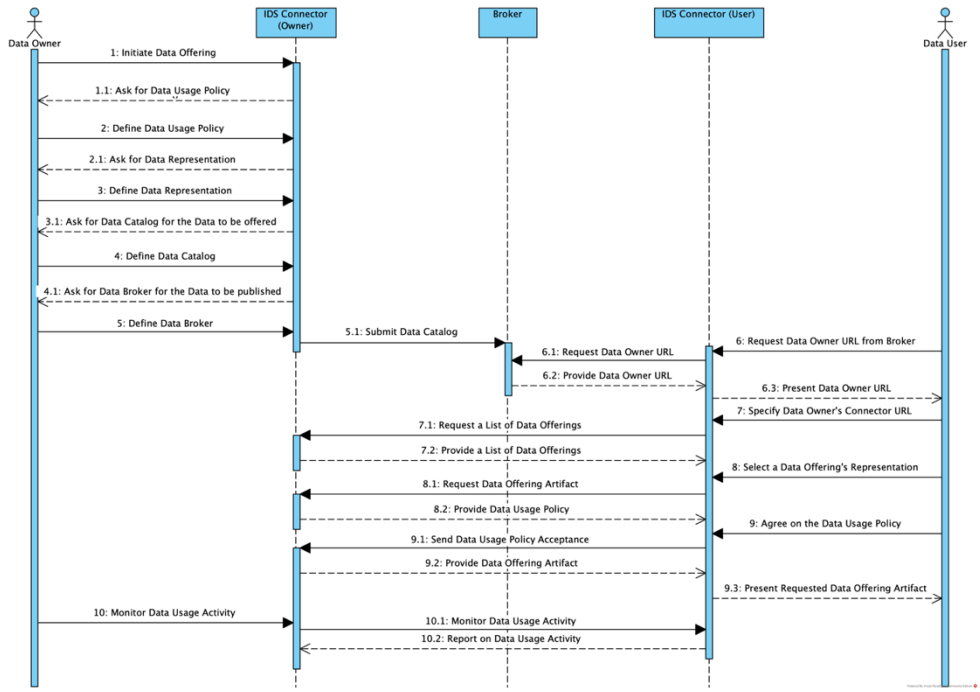


Figure 26 Data Usage Policy Definition and Enforcement between Data Owner and Data User

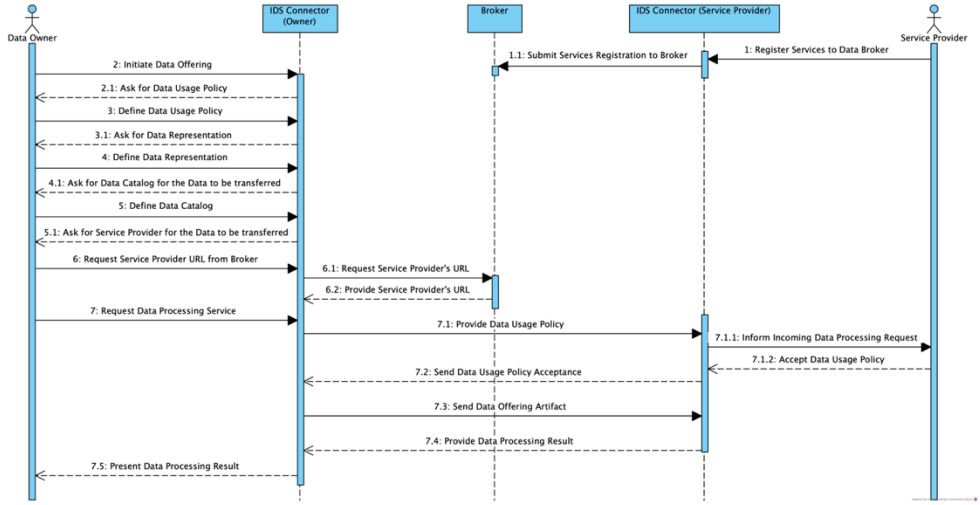


Figure 27 Data Usage Policy Definition and Enforcement Service Provider between Data Owner and Service Provider

IDS Connector Resource Offering Creation Process

The IDSA did not specify any reference processes in their documents for creating a new resource through an IDS Connector. However, they have provided a guide for creating a resource using their Dataspace Connector Core reference implementation (IDSA, 2021d), which adheres to the IDS Connector Data Model and **Figure 26** (IDSA, 2021b). This Dataspace Connector Core refers to the reference implementation presented in **Figure 21**. Based on their guide, the process involved for this purpose is listed as follows:

1. **Create a Resource**

   An IDS Connector distinguishes two kinds of Resources, which are Offered Resources, and Requested Resources. In this case, the data owner creates an Offered Resource. This step of defining the title, description, owner, and other metadata that describe the offered resource corresponds with Step 1 in **Sub-Section 5.2.3**. Based on the reference implementation (IDSA, 2021d), an example JSON snippet for creating a new resource is shown below:

Table 20 JSON Snippet - Creating a New Resource

URL	https://localhost:8080/api/offers

JSON	{
	"title": "DWD",
	"description": "Weather data",
	"keywords": ["string"],
	"publisher": "DWD",
	"language": "DE",
	"licence": "https://www.dwd.de/DE/leistungen/opendata/faqs_opendata.html",
	"sovereign": "DWD",
	"endpointDocumentation": "none"
	}

2. Create an Artifact

The artifact acts as the entity that has a 1:1 relation to the raw offered data. It describes the raw data's title, byte size, and URL to access the offered data. Together with steps 3-5 below, this step further details the process specified by Step 2 in **Sub-Section 5.2.3**. Based on the reference implementation (IDSA, 2021d), an example JSON snippet for creating a new resource is shown below. accessURL represents the endpoint's URL of the data source. As illustrated in **Figure 26**, additional attributes (e.g., username and password) can be added if the endpoint requires them to access the data.

Table 21 JSON Snippet - Creating a New Artifact

URL	https://localhost:8080/api/artifacts

JSON	{
	"title": "string",
	"accessUrl": "https://maps.dwd.de/geoserver/dwd/ows
	?service=WFS&version=1.0.0
	&request=GetFeature
	&typeName=dwd%3AWarnungen_Gemeinden
	&outputFormat=application%2Fjson",
	"automatedDownload": true
	}

3. Create a Representation

An instance of the Representation element represents the materialization of the (offered) resource in one or several representation instances. An instance of this entity entails the media content type, language, standard, etc., of the offered resource. This allows a single resource to be described and offered in several representations, for instance, in several media types (e.g., in JSON, XML, CSV, etc.) or in several languages (e.g., DE, EN, NL, etc.). An example JSON snippet for creating a new representation is shown below.

Table 22 JSON Snippet - Creating a New Representation

URL	https://localhost:8080/api/representations

JSON	{
	"title": "DWD",
	"mediaType": "application/json",
	"language": "DE",
	"standard": "???"
	}

4. Add the Artifact to the Representation

The next step is to append the newly created artifact to that representation. In the JSON snippet below, the '$representationId' indicates the UUID of

the target representation (e.g., "f62f8412-da63-4873-ae4f-e3579d9720b6"), and the "$artifactId" indicates the URI of the target artifact wrapped inside a JSON array bracket (e.g., "http://provider:8080/api/artifacts/f62f8412-da63-4873-ae4f-e3579d9720b6").

Table 23 JSON Snippet - Adding an Artifact to the Representation

URL	https://localhost:8080/api/representations/'$representationId'/artifacts
JSON	["$artifactId"]

5. Add Representation to the Resource

The following is to append the previous representation to the (offered) resource. In the JSON snippet below, the '$resourceId' indicates the UUID of the target resource (e.g., "f62f8412-da63-4873-ae4f-e3579d9720b6"), and the "$representationId" indicates the URI of the target representation wrapped inside a JSON array bracket (e.g., "http://provider:8080/api/representations/f62f8412-da63-4873-ae4f-e3579d9720b6").

Table 24 JSON Snippet - Adding a Representation to the Offered Resource

URL	https://localhost:8080/api/offers/'$resourceId'/representations
JSON	["$representationId"]

6. Create a Rule

In regard to the definition of the offered resource's usage control referring to Step 3 in **Sub-Section 5.2.3** earlier, the data model prescribes the instantiation of an object from the Rule entity. This rule, or multiple rules, will later be appended to a single contract that binds the data user and the data owner with an agreed set of rules. An instance of rule represents a machine-readable format of the data usage policy pattern that refers to the list in **Table 17**. This pattern, as shown in **Table 25**, is represented as the JSON string format inside the "value" attribute.

Table 25 JSON Snippet - Creating a New Rule

URL	https://localhost:8080/api/rule
JSON	{ "value": "{ \"@type\": \"ids:Permission\", \"@id\": \"https://w3id.org/idsa/autogen/permission/cf1cb758-b96d-4486-b0a7-f3ac0e289588\", \"ids:action\": [{ \"@id\": \"idsc:USE\" }], \"ids:description\": [(Markus & Bui, p. value)], \"ids:title\": [(Markus & Bui, p. value)] }" }

7. Create a Contract

Next to creating the rules, a contract needs to be instantiated to specify who is the owner (or the provider) of the data, who will use (or consume) it, and what is the time period that this contract will be valid. An example JSON snippet for creating a new contract is shown below.

Table 26 JSON Snippet - Creating a New Contract

URL	https://localhost:8080/api/contract

JSON	{
	"title": "DWD Contract",
	"start": "2021-05-19T10:09:59.563Z",
	"end": "2025-05-19T10:09:59.563Z"
	}

8. Add Rule to the Contract

To bind the data user and the data owner with an agreed set of rules, the previously defined rules will need to be appended to the newly created contract. In the JSON snippet below, the '$contractId' indicates the UUID of the target contract (e.g., "f62f8412-da63-4873-ae4f-e3579d9720b6"), and the '$ruleId' indicates the URI of the target rule wrapped inside a JSON array bracket (e.g., "http://provider:8080/api/representations/f62f8412-da63-4873-ae4f-e3579d9720b6"). Using this array format, multiple rules can be appended to attach multiple policy patterns to a single contract.

Table 27 JSON Snippet - Adding a Rule to the Contract

URL	https://localhost:8080/api/contracts/'$contractId'/rules
JSON	["$ruleId"]

9. Add Contract to the Resource

Subsequently, this contract has to be attached to the resource before it is offered in a catalog. The process follows a similar approach to the previous process. In the JSON snippet below, the '$resourceId' indicates the UUID of the target representation (e.g., "f62f8412-da63-4873-ae4f-e3579d9720b6"), and the "$contractId" indicates the URI of the target artifact wrapped inside a JSON array bracket (e.g., "http://provider:8080/api/artifacts/f62f8412-da63-4873-ae4f-e3579d9720b6").

Table 28 JSON Snippet - Adding a Contract to the Offered Resource

URL	https://localhost:8080/api/offers/'$resourceId'/contracts
JSON	["$contractId"]

10. Create a Catalog

Catalogs are the root elements that contain resources. It is the entity that will be exposed in an IDS Connector's self-description to indicate the catalog of data that this connector offers. Based on the reference implementation (IDSA, 2021d), an example JSON snippet for creating a new resource is shown below:

Table 29 JSON Snippet - Creating a New Catalog

URL	https://localhost:8080/api/catalogs

JSON	{
	"title": "DWD Catalog",
	"description": "A catalog with DWD resources"
	}

11. Add the Resource to the Catalog

Once the offered resource is created and a catalog is available, the next step is to append the newly created resource to that catalog. In the JSON snippet below, the '$catalogId' indicates the UUID of the target catalog (e.g., "f62f8412-da63-4873-ae4f-e3579d9720b6"), and the "$resourceId" indicates the URI of the target offered resource wrapped inside a JSON array bracket (e.g., "http://provider:8080/api/offers/f62f8412-da63-4873-ae4f-e3579d9720b6").

Table 30 JSON Snippet - Adding a Resource to the Catalog

URL	https://localhost:8080/api/catalogs/'$catalogId'/offers
JSON	["$resourceId"]

12. Publish Resource to a Metadata Broker

Upon adding the resource to a catalog, the data owner may publish this resource to a Metadata Broker to support its discoverability. So far, there are not many sources of information available that specify how an IDS Connector should perform this on the technical level (i.e., what kind of message does the Metadata Broker accept and the IDS Connector send, and

to which endpoint). The German research institute Fraunhofer ISST provided an example implementation[40]. From this example, it can be observed that the "https://broker.ids.isst.fraunhofer.de/infrastructure" endpoint is provided for this purpose. The broker itself follows Linked Data principles, as its data management approach, and utilizes a triple store as its data persistence technology (more will be discussed in the next chapter). Based on this, IDS Connectors in the data space can query the broker using SPARQL query string through the payload field of this endpoint. There is no sample SPARQL query to be found yet for publishing a resource to a broker. Therefore, **Table 31** provides a (generic) sample query that can be used to retrieve metadata of registered IDS Connectors and their offered resources, with a note that this sample query is yet to be validated.

Table 31 JSON Snippet - Metadata Broker Sample SPARQL Query

URL	https://localhost:8085/infrastructure

JSON	"
	SELECT ?subject ?predicate ?object
	WHERE {
	?subject ?predicate ?object
	}
	"

5.2.4. IDS Connector Resource Request Process

Once a data owner offered its data through its IDS Connector and submitted the metadata to a Metadata Broker, data users can begin to request the data. According to the guide on IDSA's GitHub page, the process serving this purpose is as follows:

1. **Request Provider Connector URL from Metadata Broker**

   First, data users may need to find the data they want to request, either identified by their Universal Unique Identifier (UUID) or other contextual attributes such as what is the data about or the adopted standards. For this purpose, data users can go to a Metadata Broker that publishes (1) the metadata of the offered resources and (2) the metadata of the Provider Connectors (i.e., IDS Connector that offers the resource). Within this metadata, the Metadata Broker includes an API endpoint, called the accessURL, that is used to retrieve the self-description of the Provider Connector along with the list of the resource catalog. The IDSA prescribes a convention for the path of this API endpoint as `POST <provider_connector_base_url>/api/ids/data` (IDSA, 2021d).

2. **Request Provider Connector Self-Description**

[40] https://app.swaggerhub.com/apis/idsa/IDS-Broker/1.3.1

To request the self-description of the Provider Connector, data users need to invoke the endpoint `POST <consumer_connector_base_url>/api/ids/description` on their Consumer Connector (i.e., IDS Connector that consumes the resource). Then, data users specify the accessURL of the Provider Connector as the value for the recipient URL parameter to retrieve the connector's self-description. Within this response, data users can find contextual information about the Provider Connector, for instance, the maintainer of the IDS Connector, the security profile of the IDS Connector, as well as the resource catalogs that their elements are listed.

3. **Request Provider Connector Resource Catalog Details**

To request the details of a particular catalog, data users can use the same endpoint `POST <consumer_connector_base_url>/api/ids/description`. This time, in addition to the Provider Connector's accessURL, data users fill in an additional parameter called the element ID with the URI value of the target catalog. The response returned from this operation contains a list of the offered resource under this target catalog, including the representations of each resource, as well as the artifacts of each representation.

Note: The presented artifacts in this stage do not yet expose the accessURL to access the offered data. However, important things to note in this response are the **resource id** and the **artifact id** of the resource data users want to consume, for instance:

Resource Id:

`http://provider:8080/api/offers/22a6d425-402b-4f65-9614-a03502c772be`

Artifact Id:

`http://provider:8080/api/artifacts/c0041c69-9515-4d54-8914-16380b80f236`

These identifiers will be used by the data users in the next step to negotiate a contract for consuming the artifact (IDSA, 2021d).

5.2.5. IDS Connector Contract Negotiation Process

Upon discovering the identifier of the resource's artifact to be requested, the data users may proceed to negotiate a contract for consuming it. According to the guide on IDSA's GitHub page, the process serving this purpose is as follows:

1. **Request a Contract**

To initiate the contract negotiation process, data users first need to request the contract for the requested resource. To do this, data users use the endpoint `POST <consumer_connector_base_url> /api/ids/contract` in their connector acting as the consumer. This endpoint asks for several parameters, namely:

a. **Recipient**

   To be filled in with the accessURL of the Provider Connector (e.g., http://provider:8080/api/ids/data).

b. **ResourceIds**

   To be filled in with the Id of the requested resource as specified earlier (e.g., http://provider:8080/api/offers/22a6d425-402b-4f65-9614-a03502c772be).

c. **ArtifactIds**

   To be filled in with the Id of the artifact to be consumed as specified earlier (e.g., http://provider:8080/api/artifacts/c0041c69-9515-4d54-8914-16380b80f236).

d. **Download**

   This parameter indicates the Boolean of whether the connector should automatically download the artifact's data (e.g., `true` or `false`).

2. **Specify a Contract Offer**

Next, this endpoint also asks for a request body to be filled in with a contract offer that must match the one that the resource was created with (IDSA, 2021d). This instance of the Contract Offer complies with the rules that were associated with the contract of the offered resource. Hence, the (JSON) value for this contract offer resembles the value of the machine-readable format of the data usage policy pattern used by the data owner. **Table 32** provides an example contract offer that corresponds to the data usage policy pattern used in **Table 25**.

Table 32 JSON Snippet - Specifying a Contract Offer

URL	`https://localhost:8081/api/ids/contract`

JSON	`[ {`
	`  "@type" : "ids:Permission",`
	`  "@id" : "http://provider:8080/api/rules/5041c3e5-4933-419d-87d9-474b98ced678",`
	`  "ids:description" : [ {`
	`  "@value" : "provide-access",`
	`  "@type" : "http://www.w3.org/2001/XMLSchema#string"`
	`  } ],`
	`  "ids:title" : [ {`

```
      "@value" : "Example Usage Policy",
      "@type" : "http://www.w3.org/2001/XMLSchema#string"
    } ],
    "ids:action" : [ {
      "@id" : "idsc:USE"
    }],
    "ids:target" : "http://provider:8080/api/artifacts/c0041c69-9515-
4d54-8914-16380b80f236"
} ]
```


Note: Within the request body for the contract offer, the field **ids:target** needs to contain the same artifact id that is specified in the ArtifactIds request parameter.

Subsequently, data users will receive a response for the resulting Contract Agreement as listed in **Table 33** after they execute this operation.

Table 33 JSON Response - Resulting Contract Agreement

URL	https://localhost:8081/api/ids/contract

JSON	{
	"creationDate": "2021-07-06T05:54:54.980+0000",
	"modificationDate": "2021-07-06T05:54:54.980+0000",
	"remoteId": "http://provider:8080/api/agreements/7a78b3e6-9efd-40f8-9427-b03743ec4980",
	"confirmed": true,
	"value": "{\n \"@context\" : {\n \"ids\" : \"https://w3id.org/idsa/core/\"\n },\n \"@type\" : \"ids:ContractAgreement\",\n \"@id\" : \"http://provider:8080/api/agreements/7a78b3e6-9efd-40f8-9427-b03743ec4980\",\n \"ids:prohibition\" : [ ],\n \"ids:obligation\" : [ ],\n \"ids:contractStart\" : {\n \"@value\" : \"2021-07-06T05:54:54.172Z\",\n \"@type\" : \"http://www.w3.org/2001/XMLSchema#dateTimeStamp\"\n },\n \"ids:contractDate\" : {\n \"@value\" : \"2021-07-06T05:54:54.171Z\",\n \"@type\" : \"http://www.w3.org/2001/XMLSchema#dateTimeStamp\"\n },\n \"ids:provider\" : {\n \"@id\" : \"https://w3id.org/idsa/autogen/baseConnector/provider\"\n },\n \"ids:permission\" : [{\n \"@type\" : \"ids:Permission\",\n \"@id\" : \"http://provider:8080/api/rules/5041c3e5-4933-419d-87d9-474b98ced678\",\n \"ids:action\" : [{\n \"@id\" : \"idsc:USE\"\n }],\n \"ids:description\" : [{\n \"@value\" : \"provide-access\",\n \"@type\" : \"http://www.w3.org/2001/XMLSchema#string\"\n }],\n \"ids:assigner\" : [{\n \"@id\" : \"https://w3id.org/idsa/autogen/baseConnector/provider\"\n }],\n \"ids:title\" : [{\n \"@value\" : \"Example Usage Policy\",\n \"@type\" : \"http://www.w3.org/2001/XMLSchema#string\"\n }],\n \"ids:assignee\" : [{\n \"@id\" : \"https://w3id.org/idsa/autogen/baseConnector/consumer\"\n }],\n \"ids:target\" : {\n \"@id\" : \"http://provider:8080/api/artifacts/563ddf04-43ee-4eae-9634-16b70cc65ca7\"\n }\n }],\n \"ids:consumer\" : {\n \"@id\" : \"https://w3id.org/idsa/autogen/baseConnector/consumer\"\n }\n}",
	"_links": {
	"self": {
	"href": "http://localhost:8081/api/agreements/f1acf392-8c87-49ea-afa5-4c18b2751f07"
	},
	"artifacts": {
	"href": "http://localhost:8081/api/agreements/f1acf392-8c87-49ea-afa5-4c18b2751f07/artifacts{?page,size}",
	"templated": true}}}

5.2.6. IDS Connector Artifact Consumption Process

Once the contract negotiation is done as described above, the link to the artifact to be consumed can be found in the resulting Contract Agreement response. In the example provided above, the "href" link under the "artifacts" object can be used to retrieve the artifact associated with the negotiated agreement. Hence, by accessing this (example) link http://localhost:8081/api/agreements/f1acf392-8c87-49ea-afa5-4c18b2751f07/artifacts (in a browser for instance), then data users will receive a response resembling the one listed below.

Table 34 JSON Response - Artifact of a Contract Agreement

URL	https://localhost:8081/api/ids/contract

JSON	{
	"_embedded": {
	"artifacts": [
	{
	"creationDate": "2021-07-06T05:54:55.364+0000",
	"modificationDate": "2021-07-06T05:56:43.034+0000",
	"remoteId": "http://provider:8080/api/artifacts/563ddf04-43ee-4eae-9634-16b70cc65ca7",
	"title": "string","numAccessed": 1,
	"byteSize": 475752,"checkSum": 0,
	"additional": {
	"ids:byteSize": "0", "ids:checkSum": "0",
	"ids:creationDate": "2021-07-05T13:50:22.717Z"
	},
	"_links": {
	"self": {
	"href": "http://localhost:8081/api/artifacts/3063b06d-56ed-4d3f-9651-1a526c9c1b3d"
	},
	"data": {
	"href": "http://localhost:8081/api/artifacts/3063b06d-56ed-4d3f-9651-1a526c9c1b3d/data"
	},
	"representations": {
	"href": "http://localhost:8081/api/artifacts/3063b06d-56ed-4d3f-9651-1a526c9c1b3d/representations{?page,size}",
	"templated": true
	},
	"agreements": {
	"href": "http://localhost:8081/api/artifacts/3063b06d-56ed-4d3f-9651-1a526c9c1b3d/agreements{?page,size}",
	"templated": true
	}}}]
	},
	"_links": {
	"self": {
	"href": "http://localhost:8081/api/agreements/f1acf392-8c87-49ea-afa5-4c18b2751f07/artifacts?page=0&size=30"
	}
	},
	"page": {
	...
	}}

Upon receiving this response, data users can use the "href" link that ends with /data listed under the "data" object to consume the artifact. This way, the data value returned from the artifact's accessURL will be downloaded or shown to the data users.

5.3. Summary and Conclusion

This chapter explains how an IDS Connector and IDS Data Apps can be designed according to the IDSA specification to facilitate data interoperability and sovereignty. The design process is done by referring to several documents and reference implementations from the IDSA. From the requirement elicitation and architecture design processes, this chapter also covers the SRS and produced system architectures that visualize these two essential application components in accordance with the IDSA specification in particular, we focus on how the IDS Connector can be designed to orchestrate the IDS Data Apps, and also on how the IDS Data Apps should function based on their profiles and types. Specifications related to how the IDS Connector can enforce the data usage control are also presented. These system specifications and architectures serve as a guideline to develop IDS Connector and IDS Data Apps prototypes. Then, these prototypes will be implemented in several business cases to be evaluated their contribution to managing data interoperability and sovereignty. In the next chapter, we discuss how a Connector Store can be designed to support the discovery and selection of these components along with the offered resources, which completes the **Treatment Design** phase of our design methodology.

6 The Design of a Connector Store for a Logistics Data Space

This chapter is presented as part of the endeavor to investigate a suitable design of an application component responsible to support the discoverability of the IDS Connectors, data sources, and participants active in a logistics data space. Following the same approach in the previous chapter, we initiate the design process of a Connector Store by eliciting its requirements before constructing its architectural viewpoint. Thereupon, the produced design can be used to develop its prototype as part of the **Treatment Validation** phase in the next chapter. Therefore, **Section 6.1** triggers the discussion by describing the business role in the IDS that is relevant to the Connector Store. Next, **Section 6.2** discusses the development of the mechanism to support semantic discovery and selection of IDS Connectors for an IDS ecosystem. Lastly, **Section 6.3** concludes the chapter by describing the next step to instantiate the prototypes based on this design and validate their contribution towards the defined goals.

6.1. The Role of a Broker Service Provider in International Data Spaces: An Enterprise Architecture Viewpoint

According to the IDS RAM, a Broker Service Provider (BSP) refers to an intermediary entity that registers, publishes, and supports the search for metadata about data sources and services available in an IDS ecosystem (IDSA, 2019). A BSP adds value to a data space by providing services to leverage the discoverability of IDS Connectors and resources offered by participants (Bader, Bruckner, et al., 2020). It is necessary to have at least one BSP operating per business domain (e.g., the logistics sector). Thus, multiple BSPs could also simultaneously serve as a cross-domain application. The IDS RAM described that a BSP may also assume other business roles (e.g., a Clearing House that is responsible for keeping logs of all activities related to data exchange in an IDS ecosystem). However, the BSP's responsibilities are limited to supporting data users and owners with the management of the metadata about a particular resource or service. Therefore, the direct data exchange and usage negotiation processes involving only data users and owners are not part of the responsibilities of a BSP.

Figure 28 depicts an ArchiMate model of a BSP, specifying how this entity interacts with other actors and components of an IDS ecosystem through an IDS Metadata Broker. The architecture conforms to the technical specifications of the IDS RAM and the IDS Metadata Broker component (IDSA, 2019; IDSA Certification Working Group, 2021). The BSP (colored orange) is a business role that develops, hosts, and maintains the IDS Metadata Broker. The IDS Metadata Broker acts as a metadata repository that exposes GUIs and APIs to facilitate metadata publication services.

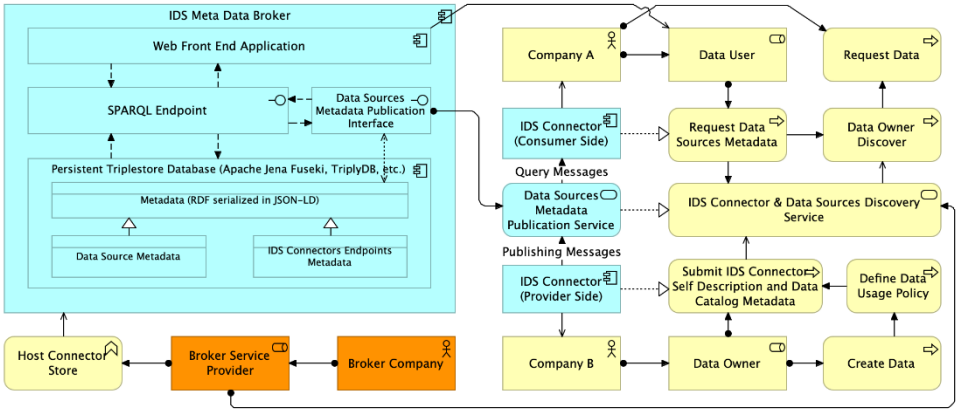


Figure 28 Broker Service Provider Ecosystem and System Architecture Viewpoint

To carry out its metadata management responsibility, a BSP must provide an interface for data owners to publish their metadata, including descriptions of their IDS Connectors and the data catalogs accessible through those connectors. The metadata can be stored in the BSP's internal repository and made available for structured queries submitted by the data user. In addition to supporting the data users in retrieving the metadata of participating IDS Connectors or cataloged data resources, the IDS Metadata Broker should also help the data owners register, update, or remove metadata entries (IDSA Certification Working Group, 2021). Additionally, it should provide an interface describing additional information about its functionalities and indexing services, such as supported query languages, available add-on services, and their data endpoints.

By hosting the IDS Metadata Broker, the BSP offers its service to the data space to support data users in finding and discovering IDS Connectors and data sources provided by the data owners. Two processes must take place for this service to deliver its full potential. First, before enacting any data or metadata exchange, the data owners and users should already be in control of a certified IDS Connector. Acquiring the so-called IDS-ready labels for software components is one of the requirements for business actors to participate in an IDS ecosystem after being approved on the organizational level (IDSA, 2019, 2020c). Secondly, the data owners could submit the self-description and the metadata describing the data used by their data connectors to the IDS metadata broker via the exposed interface based on a standardized protocol (e.g., REST API, OpenAPI 3.0, etc.) (Bader, Bruckner, et al., 2020; IDSA, 2020b; IDSA Certification Working Group, 2021). This process occurs after they create the data and define their data usage policies. Next, the data users could discover these catalogs by browsing the IDS metadata broker's metadata based on contextual information (e.g., keywords, language, usage policies, maintainer, etc.). Finally, the data users could receive the information required to access the data owner's IDS Connector to request the desired data.

6.2. Semantic Discovery and Selection of IDS Connectors in International Data Spaces

The metadata broker specification document states that different Metadata Broker implementations may be developed and made available by various providers in International Data Spaces (Bader, Bruckner, et al., 2020). This thesis proposes the Connector Store as an extension of the IDS Metadata Broker by providing additional functionality to support the semantic discovery and selection of IDS Connectors (Firdausy et al., 2022c). It aims, therefore, to help data owners and users discover and select the connectors that are most suitable for their needs and capabilities based on information about the context in which the connectors could operate.

Such functionalities of an IDS Metadata Broker are prescribed by the IDSA to be supported by the adoption of Semantic Web and Linked Data technology. This technology has been implemented in an increasing variety of contexts in recent years to enhance the discoverability and accessibility of resources on the Web (Janowicz et al., 2015). One of the building blocks that constitute the Semantic Web is the Ontology, which is a formal and explicit specification of a concept that works by adding a layer of metadata to the described resources (Salma et al., 2019). This procedure makes the Web more accessible and understandable for more refined search results by software agents in providing information to human agents. Therefore, to facilitate the discoverability and selection process of the IDS Connectors, the development of the Connector Store should incorporate the Semantic Web technology that begins with the development of an Ontology.

6.2.1. Ontology Development and Requirements Specification Methodological Guidelines

The first step in developing an ontology-based software application is the formulation of the Ontology Requirements Specification Document (ORSD). In this thesis, we adopt scenario-based NeON Methodology, which emphasizes the reuse of existing ontological and non-ontological resources in developing the ontology (Gómez-Pérez & Suárez-Figueroa, 2009). In addition to the requirements specification activity guidelines, this methodology also provides a template to formulate the ORSD as a filling card that describes the purpose, scope, implementation language, intended end-user, intended uses, requirements, and pre-glossary terms of the ontology under design (Suárez-Figueroa et al., 2009).

To maintain interoperability with the domain reference ontology, NeON suggests a quick search of knowledge resources for possible reuse during development. For this purpose, the IDSA has published the IDS IM that describes the fundamental concepts of the IDS, covering entities from the participants to the infrastructure components (IDSA, 2021f). This IDS IM grounds the ontology proposed in this work. The resulting conceptual model is depicted in OntoUML (Guizzardi, 2005). This model serves as the basis for further implementation into OWL to describe the IDS

Connectors and distinguish them with the subject-predicate-object triples according to the RDF format (Berners-Lee et al., 2001). Through this semantic annotation, several sentences can be formed to explain the IDS Connectors. For instance, company A maintains an IDS Connector X, IDS Connector Y is offered in a flat-rate pricing model, or IDS Connector Y complies with GS1 standards. As a result, software agents will be able to discover the IDS Connectors that are appropriate to their data exchange demands.

6.2.2. Connector Store Ontology Requirement Specification

The requirements specification identifies the ontology's purpose, scope, and implementation language. **Table 35** presents the three main end-users that will take advantage of the knowledge given by the Connector Store ontology. The business representatives are the first target users due to their interest in spotting potential business opportunities in the current business landscape. For the potential IDS participants, the presence of their partners and the prospect of securing a strategic partnership with other existing participants signal the value of participating in the data space. Such a scenario might influence their willingness to consolidate into the IDS ecosystem.

Table 35 Connector Store Ontology Requirements Specification Document

Purpose	
To describe IDS Connectors for potential participants of an IDS ecosystem	
Scope	
Contextual information about the business ecosystem where the data connector will operate, e.g., business domain, pricing model, and enforced data access policy	
Implementation Language	
The ontology is represented in OntoUML, with further translation into OWL.	
Intended End-Users	
User 1.	Business representatives of potential and existing IDS participants
User 2.	IT representatives of potential and current IDS participants
User 3.	Software and service providers who develop and supply IDS Connectors
User 4.	Scholars exploring the ontology's knowledge representation capabilities
Intended Uses	
Use 1.	Software and service providers publish their offered data connectors' metadata on the IDS Connector Store to make their data connectors discoverable.
Use 2.	Business representatives search for IDS-compliant partners operating in the same business domain, complying with common standards, etc.
Use 3.	IT representatives search for data connectors that match their needs and capabilities.
Use 4.	Scholars search and import the ontology into their IDS proof-of-concept tools.
Ontology Requirements **Non-Functional Requirements**	
NFR 1.	The ontology must at least use English.
NFR 2.	The ontology must comply, reuse, and integrate with the existing IDS Ontology specified under the IDS IM.
Functional Requirements: Competency Questions	
CQ 1.	*What software provider offers IDS Connectors?*
CQ 2.	*Which IDS Connectors are developed for a specific business domain?*

CQ 3.	*Which IDS Connectors are complying with a particular standard?*
CQ 4.	*Which IDS Connectors are offered in this pricing model?*
CQ 5.	*Which IDS Connectors support these data usage agreements?*
CQ 6.	*Which IDS Connectors were developed in which development framework?*
CQ 7.	*Which IDS Connectors are offered in this deployment context?*
CQ 8.	*Which IDS actors use a particular IDS Connector from a specific software provider?*
CQ 9.	*Which IDS actors operate in a particular business domain?*
CQ 10.	*Which IDS actors comply with a particular standard?*
Terms from Competency Questions & Frequency	
Business Domain, Data usage agreement, deployment, IDS Connector, participant, pricing mode, software provider, standards, technology	
Objects and Terms for Answers	
- Gatewise IDS Connector, Supplydrive IDS Connect-or, TradeCloud IDS Connector;	
- Transport Logistics, Glass Manufacturing, Steel Manufacturing;	
- Delete After Interval, Connector-restricted Agreement, Logging Agreement;	
- Vandaglas B.V., Van Egmond Groep, Meijer Metal;	
- ECI Software Solutions, Tradecloud, OTM, GS1, EDI4STEEL;	
- Flat Rate, Freemium, Pay per User, Pay per Feature;	
- Java, Spring Boot, JavaScript, NodeJS, VueJS, Python, On-Premise, cloud SaaS.	

Conversely, the interests of the existing participants can take many forms. One example is to find other prospective partners to engage in strategic information exchange to leverage their value chain performance. The IT representatives will further translate these business strategies into IT implementation strategies by investigating the suitability of the IDS Connector that matches their needs and capabilities. Such a demand leads to concerns about which IDS Connectors fit their business domain or adopted industrial standards for data exchange. In response, software and service providers will be interested in making their IDS Connectors discoverable by external software applications. Additionally, frequent terms are extracted from the CQs, leading to the enumeration of objects for answering the end user's query. We instantiate the entities listed in **Table 35** by referring to the literature (Bol Raap et al., 2016; Lopes-Martínez et al., 2018), industrial standards (GS1, 2021; INAD Industrie Software B.V, 2022; OpenTripModel, 2021), IDSA documentation and publications (IDSA, 2019, 2021f, 2022), and the publication of the SCSN, one of the IDS forerunners in the Dutch manufacturing supply chain (Stolwijk & Berkers, 2020; TNO, 2020).

6.2.3. Preliminary Connector Store Ontology Conceptual Model

Using the IDS RAM and the IDS IM as a starting point, we identified several concepts relevant to answering the CQs above, namely the Participant and the Connector concepts (IDSA, 2019, 2021f). As shown in **Figure 29**, we identify the concept of the Participant as the IDS Actor and extend it further into two specializations. The Core IDS Actors refer to the participants who either own and provide or request and use data. Whereas the IDS Supporting Actors are associated with parties that ensure the continuation of the data-sharing ecosystem. The software and service provider carries out its duty by providing essential application components for participating in the data space. Meanwhile, the Broker Service

Provider supports the core actors with the function to look up the other actors as well as their IDS Connectors through the functionality offered by the Connector Store. In addition, the Supporting IDS Actor also covers other roles, such as the Clearing House and Identity Provider. However, as the IDS RAM describes, these roles can be assumed by the same organization that takes the part of the Broker Service Provider (IDSA, 2019).

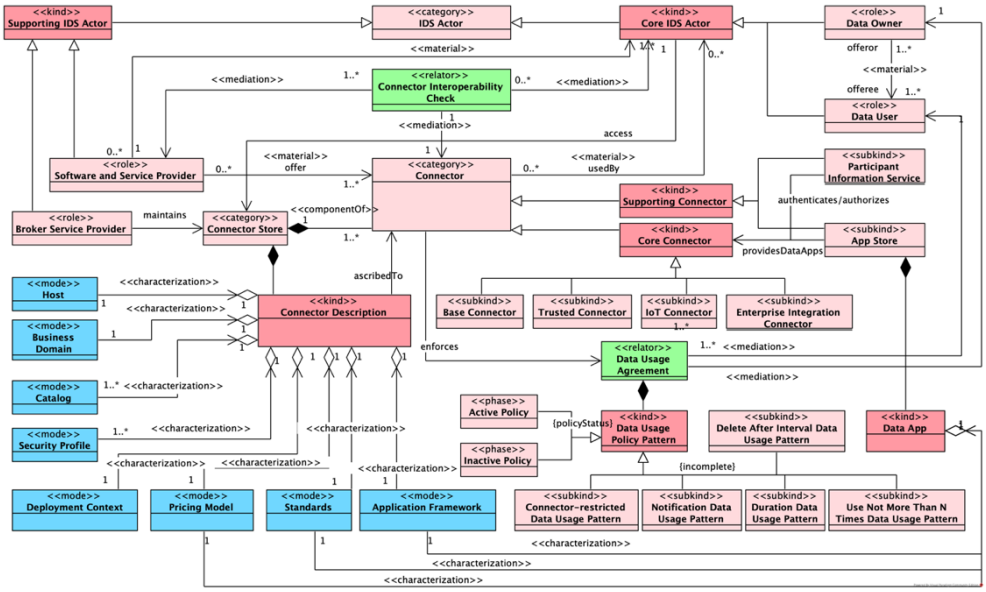


Figure 29 Preliminary Connector Store Ontology Conceptual Model

The IDSA expresses the IDS Connectors from several different perspectives. On the one hand, the IDS IM describes the concept of a Connector to be the generalization of the Base Connector, Trusted Connector, App Store, and Participant Information Service (ParIS) (IDSA, 2021f). Here, we distinguish these Connectors into the Core Connector and Supporting Connector, each used by the corresponding type of role. The IDS RAM justifies this distinction by describing the functions of the supporting category (i.e., the App Store Provider, Broker Service Provider, and Identity Provider) to be relying on the Connector technology (IDSA, 2019). On the other hand, the IDS RAM also characterizes the Connector from its Deployment Context, Security Profile, Catalog, and Host. The Deployment Context designates the Connector's deployment environment (i.e., on-premises or cloud-based). Security Profile explicates the Connector's capability to enact a secure data exchange and processing environment. The host signals the communication protocol supported by the Connector to expose resources (i.e., HTTPS URLs, MQTT topics, etc.). Whereas the Catalog facilitates the participant discovery in the ecosystem based on the digital resources provided by the Connector.

We extend the Connector concept with additional properties to facilitate its discovery and selection. The Business Domain describes the context where the

Connector is specialized. Standards refer to the criteria with which the connector complies. The Pricing Model indicates how the end-users are expected to pay for the Connector's usage and acquisition. The Application Framework specifies which technology stacks are used to develop and support the Connector's runtime. Finally, the Data Usage Agreement is understood as the contract composed of the Data Usage Policy Pattern and agreed upon by the interacting Core IDS Actor to govern the data usage. As of now, five types of data usage patterns are supported by the Connector, and more designs may be added in the future.

6.2.4. Connector Store Ecosystem and System Architecture

The ORSD presented earlier, and the ontology conceptual model derived from it grounded the development of the Connector Store by providing a taxonomy of IDS Connectors and properties that characterize their operational context. These properties are defined to answer a list of ontology CQs related to the discovery and selection of IDS Connectors and their respective software providers, data owners, and users. By complying with this ORSD, the Connector Store aims to recommend connectors that are: (1) developed by a specific service or software provider; (2) developed for a specific business domain; (3) offered in a specific pricing model; or (4) developed using a particular technology. **Figure 30** depicts an alternative ArchiMate viewpoint detailing the internal system infrastructure of the Connector Store. It exposes the IDS Connector Provisioning Metadata Publication Service through its provisioning interface, which extends the metadata publication service defined primarily as a standard to implement a Metadata Broker. This service enables software and service providers to register, update, passivate, and delete their metadata entries and the connectors they offer.

The Connector Store combines Linked Data principles and Semantic Web technologies to store and provide metadata to describe IDS Connectors and their providers. Such technologies allow the integration of disparate open data sources in a standardized way (Soylu et al., 2020). This design decision connects well with the IDSA technical specifications, which indicate that a Metadata Broker should allow the discovery of data and other resources based on Linked Data principles (Bader, Bruckner, et al., 2020; IDSA, 2019). Therefore, the Connector Store uses the RDF format to describe the metadata of the IDS Connectors and data sources by annotating them with a layer of semantics to form subject-predicate-object triples (Berners-Lee et al., 2001). Examples of relevant triples can include: “Company A uses IDS Connector X”; “Software Provider B develops IDS Connector Y”; “IDS Connector X is specialized in the Transport Logistics sector”; or “IDS Connector Y is offered in a flat-rate pricing model”, as explained in **Sub-Section 6.2.1**. These knowledge representation triples could therefore support the IDS actors in discovering resources of interest based on semi-automated machine reasoning.

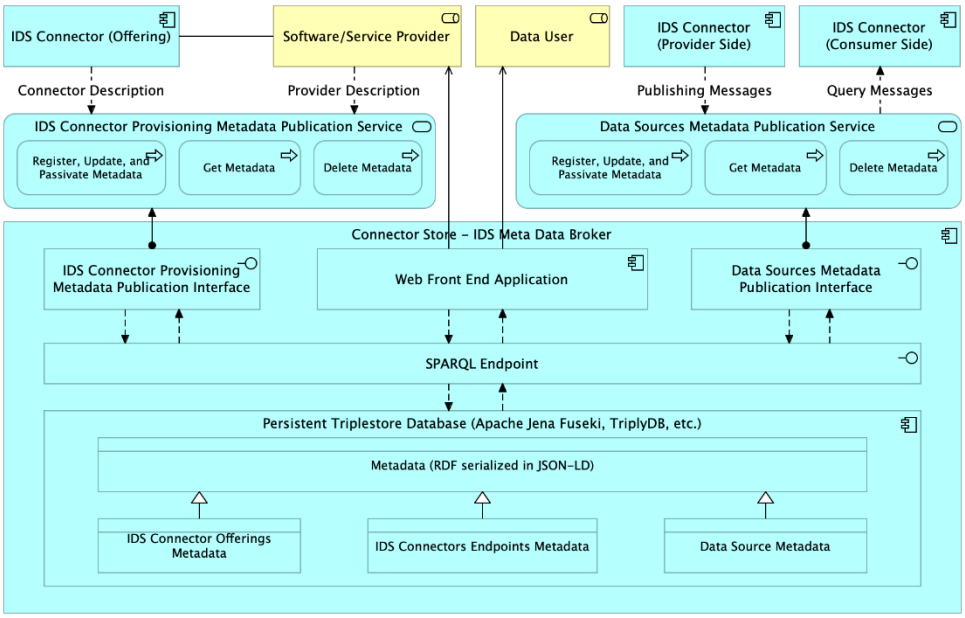


Figure 30 Connector Store – Internal System Architecture Viewpoint

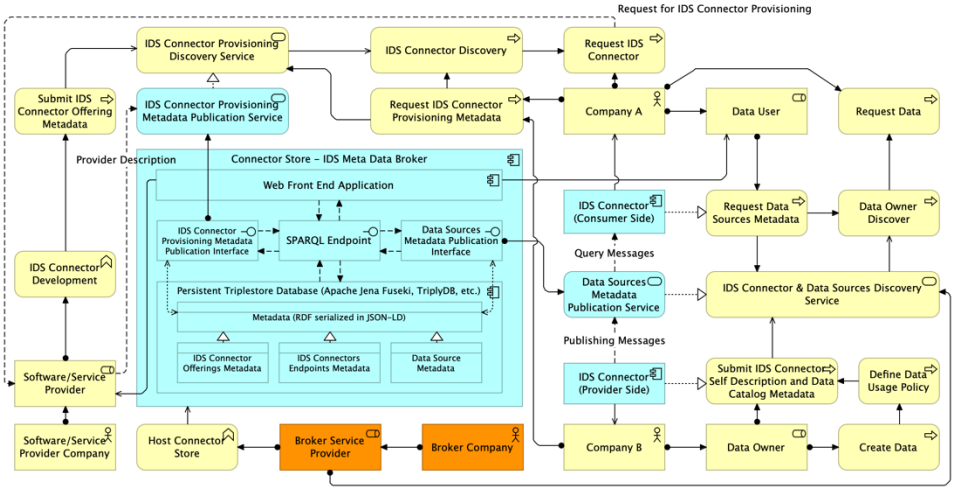


Figure 31 Connector Store – Ecosystem Interaction Viewpoint

To store these metadata represented in RDF, the IDSA suggests the use of a triple store database (e.g., Apache Jena Fuseki, TriplyDB, etc.) or any other storage back end that fits the purpose (Bader, Bruckner, et al., 2020; IDSA, 2020b). The Connector Store also needs to provide a SPARQL endpoint to allow data owners and users to (1) accept and send messages that comply with the IDS IM and (2) execute the metadata operation queries (IDSA, 2021f; Pérez et al., 2006; Soylu et al., 2020). These

IDS IM-compliant messages refer to the publish message that pushes metadata into the repository and the query message that pulls metadata from it (Bader, Bruckner, et al., 2020).

Finally, the Connector Store supports the ecosystem interaction shown in **Figure 31** which is based on the architecture depicted in **Figure 30**. It helps software and service providers submit the metadata describing the IDS Connectors they develop and offer in this ecosystem. Accordingly, through its IDS Connector Provisioning Service, the Connector Store allows data owners and users to find and acquire the best fit IDS Connectors based on their contextual information (in addition to the Data Sources Metadata Publication Service essential for the Metadata Broker to provide). Its mechanism enhances the process of discovering and selecting IDS Connectors for the participants, enabling them to quickly onboard to an IDS ecosystem.

6.3. Summary and Conclusion

This chapter explains how the Connector Store can be designed to support the discovery and selection of the IDS Connectors, data sources, and participants active in a logistics data space. The design process is initiated by an investigation of (1) the role of the BSP in IDS, (2) how this role supports and interacts with the other core participants, and (3) how its application component (i.e., the Metadata Broker) system architecture looks like. With this background knowledge in mind, this chapter shows how the Connector Store acts as a Metadata Broker and extends its functionality with the semantic discovery and selection of IDS Connectors. This is carried out by, first, specifying the requirements of the ontology that describes IDS Connectors that will be used as a façade of a Connector Store, and second, by visualizing the ontology conceptual model using OntoUML based on these requirements. This design phase, together with the previous chapter, traces back to the realization of the IDS adoption for the Dutch Logistics Sector discussed in **Section 4.3**. Next, two more ArchiMate viewpoints are presented to visualize (1) how the Connector Store can be assembled to utilize the proposed Connector Store ontology and (2) what kind of functionalities the Connector Store can provide to other core participants in a data-sharing environment by utilizing the proposed ontology. To operationalize and validate these design artifacts, the next step is to instantiate them into prototypes that make up the logistics data space demonstrator. Therefore, in the next chapter, we discuss how the prototypes can be developed based on the proposed designs will be discussed. This marks the end of the **Treatment Design** phase and, simultaneously, starts the **Treatment Validation** phase.

PART 3
TREATMENT VALIDATION

7 The Development of Application Prototypes for a Logistics Data Space

This chapter demonstrates the instantiation of the designs presented in the previous phase into working prototypes for validation purposes. Three application components' designs were elaborated on earlier, which also determine the structure of this chapter. Thus, **Section 7.1** discusses the development process of an IDS Connector based on the specification prescribed in **Chapter 5**. Next, the development of several IDS Data Apps providing several functions will be described in **Section 7.2**. **Section 7.3** describes the development of the Connector Store application to operationalize the ontology and system architecture presented in **Chapter 6**. Finally, **Section 7.4** presents the conclusion of this chapter.

7.1. The Development of an IDS Connector Prototype

The development of this IDS Connector prototype serves as a proof-of-concept for the requirements elicited from IDSA's documents and reference implementations. The main goal here is to demonstrate the technical feasibility of such an artifact to support data owners and users in managing data interoperability and sovereignty. Therefore, the development effort put into this prototype is spent mainly on replicating the relevant black box functionalities instead of replicating the accurate internal white box specifications. In the following sub-sections, we will discuss the development process of this prototype starting from preparing the deployment environment and how it can support the user to transform and route the requested resources to its user's internal enterprise system.

7.1.1. IDS Connector Operating System and Containerized Deployment Environment

Referring to the specifications discussed in **Section 5.1.2**, the IDS Connector is composed of several containerized applications managed under an application container management layer that runs on top of an operating system. These containerized applications comprise the core connector as the backend for application logic and a GUI as the front-end for user interactions. This thesis develops the IDS Connector prototype with Mendix[41], a low-code web application development platform that streamlines the front and back-end application development (and database design) into a single application. This alignment accelerates the development process of the prototype. It also supports REST (and SOAP) API protocol for data exchange and integration with other applications, making it suitable for this purpose.

[41] https://www.mendix.com/

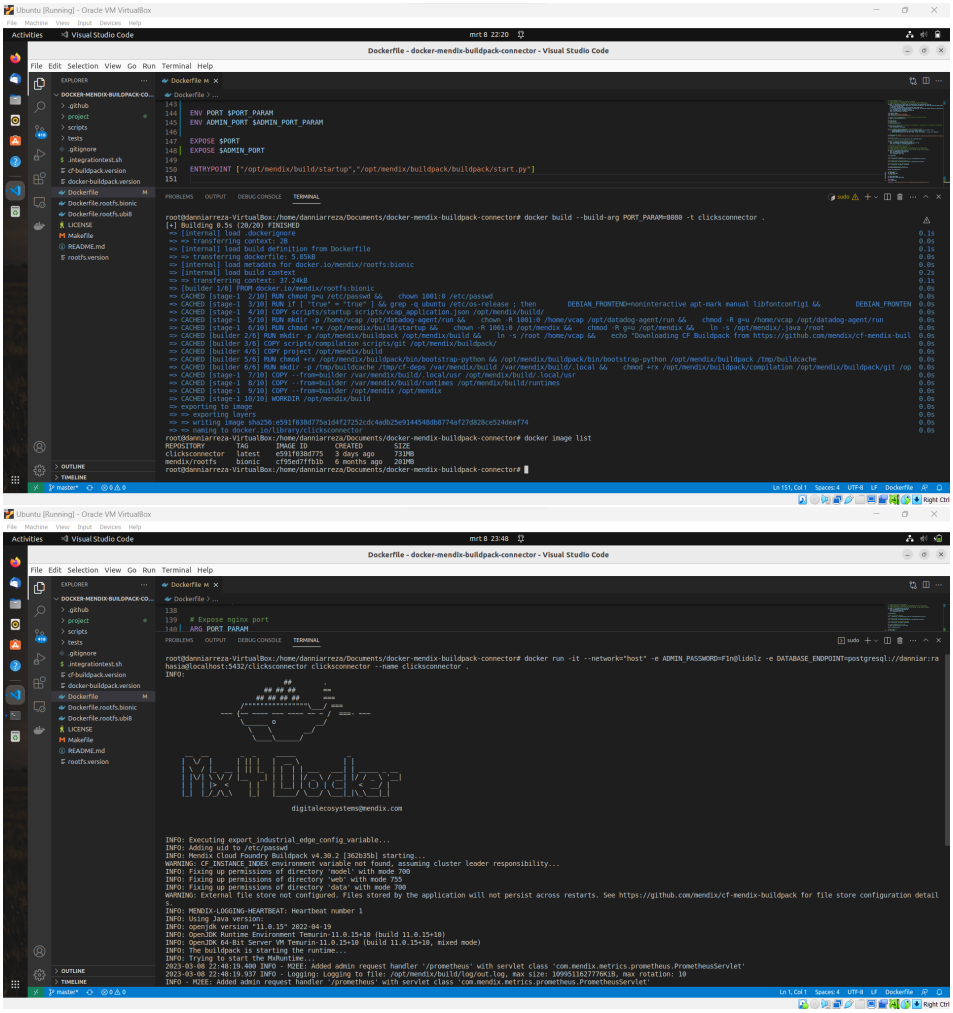


Figure 32 IDS Connector Prototype – Build Image and Run IDS Connector Container

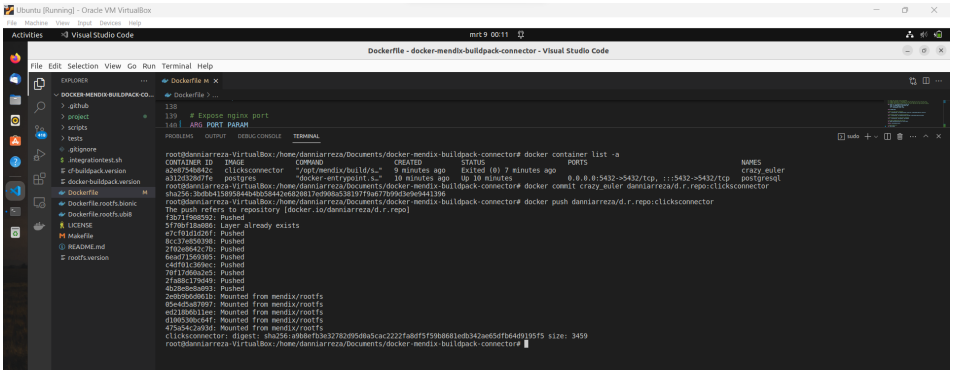


Figure 33 IDS Connector Prototype – Commit & Push Image to DockerHub Repo

Mendix also allows development in a local environment and provides one-click deployment on a free cloud environment[42]. By default, this free cloud environment is powered by the AWS[43] platform. However, there is little information that can be found on how this environment is configured, especially in relation to the containerized application deployment environment. Therefore, to ensure that the prescribed design is satisfied, we followed an online guide on how to deploy and run this Mendix-based prototype on a Docker environment[44]. This requires us to, first, build an image of the developed Mendix application so then later we can run it in a container on any Docker environment, demonstrated in **Figure 32**. For this, we used the Mendix Build Pack for Docker[45] published in GitHub which works by encapsulating the Mendix project file with configurations relevant to assembling and deploying a Docker image. Afterward, we commit and push the image to the DockerHub repository (with the repo name danniarreza/d.r.repo:clicksconnector) after generating it, as **Figure 33** shows. **Table 36** elaborates on the Docker commands that we used for this task, from generating the image to pushing it to the repository.

Table 36 Docker Commands to Build and Push IDS Connector Prototype Image

No	Command	Description
1	`docker build --build-arg PORT_PARAM=8080 -t clicksconnector .`	(Local) Build IDS Connector's image.
2	`docker run --name postgresql -p 5432:5432 -e POSTGRES_USER=<db_username> -e POSTGRES_PASSWORD=<db_password> -e POSTGRES_DB=clicksconnector -d postgres`	(Local) Run a container of PostgreSQL for IDS Connector's database.
3	`docker run -it --network="host" -e ADMIN_PASSWORD=<admin_password> -e DATABASE_ENDPOINT=postgresql://<db_username>:<db_password>@localhost:5432/clicksconnector clicksconnector --name clicksconnector .`	(Local) Run a container based on IDS Connector's image. The name of this container will be used in Step 5 below.
4	`docker login`	(Local) Optional, authentication for pushing the image to DockerHub repository.

[42] https://clicksidsconnectorv1-sandbox.mxapps.io/
[43] https://aws.amazon.com/what-is-aws/
[44] https://docs.mendix.com/developerportal/deploy/run-mendix-docker-image/
[45] https://github.com/mendix/docker-mendix-buildpack

5	`docker commit <ids_connector_container_name> danniarreza/d.r.repo:clicksconnector`	(Local) Commit the image of the running IDS Connector's container to the local repository.
6	`docker push danniarreza/d.r.repo:clicksconnector`	(Local) Push the committed image from the local to DockerHub repository.
7	`docker run --name postgresql -p 5432:5432 -e POSTGRES_USER=<db_username> -e POSTGRES_PASSWORD=<db_password> -e POSTGRES_DB=clicksconnector -d postgres`	(Cloud) Run a container of PostgreSQL for IDS Connector's database.
8	`docker run -it --network="host" -e ADMIN_PASSWORD=<admin_password> -e DATABASE_ENDPOINT= postgresql://<db_username>:<db_password>@localhost:5432/clicksconnector danniarreza/d.r.repo:clicksconnector --name clicksconnector .`	(Cloud) Run a container based on the pushed IDS Connector's image.

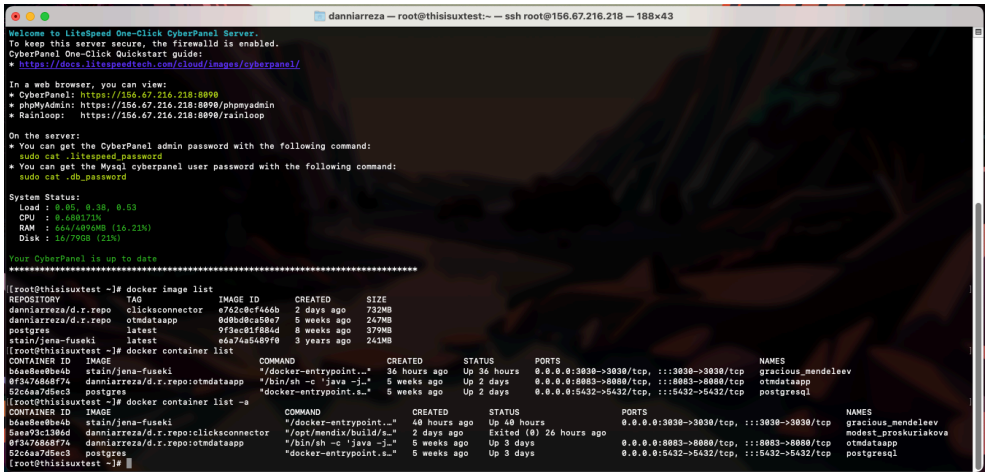


Figure 34 IDS Connector Prototype – Docker on VPS Environment for Deployment

Having the image of the IDS Connector compiled and ready to be instantiated into a container, we proceed with preparing the Docker environment to host its deployment along with other relevant components (e.g., database system, IDS Data Apps, etc.). For this, we used a VPS hosting provided by Hostinger[46]. The server

[46] https://www.hostinger.com/vps-hosting

runs a clean Linux CentOS[47] installation as a starting point. Therefore, we follow the guide from Docker's documentation page[48] to install Docker Engine on this server. Upon successful installation, the VPS environment is now ready to execute and spin application containers based on the images we pushed to the DockerHub.

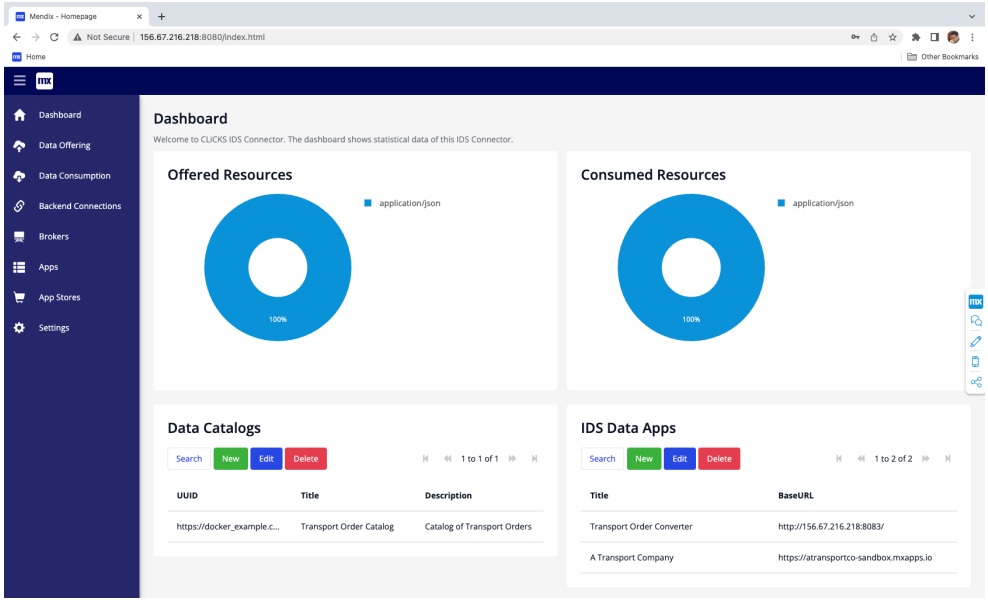


Figure 35 IDS Connector Prototype – Running on a Docker Container

Figure 34 depicts the Docker operation on the VPS environment operated through the command-line interface of a local machine via an SSH connection. In addition, **Figure 34** also shows the list of (1) images downloaded from DockerHub and (2) running containers based on these images, which are executed from the commands described in **Table 36**. Note that we annotated this table's Description column with (Local) or (Cloud) to indicate in which environment these commands are executed. **Figure 35** is then presented to clarify how this IDS Connector prototype runs on a Docker container in the VPS environment with the Base URL and port of http://156.67.216.218:8080. Note that this address might not work in the future after this thesis is published due to the hosting service being a temporary subscription. With the deployment environment prepared, the development of the core connector application itself can be initiated. Subsequently, this will be discussed in more detail in the next sub-section.

[47] https://www.centos.org/

[48] https://docs.docker.com/engine/install/centos/

7.1.2. IDS Connector Data Model Implementation

We start the development of the core connector from the implementation of the data model prescribed by the IDSA as discussed in **Section 5.2.2**. Seven main data entities were enlisted, namely, the Catalog, Resource, Representation, Artifact, Contract, Rule, and Agreement. **Figure 36** presents the implemented data model in Mendix Studio, which is called the domain model in the studio. Among these data entities, the seven prescribed data entities are the ones that are highlighted with red boxes. Not all attributes are visible in **Figure 36**. An example is the creationDate and modificationDate since they are embedded as default attributes within the entity. Another example is remoteId in the Representation entity, due to its usage is found to be still questionable.

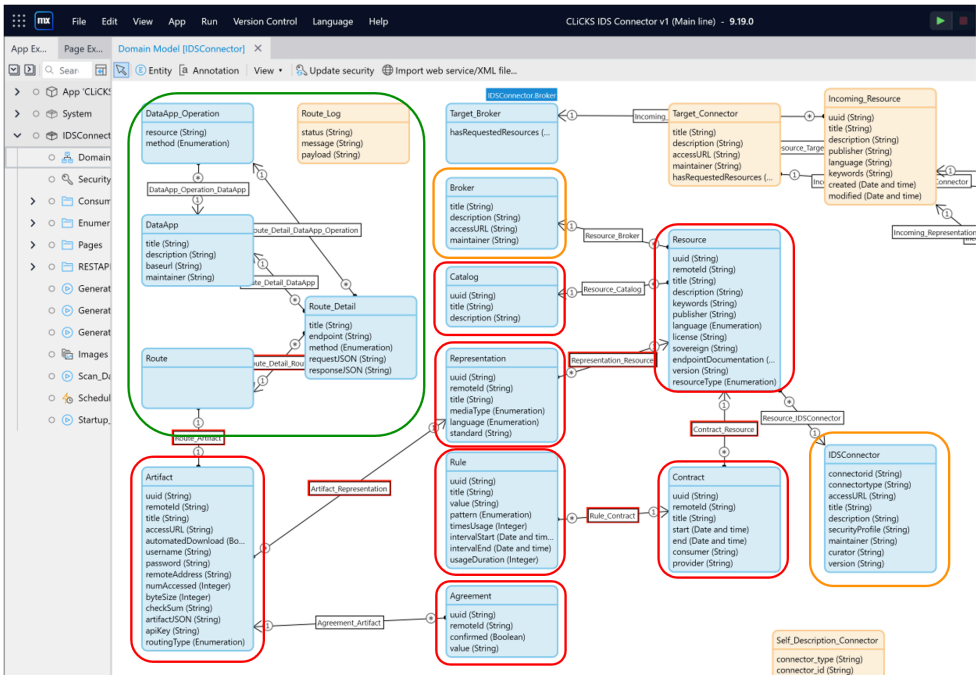


Figure 36 IDS Connector Prototype – Data Model Implemented in Mendix

Next to that, we extend the domain model with several data entities that are relevant for supporting the other functionalities. An example of them is the Broker and IDS Connector entities, highlighted in orange boxes, that are meant for identifying from which IDS components the requested resources are coming. The IDS Connector entity is also used to describe this IDS Connector itself, either when another IDS Connector requests the self-description of this connector or when this connector publishes metadata to a Broker. On the other side, the data entities inside the green box are used for supporting data routing and IDS Data Apps orchestration. Artifact objects have a Route that is defined with several destination points, called Route Details. Each of these Route Details is associated with a Data App and refers to this

app's operation endpoints. **Section 7.1.3** will discuss in more detail how these data entities are operationalized with application logic and user interfaces.

7.1.3. IDS Connector Resource Offering and Metadata Publication

The first functionality discussed here is about how the IDS Connector can support data owners to offer a data resource for other data space participants. This functionality includes defining this resource's data usage policies, representation and its artifacts, and catalog, as well as the Broker to whom the data owner wishes to publish this resource. Referring to the initial design explained in **Section 5.2.3**, the first step in offering a resource is to define some metadata that describes it. This step, in this prototype implementation, is depicted in **Figure 37**. Within the "Data Offering" menu, the user is provided with the "New" or "Edit" page to create a new data offering. On this page, the user, acting as the data owner, describes the general information of the data that it will offer.

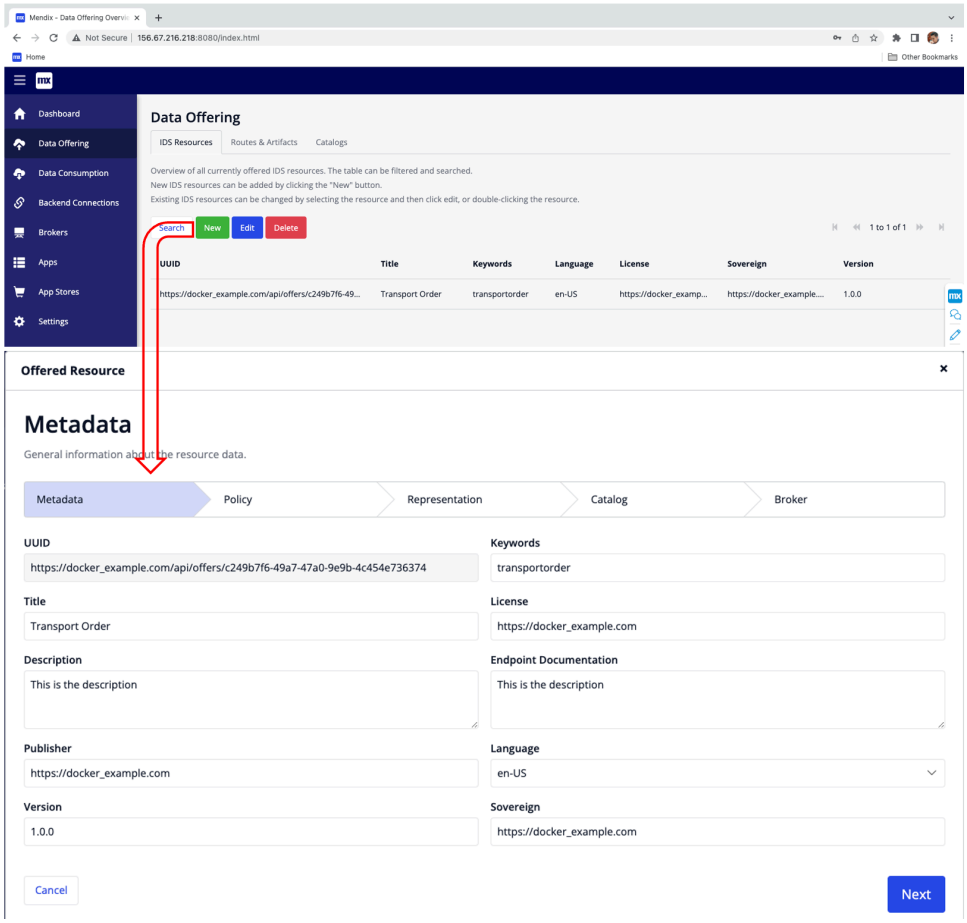


Figure 37 IDS Connector Prototype – Data Offering's Metadata Description

Next, the data owner defines the data usage policy that is binding the offered resource in concern. Note that the data usage policy discussed here corresponds to the Contract object described in the domain model in **Sub-Section 7.1.2**. On this page, the data owner selects and attaches the data usage rules that make up the data usage policy. **Figure 38** shows the interface for the user when selecting the rule for the data usage policy. On the right side of the image is the list of data usage rule templates that the user can select. By the time this prototype is developed, the first six rules are supported (i.e., from "Provide Access" to "Usage Until Deletion"). The rest are not yet developed as they require more investigation that is outside the scope of this thesis (i.e., investigation on the design and development of other IDS components).

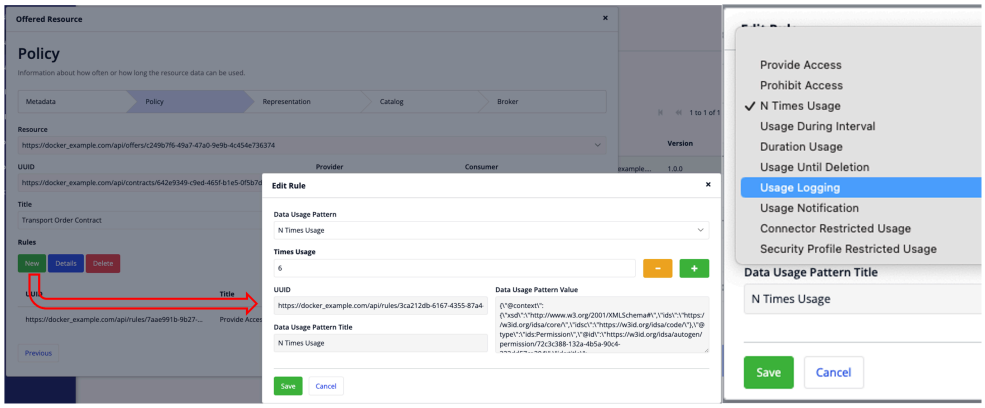


Figure 38 IDS Connector Prototype – Data Offering's Data Usage Policy and Rules

Additionally, in this image, the "Usage Logging" rule is highlighted. To function, this rule connects with a Clearing House that, if implemented, may further enhance the data sovereignty capability of participants in the data space. This rule can be implemented by specifying the IDS Connector to log and report to a Clearing House regarding every usage of the data within the data space boundary. Though the IDSA has provided some guidelines regarding its functions and responsibilities, little to no research works are found that describe and investigate the design of such a component, particularly in the logistics sector. Therefore, in this thesis, we exclude the development of this data usage rule and open this research topic to be further investigated as a master thesis.

Following the data usage policy definition, the data owner proceeds to describe the representation and its artifacts. In this page, shown in **Figure 39**, the data owner describes the representation with a title, language, and media type. In this current development of the prototype, these attributes do not directly influence any functionalities of the connector. However, it is foreseen that this attribute is reserved for describing the offered resource with several representations of different languages (e.g., in de_DE, nl_NL, etc., in addition to en_US) and media types (e.g., in XML, CSV, etc., in addition to JSON).

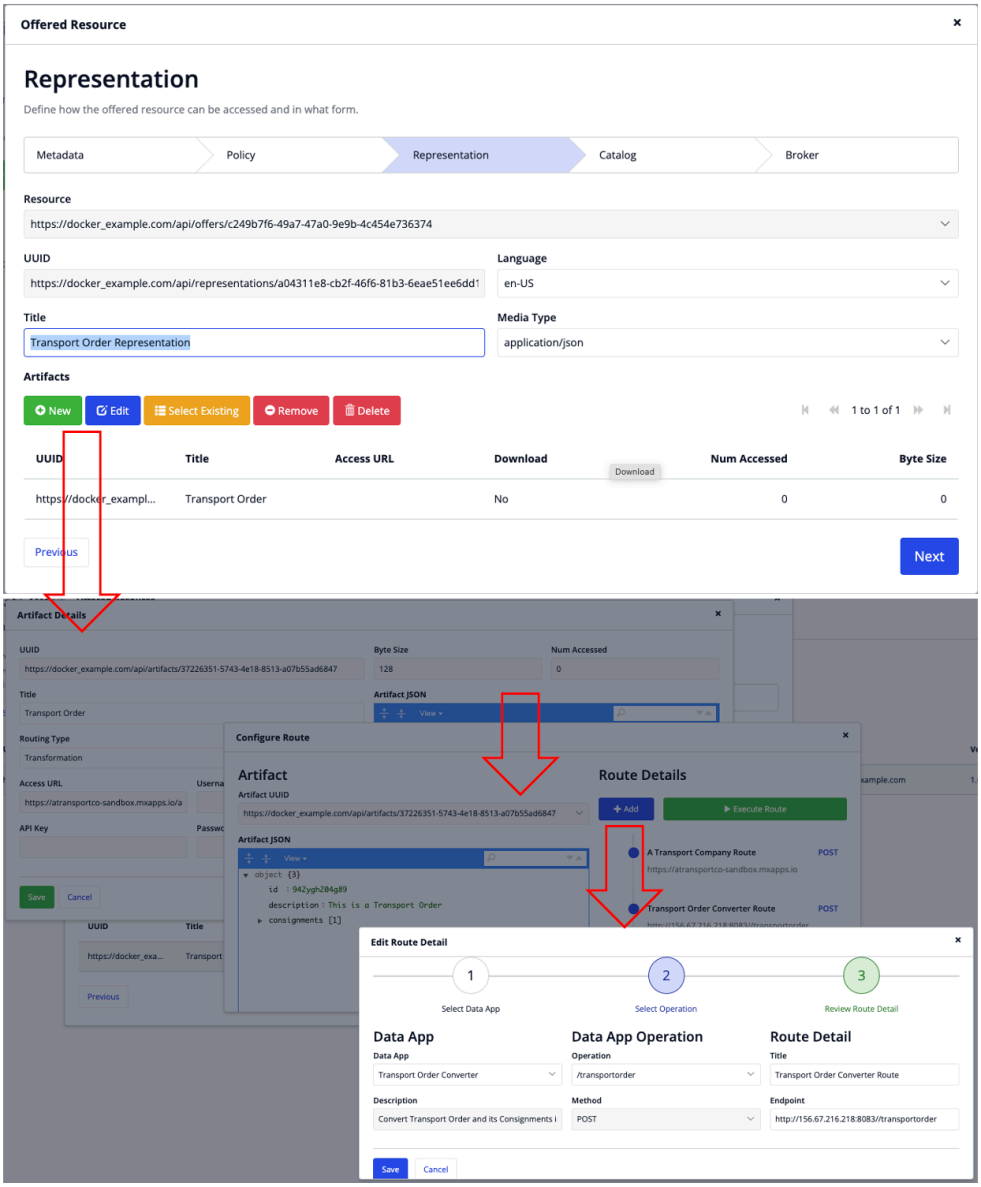


Figure 39 IDS Connector Prototype – Data Offering's Representation, Artifacts, and Data Apps Routing

Next to that, the data owner attaches artifacts that are relevant to the described representation by either creating a new one or selecting an existing one. First of all, creating a new one requires the data owner to fill in the artifact's title. Secondly, we provided an additional attribute, the routing type, for the data owner to decide whether the data for this artifact (1) will be directly pulled from a data source (e.g.,

DBMS or other application via REST API) or (2) will require routing and transformation functionality before offered to the data space. For the first scenario, the data owner can select the "Passthrough" as the routing type, which allows him to specify the Access URL where the data can be pulled from and API Key (as well as Username and Password) if required for authentication purposes to this source. He can retrieve the data from this URL and store it in the Artifact JSON attribute by clicking the "Synchronize" button. For the second scenario, on the other hand, the data owner can select the "Transformation" routing type to make use of the route object associated with this artifact.

In **Figure 39**, this latter type is selected so the data owner gets access to the page for configuring the route. On the right side of this page, he is provided with the "Add" button to add a route detail, which will open a page for him to select a Data App and its operation to fill in the required attributes of the route detail. These route details will line with one another vertically, indicating the flow of the route when executed. To execute this route, he can click the green "Execute Route" button. The result of this route execution (i.e., the payload under the Route_Log data object in **Figure 36**) will then be stored in the Artifact JSON attribute of the Artifact object to be offered as the offered resource.

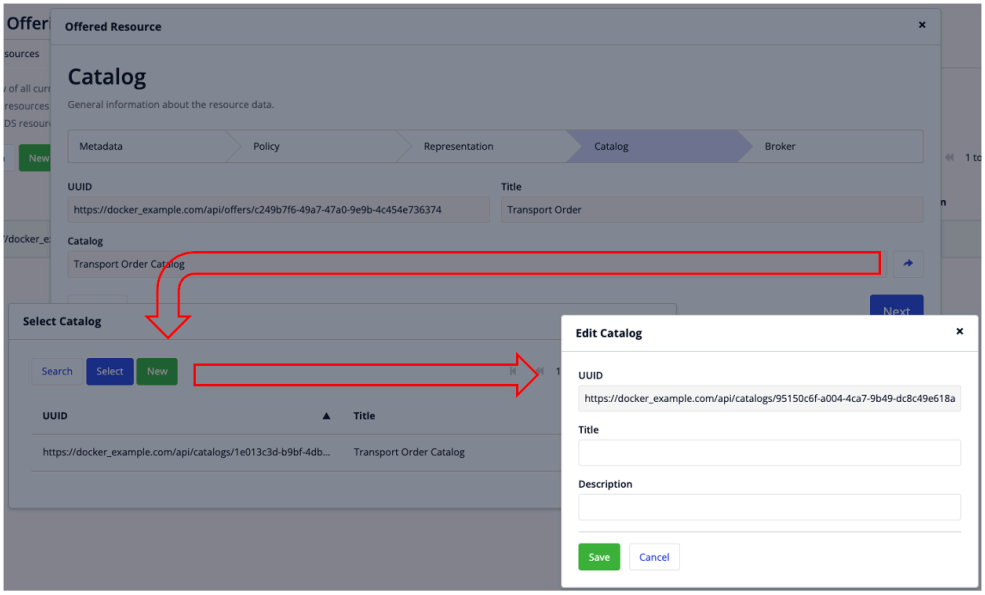


Figure 40 IDS Connector Prototype – Data Offering's Catalog

The last two steps are optional with respect to the whole process of creating an offered resource. **Figure 40** illustrates the page where the data owner can attach this offered resource to a particular data catalog. This step is provided if the data owner intents to categorize their offered resources under a certain grouping entity. Consequently, this step satisfies the data model discussed earlier in **Figure 25** and

Figure 36. Interestingly, this design correlates with the RDF-based DCAT[49] vocabulary that prescribes the data structure for publishing data catalogs on the Web using their metadata.

Lastly, to facilitate resource and service discovery, the data owner can finalize the whole resource offering process by registering this resource to an active Broker of choice. **Figure 41** presents the page where the data owner can select the Broker to publish the metadata of this resource. There is a possibility that this IDS Connector has not yet registered itself with a Broker or has not yet stored the information of a Broker in its database. Therefore, the data owner can find a Broker based on the Broker's access URL.

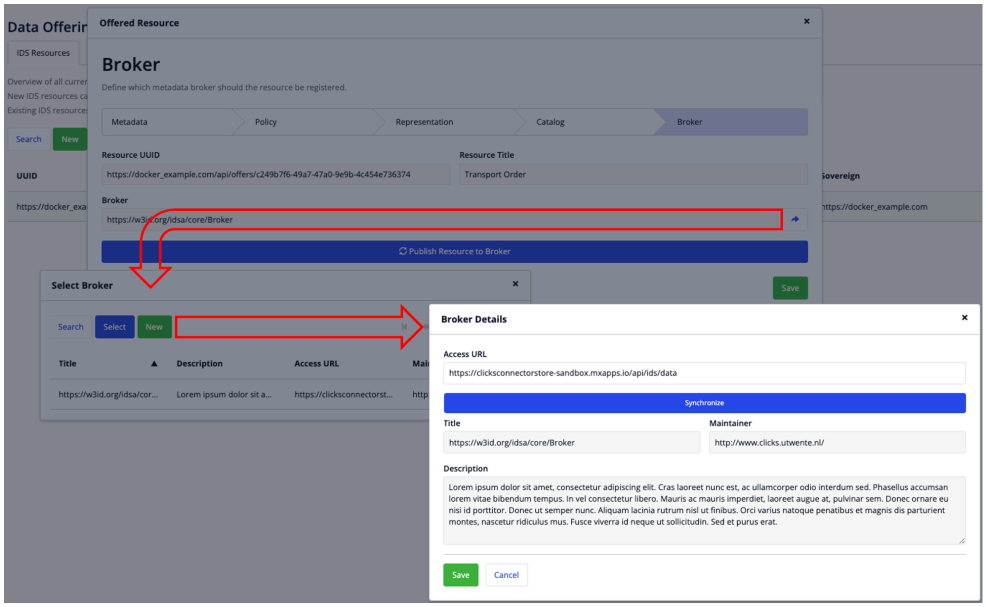


Figure 41 IDS Connector Prototype – Data Offering's Publishing to a Broker

For this, the data owner can fill it in with this URL template https://<base_url>/api/ids/data since, based on the IDSA's reference implementations, most IDS components can be reached and requested through this endpoint. Once the data owner hits the "Synchronize" button to retrieve the Broker's self-description and save it, he can select it and publish the resource to this Broker by clicking the "Publish Resource to Broker" button. Finally, he can click the "Save" button on the bottom right to close the Offered Resource page.

[49] https://www.w3.org/TR/vocab-dcat-2/

7.1.4. IDS Connector Resource Request and Contract Negotiation

Once a resource has been offered by a data owner to the data space, data users can start to find and request it. Adhering to the enterprise architecture in **Figure 15** and the sequence diagram in **Figure 26**, data users can discover the offered data resources through the Broker or via a query sent to the Broker. **Figure 42** illustrates this functionality. The users of the IDS Connector, acting as data users, can go to the "Data Consumption" menu, where they can find the "Request Resource" button under the "IDS Resources" tab.

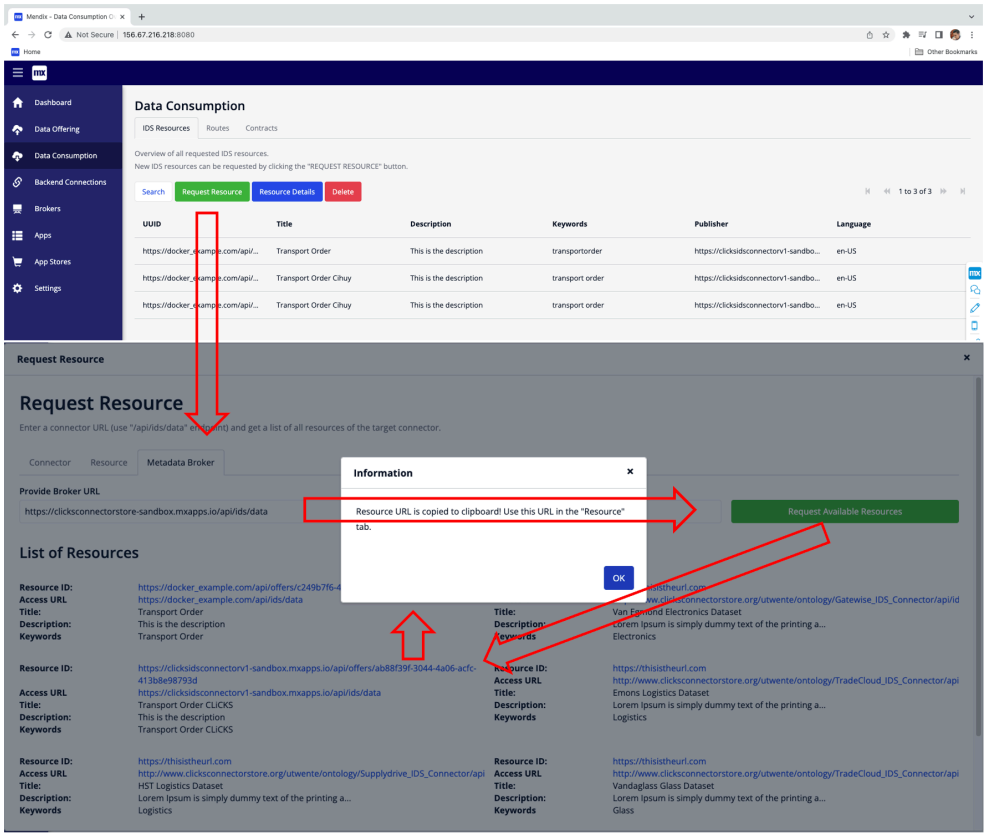


Figure 42 IDS Connector Prototype – Data Consumption's Metadata from Broker

The pop-up page shown in **Figure 42** provides three tabs, namely the "Connector", "Resource", and "Metadata Broker". Multiple scenarios can be considered here. Assuming that the data user is not aware of the available resources offered in a data space, he can first request the resources' metadata from a Metadata Broker by sending a query to the Broker's access URL (as mentioned earlier, the access URL is indicated by the "/api/ids/data" endpoint). In this prototype implementation, we simplify this process by just providing the Broker's access URL and the user can hit the "Request Available Resources" green button. The Metadata Broker will supply

this IDS Connector with a list of the offered resources published by this Broker. This provides the user with the ID of the resource or the access URL of the offering IDS Connector that can be used in the "Connector" or "Resource" pages, which will be discussed in **Figure 43**.

Once the data user obtains either the data owner's IDS Connector's access URL or the resource ID, he can use it to request the list of offered resources or the list of representations respectively. In essence, such a resource ID represents the UUID that is used to request the resource's metadata through the web service endpoint of the offering IDS Connector. **Figure 43** shows an example of how to request the offered resource using the provided input field after the data user obtains the resource ID. Once the data user hits the "Request Available Representations" button, the IDS Connector will request the available representations of this resource from the offering IDS Connector. He can select one of the representations and click the green lock button to initiate the contract negotiation process.

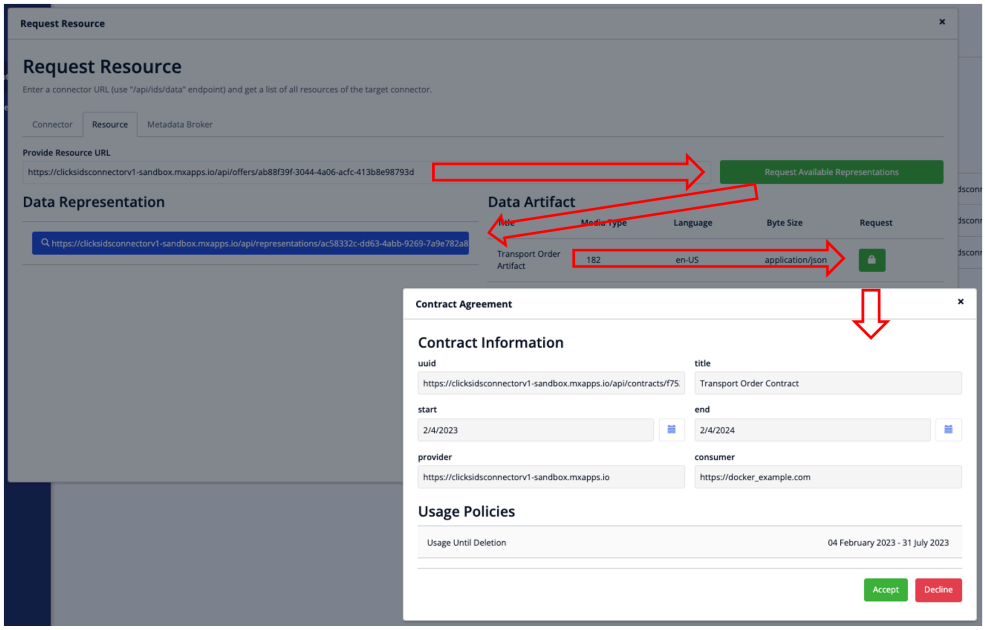


Figure 43 IDS Connector Prototype – Data Consumption's Resource Request and Contract Negotiation

Clicking this button opens a new pop-up page that shows the contract detail regarding the data request and usage policy agreement. This contract binds the providing and requesting parties under the provider and consumer fields. Note that in this example, the provider refers to the IDS Connector that is deployed using the default Mendix free cloud deployment environment https://clicksidsconnectorv1-sandbox.mxapps.io. On the other hand, the consumer that is currently described as

the https://docker_example.com refers to the IDS Connector deployed on the VPS environment under the http://156.67.216.218:8080 address. Next to this, the contract also informs the list of rules that binds the usage of this requested data. Therefore, upon accepting it, the data user is expected to provide its agreement to this contract. Once he accepts, the metadata of the requested resource will be downloaded and made available on the data user's IDS Connector. **Figure 44** shows this metadata, which provides all relevant information regarding the resource. The next section will discuss how this requested resource can be used by the data user.

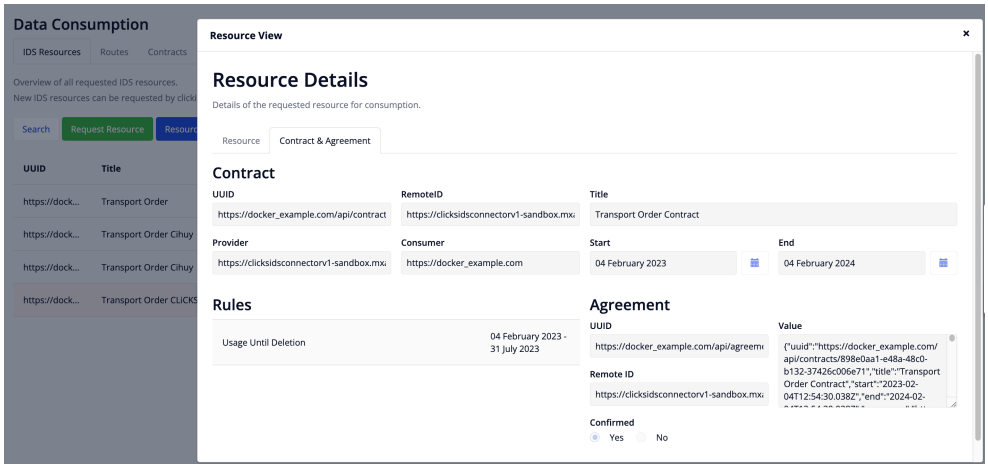


Figure 44 IDS Connector Prototype – Data Consumption's Requested Resource

7.1.5. IDS Connector Artifact Consumption and Route Execution

Upon receiving the metadata of the requested resource, the data user can start the process of consuming it. For this purpose, he can navigate to the "Resource" tab under the same pop-up page shown in **Figure 44**. Before the data user can use or consume the requested resource for his internal enterprise system or other data sink, he can click the blue "Sync" button to trigger a REST API call to the artifact's access URL and store the returned response under the artifact object as prescribed by the data model in **Figure 25** and **Figure 36**. As mentioned earlier, in the current version of the prototype, only a JSON-formatted message is provided. The idea is that the media type attribute on the representation object of this requested resource will help the IDS Connector to determine how this response will be stored, which in this case, is stored under the Artifact JSON attribute.

Figure 45 illustrates this process by showing how the Artifact JSON field is filled in with the data value after triggering the "Sync" function. Note that there are also noticeable changes in Num Accessed and Byte Size between before and after the sync. This Num Accessed increment correlates with one of the data usage rules discussed in **Figure 38** as the "N Times Usage". Assuming that this requested resource has the rule of N Times Usage with N being 5, then after 5 synchronizations the data user can no longer synchronize or request the latest data. The example

shown in **Figure 44** shows that the defined rule is "Usage Until Deletion". This means that the value under the Artifact JSON will be automatically deleted after the determined data, which in this case is after 31 July 2023.

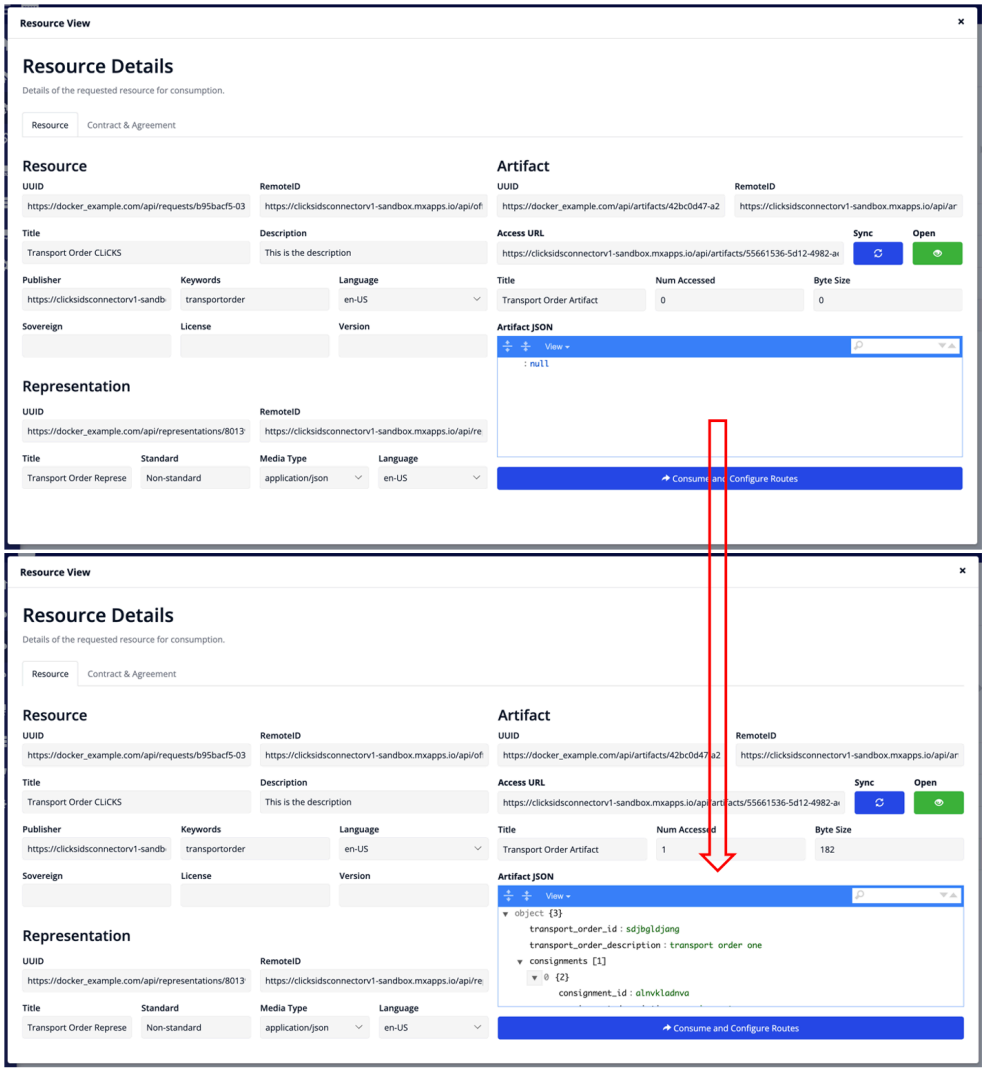


Figure 45 IDS Connector Prototype – Data Consumption's Artifact Synchronization

To consume this resource for the use of the data user's enterprise system or data sink, the data user proceeds with the "Consume and Configure Routes" button to configure the data app's orchestration and message routing. Upon clicking this button, the "Configure Route" pop-up page as shown in **Figure 46** will be provided. The idea here is that the IDS Connector facilitates the data user to configure the route

for consuming the requested data resource under the artifact's JSON. Doing so might require some transformations or other data processing capabilities before the resource can finally be retrieved by the enterprise system. Alternatively, it may also be the case that this requested resource is to be used by more than one target system (i.e., multiple endpoints) along the way.

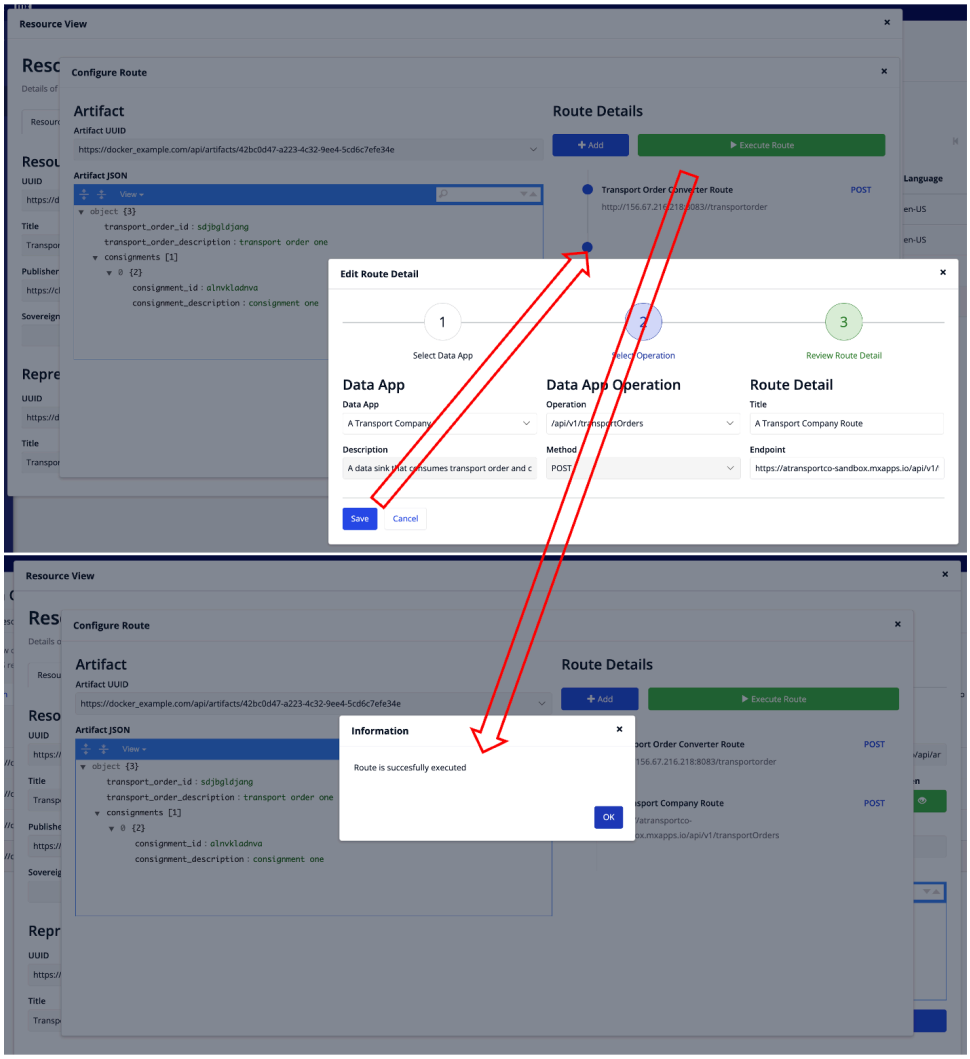


Figure 46 IDS Connector Prototype – Data Consumption's Route Configuration and Execution

Therefore, to facilitate these use cases, a route is defined to have more than one route detail. On the left side of **Figure 46**, one can observe the artifact JSON field box that indicates the JSON value of the artifact prior to the route execution. Here, we demonstrated how the data user can add a new route detail. He can select a data

app and its operation (i.e., endpoint) in order to assemble the route detail (i.e., refer to Step 3 within the pop-up page). Note that, in the example provided above, the "Transport Order Converter Route" represents a Data App that can return a transport order and its consignments from a non-compliant format to an OTM-compliant one. On the contrary, the "A Transport Company" app is in fact an endpoint of a mockup enterprise system. This means that the data user does not necessarily have to choose a data processing or transforming app for the route detail. Upon saving the route detail and reviewing all details of the whole route, the data user can hit the green button to execute the route. Once the route execution for this particular scenario is done, the data value described under the artifact JSON is sent to the converter app to be converted into OTM and then forwarded to the target "A Transport Company" app. **Figure 47** shows the API documentation of the app (a simple Mendix app accessible online on this address https://atransportco-sandbox.mxapps.io), which expects id and descriptions as the attributes of a transport order and its consignment as opposed to transport_order_id and transport_order_description. At the same time, this figure also reports that the data is received on its end.

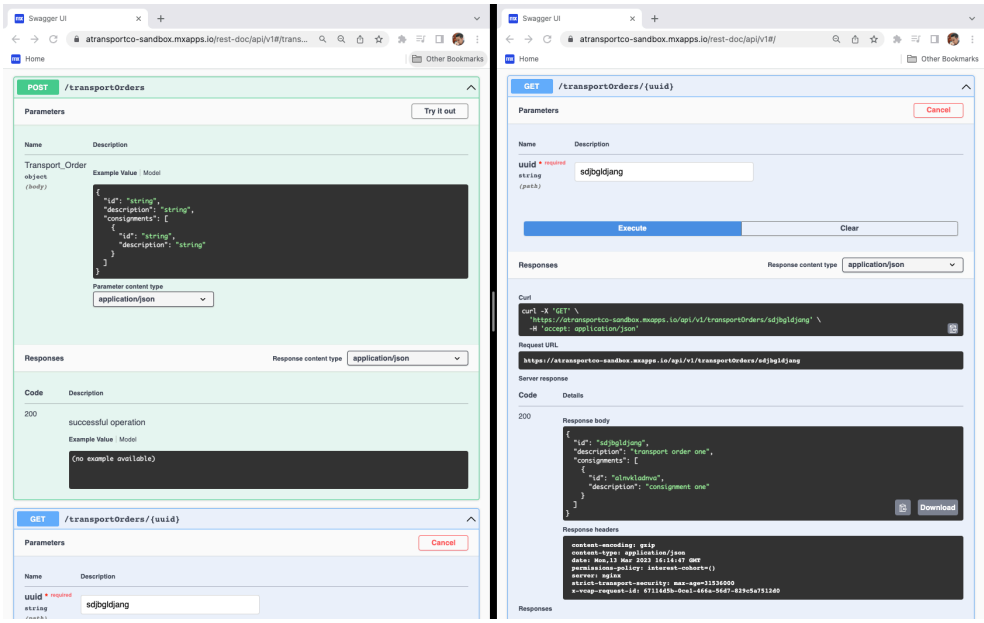


Figure 47 Mockup of a Transport Company Enterprise System

7.1.6. IDS Connector IDS Data Apps Registration and Management

In the previous sub-sections, we demonstrated how the IDS Connector prototype can support its users with data transformation and message routing through the data app orchestration. This orchestration follows the concept of service orchestration, in which, the services being orchestrated do not always have to be

IDS Data Apps. We allow the user to register and orchestrate the services under their internal environment with this IDS Connector prototype. **Figure 48** is presented to show how the user can do this through the "Apps" menu.

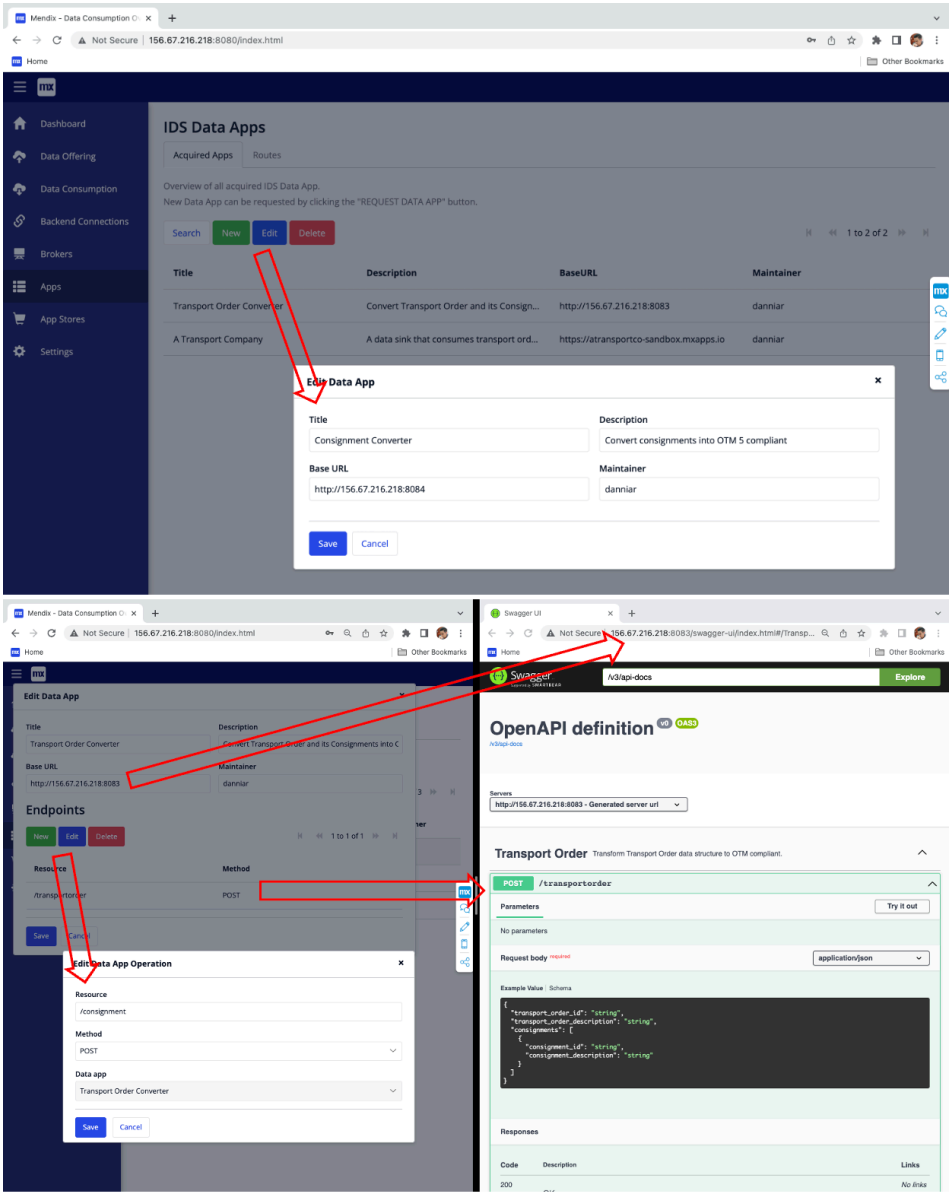


Figure 48 IDS Connector Prototype – IDS Data App's Registry and Management

In this menu, the user is presented with the list of data apps, or services, that he has registered to his IDS Connector. In registering a new service, the user fills in the title of this service, base URL, description, as well as the maintainer. Upon saving it, the

user can proceed to add endpoints that are supported by this service by clicking the "Edit" button or double-clicking a service. On the bottom part of **Figure 48**, we can see the pop-up page being extended for the user to manage the endpoints. In this extended part, the user can register a new endpoint related to this service by clicking the "New" button. This will open a new pop-up page that requires the user to type in the endpoint's address (which in the provided example is "/consignment" or "/transportorder") and the HTTP Request Method (i.e., GET, POST, PUT, DELETE, etc.). These two fields are provided to match with the endpoints and request method of the target system, as shown in the bottom right part of **Figure 48**. Having configured these services in this menu, the data user can reuse these configurations in setting up the route details as discussed in **Sub-Section 7.1.5** earlier. In the following **Section 7.2**, we will elaborate on the process that we followed to develop the prototypes of IDS Data Apps for this thesis. Next to that, considering its relevance for this design project, the next chapter will also discuss how we can configure and utilize the eMagiz platform to address data interoperability and schema mapping issues.

7.2. The Development of IDS Data Apps Prototype

Having the IDS Connector prototype demonstrated and able to orchestrate IDS Data Apps, the next step is to develop these apps to offer data processing and other additional capabilities for the connector. Similar to the development of the IDS Connector prototype, this development of the IDS Data Apps only serves as a proof-of-concept to demonstrate how they can be designed to alleviate data interoperability issues by providing additional functionalities for the IDS Connector. In the following sub-sections, we will discuss the development of two examples of data apps that provide 2 different data transformation specifications. In addition, we will also demonstrate how we can configure and set up a data transformation app using the eMagiz platform as an alternative. This alternative can be relevant for companies who need to procure their own data transformation and message routing component since eMagiz provides the means for clients to develop and configure their own system integration landscape in a low-code manner.

7.2.1. Developing IDS Data App for OTM Data Transformation

The first IDS Data App prototype that we built is a data app to transform transport order and consignment data from a proprietary format into a format that complies with the OTM standard. In this development, it is assumed that a GUI for the data apps is not crucial, as their operations will be orchestrated by the IDS Connector itself. Therefore, we used Java Spring Boot framework to develop it, as it provides extensive support for REST API through its Spring Web dependency, among other dependencies such as Spring Doc UI dependency for Swagger API documentation[50].

[50] https://springdoc.org/#getting-started

Referring to the specifications listed in **Sub-Section 5.1.3**, especially under requirement APS_5, that an IDS Connector supports IDS Data Apps to be delivered and installed as independent software containers (i.e., IDS Data Apps bring along possible dependencies of, e.g., software modules themselves and can be used irrespective of the IDS Connector's configuration). Similar to the IDS Connector, this signifies the importance of the IDS Data Apps to be developed and deployed as a containerized application (e.g., for a Docker environment) in mind. Therefore, in **Figure 49**, we prepared a Dockerfile configuration file to build the Docker image out of the built jar file. This jar file is built using the "./gradlew bootJar" command and then stored inside the build->libs folder. To build the docker image of this jar file and push it to the Dockerhub repository, we used this command:

"docker buildx build -t danniarreza/d.r.repo:otmdataapp --platform linux/amd64,linux/arm64/v8 --push ."

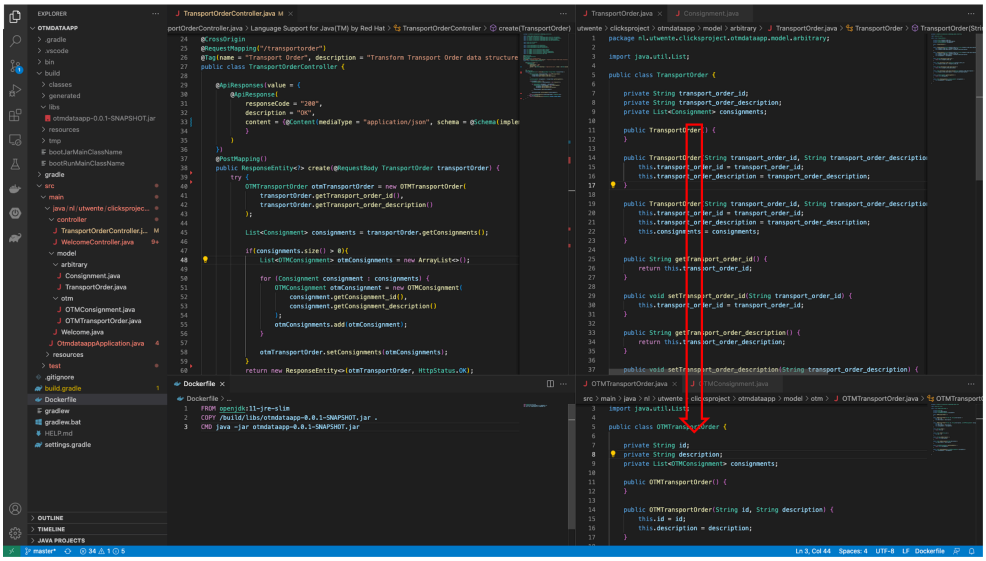


Figure 49 IDS Data Apps Prototype – OTM Converter Source Code

In this project, we developed this data app to be a simple application that does data transformation to OTM-compliant transport order data, including its consignments, from a non-compliant one. As illustrated on the left side of **Figure 49**, this works by taking in the non-compliant transport order data through a function, called create() and exposed as a web service operation with the "/transportorder" endpoint, and then returning to the caller the compliant format. This design is made to comply with the requirement IDA_2 specified in **Table 15** earlier. In this example, the non-compliant format refers to the name of the attributes that follow the snake case style. Whereas, the OTM-compliant format refers to the ones that follow the camel case style.

In **Figure 49**, the data structure being transformed is indicated by the red arrow. This way, the functionality provided by this data app is documented on the Swagger UI page as shown in **Figure 50**. As shown in **Figure 50**, in this development, this data app prototype is still limited to transforming the data structure shown on the left side into the one returned on the right side. However, this opens up the possibility for future investigation and development to support more operations and data structure if the demand exists.

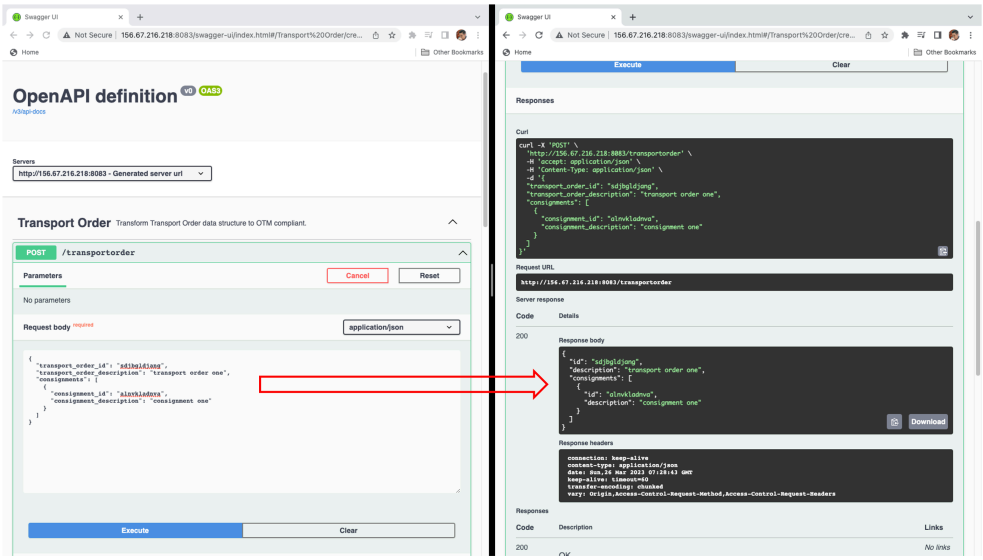


Figure 50 IDS Data Apps Prototype – OTM Converter Swagger API Documentation

To provide more examples of the component, the second IDS Data App prototype we developed is a data app to transform consignment data from the OTM-compliant profile into the specification that fits a custom backend system. This second data app, as well as the mentioned backend system[51], will be discussed in its usage for the Transport Trip Carbon Emission Tax Reporting mockup case in the next chapter as part of the validation of this thesis. Therefore, as shown in **Figure 51**, we call this component the Regulatory Body Data App.

Similar to the OTM Converter Data App, this data app is developed with Java Spring Boot framework, in tune with the same Spring Web and Spring Doc UI dependencies for REST API support and documentation. We also prepare a Dockerfile configuration file resembling the previous data app to build and push the Docker image to Dockerhub, which requires a Docker command resembling the previous data app to execute.

[51] https://regulatorycomplianceproject-sandbox.mxapps.io/

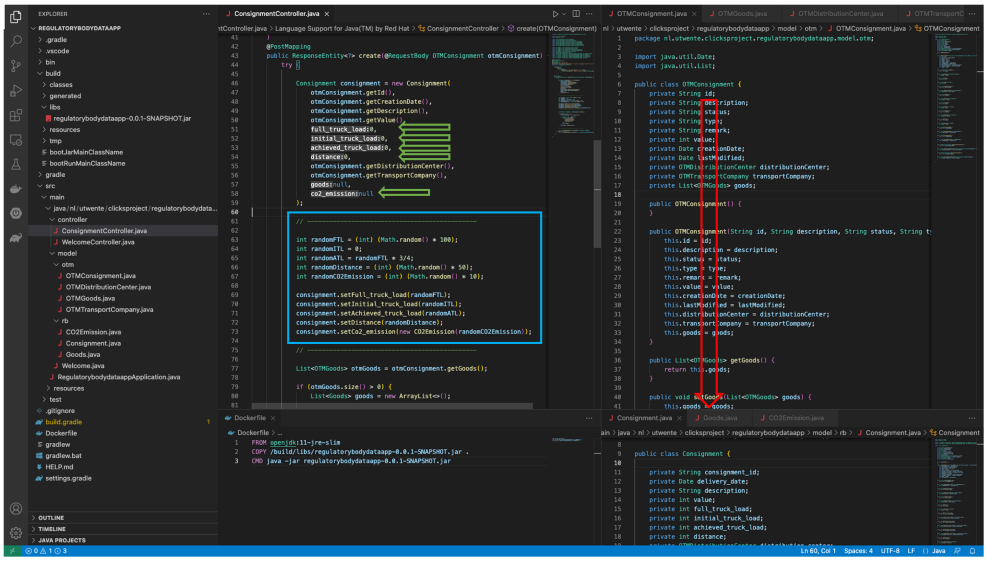


Figure 51 IDS Data Apps Prototype – Regulatory Body Source Code

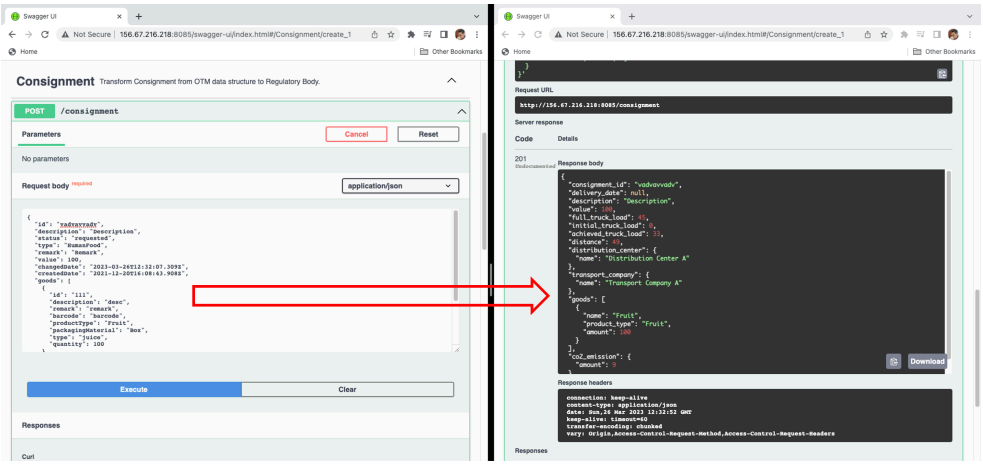


Figure 52 IDS Data Apps Prototype – Regulatory Body Swagger API Documentation

We developed this data app to transform data from OTM-compliant consignment data to a data structure that is supported by a custom Regulatory Body backend system. As shown on the left side of **Figure 51**, this works by receiving the OTM-compliant consignment data on a function, which is called create() and exposed as a web service operation with the "/consignment" endpoint, and then returning to the caller the needed format. What differs with the previous OTM Converter data app is that, in this case, not all attributes required by the custom backend system are available on the incoming data, as indicated by the green arrows. Therefore, to fill in this gap, we supplemented the function with logic to generate random numbers, shown inside the blue box. **Figure 23** demonstrates the "Data Aggregation" application function. In a real-world setting, this part might be replaced with a logic

of calling or consuming an external web service. As a result of this function, this data app enables its user to make up missing attributes. The resulting interface to invoke this function is presented in **Figure 52**.

7.2.2. Developing IDS Data App with eMagiz iPaaS Platform

The concept of an IDS Data App is defined as a small containerized application that has a specific data processing or transformation functionality that can be accessed by an IDS Connector through clearly defined web service endpoints for data input or output (refer to requirements IDA_1 and IDA_2). This means that any kind of containerized applications that provide such functionality and can be accessed by the IDS Connector can be seen as IDS Data Apps. eMagiz fits this purpose since it allows its users to capture, design, create, deploy, and manage the system integration of their IT landscape as well as message routing and transformation in a low-code manner. Low-code, in this context, refers to minimizing the effort for users to develop, configure, and set-up the integration components (e.g., connectors to transform and route message flows) using traditional syntactical coding. Deploying the integration runtime using eMagiz means deploying containerized application runtimes on a deployment environment (i.e., AWS cloud or the on-premise server that has Docker running in it) based on the integration landscape designed and created in the previous phases. In general, eMagiz supports 3 types of integration patterns, namely, Messaging, API Gateway, and Event Streaming. From there on, the deployed runtimes, respective to the selected integration patterns, will then expose web service endpoints that can be invoked by users or systems registered within the integration landscape. In this thesis, we demonstrate the usage of the eMagiz platform to design and deploy such a data transformation and message routing app with the API Gateway integration pattern.

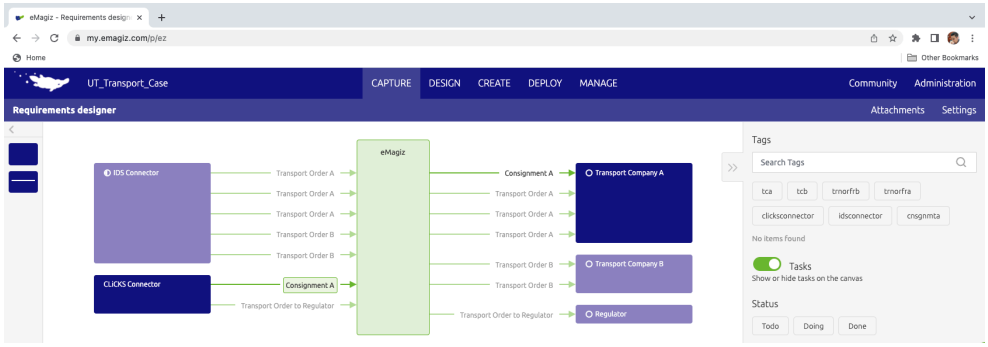


Figure 53 eMagiz Platform Create Phase – System Integration Landscape

In this section, we try to develop a simple eMagiz runtime for data transformation using eMagiz's API Gateway pattern to be integrated into the data flow of the IDS Connector prototype (again, refer to requirement IDA_2). The integration development within the eMagiz platform starts from the Capture phase. As shown in **Figure 53**, in this phase, we specify the systems involved in an integration

landscape and message types used by the communicating systems using the integration flows that are connecting them. In this landscape, we have defined several systems to interact with one another. To provide an example, we highlight two systems here, namely, the CLiCKS Connector (on the left side) and Transport Company A (on the right side). From the CLiCKS Connector, we draw an integration line that goes to the eMagiz API Gateway (green block in the middle) carrying the Consignment A message. From the gateway, we draw another integration line carrying the same message that goes to Transport Company A. This tells us that the CLiCKS Connector is given the permission to invoke an operation endpoint exposed by the API Gateway to make a request to (and expecting a response from) the Transport Company A.

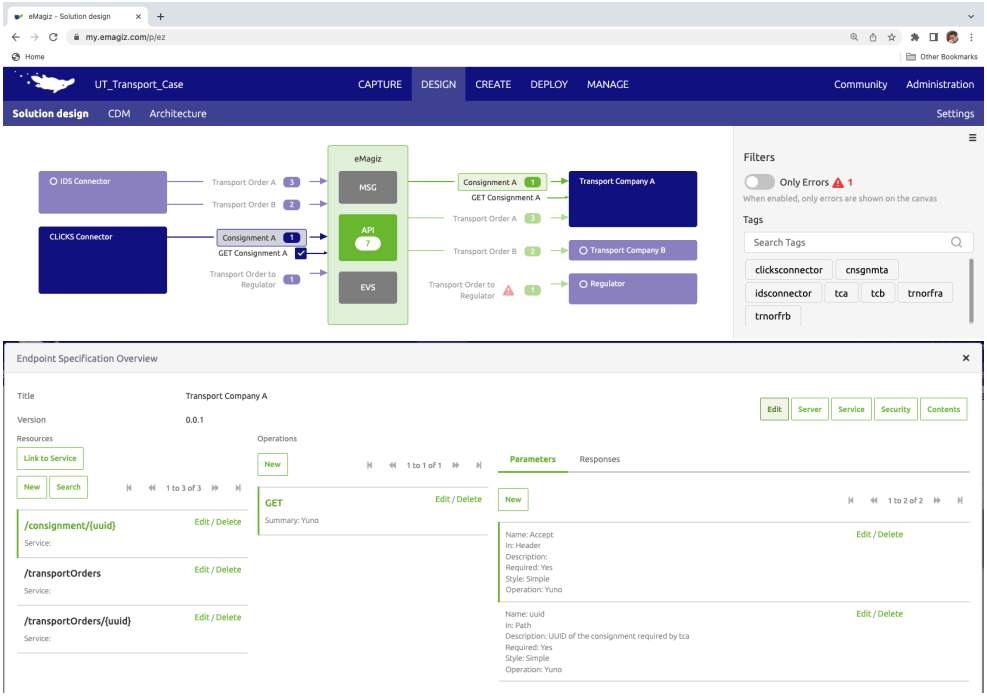


Figure 54 eMagiz Platform Design Phase – Endpoint Specification Overview

In the Design Phase, we further specify the operations supported by the target systems (i.e., systems that the requests are going to). We also specify which of these operations can be invoked by the source system (i.e., systems that the requests are coming from). Considering the previous examples, we consider the CLiCKS Connector to be the source system and Transport Company A as the target system. Next, as shown in the bottom part of **Figure 54** we specify the endpoints exposed by the Transport Company A system, along with the HTTP methods being used. This way, we can give permission to CLiCKS Connector to send a consignment message to this endpoint by giving check marks on the integration line coming from it.

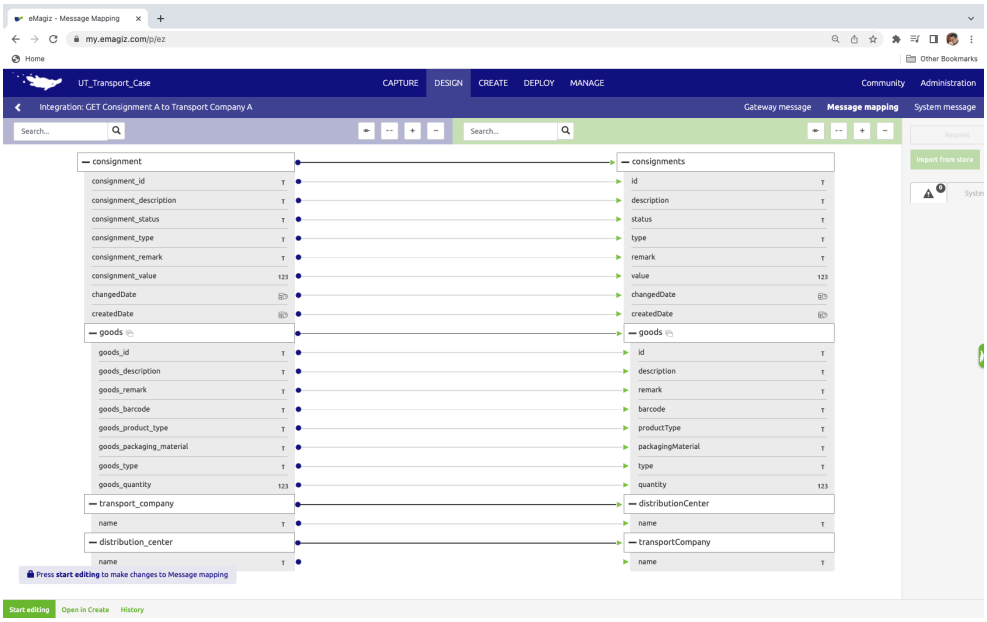


Figure 55 eMagiz Platform Design Phase – Response Message Mapping

The next step is to configure the message mapping for this request. An HTTP call consists of a request message and the response message. In this example, we do not send a message in the request, other than the id of the consignment that is attached to the endpoint's PATH parameter. Therefore, we only configure the response message coming from the target system. In this example, we try to transform the response message that is returned from the target system (blue system on the left) to the one that will be exposed on the API Gateway and consumed by the source system (green system on the right). The format used by the target system follows the snake case style, whereas the one required by the source system complies with OTM, which follows the camel case style. In **Figure 55**, we demonstrate how such a rather simple transformation can be performed within the eMagiz platform.

Upon wrapping the configurations in the Design Phase, the next step is to create the designed integrations in the Create Phase. The left side of **Figure 56** illustrates the overview of the integrated target systems with the API Gateway that will be created by eMagiz. Here, the user can give a check-mark on the Consignment A integration line, telling eMagiz to create the exit gate to the endpoint of this target system from the API Gateway. Afterward, the user can configure the exit gate going to this target system by double-clicking the GET Consignment A operation. This redirects the user to the page on the bottom part of **Figure 56**, and then set up (1) the accept header (required for the consignment's UUID specified for the request's PATH parameter) and (2) the HTTP Outbound Gateway (written as the send.cnsgnmta on the right end of the flow).

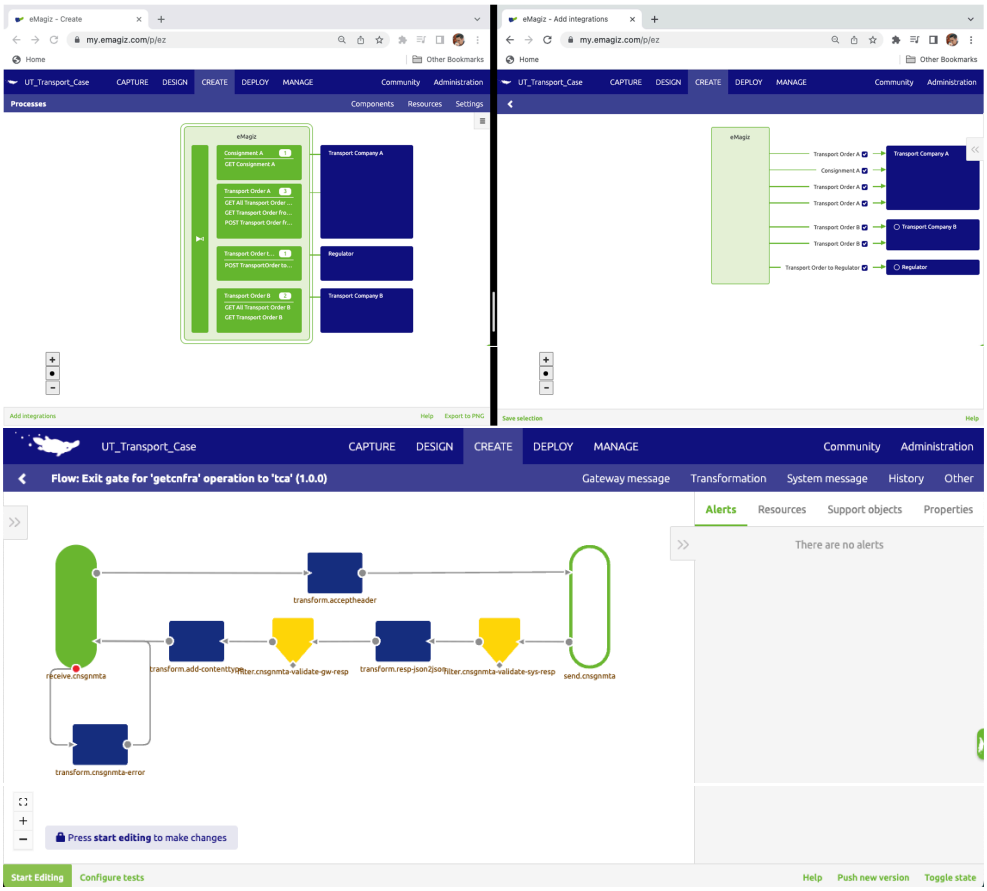


Figure 56 eMagiz Platform Create Phase – Adding Integration and Configuring Exit Gate

Lastly, in the Deploy Phase, we release the created integration solution to a deployment environment. In **Figure 57**, after we made a release and deploy it to the deployment environment, we can see the deployed containerized application that acts as the API Gateway. In this example, we specified that the released API Gateway is deployed on eMagiz's cloud environment that is provided by AWS. Another option is to deploy it on the user's own on-premise environment (e.g., local or VPS) that has Docker installed and running.

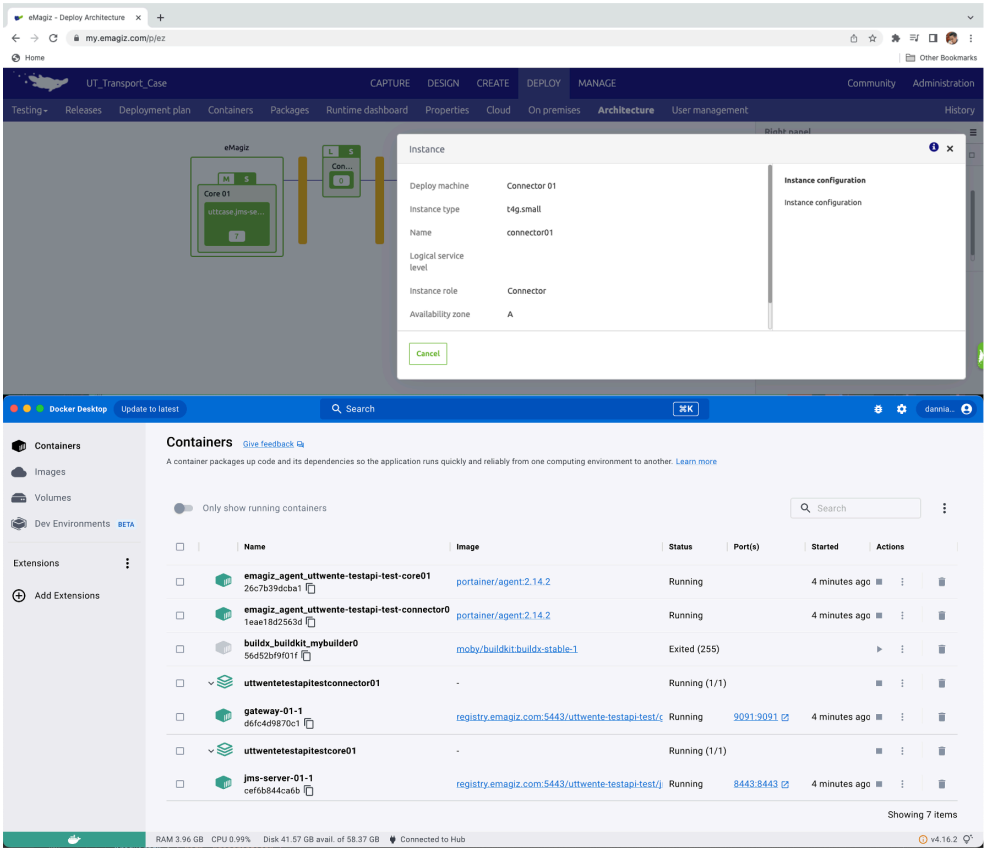


Figure 57 eMagiz Platform Deploy Phase – Deployment Architecture and Docker Environment

In the bottom part of **Figure 57**, we provide an example of how this designed integration solution looks when deployed on a local Docker environment. From this point, the deployed API Gateway container exposes its endpoints that are accessible by the CLiCKS Connector to perform the operations described above. This, then, concludes the development of an IDS Data App with the eMagiz platform.

7.3. The Development of a Connector Store Prototype

In this stage of the prototype development, we demonstrate how the application component responsible for service and resource discovery in a data space, as well as for supporting data space participants' onboarding, can be developed. As mentioned previously, adhering to the principles of the IDS, the Connecor Store is developed to take the role of and extend the functionalities of, the Metadata Broker. According to the proposed design, the Connector Store combines multiple application components and interfaces, comprising a front-end application, APIs,

and a triple store database. In the following sub-sections, the development activities involved in the Connector Store prototype will be discussed.

7.3.1. Instantiating the Connector Store Ontology: From Conceptual Model to Operational Ontology

As illustrated in **Figure 30**, the internal system of a Connector Store necessitates the instantiation of an operational ontology, serialized in OWL, based on the proposed preliminary conceptual model to be stored in and exposed by a persistent triple store database. To do this, we used Protégé, which supports the design of an OWL ontology and the verification of its axioms. Then, ontology individuals are, created manually by referring to the objects and terms listed in the OSRD in **Table 35** to reduce initial development complexity. **Figure 58** depicts parts of the operational ontology design using the Protégé tool. **Figure 59** shows a graphical visualization of the ontology with the OntoGraf plugin for Protégé, and showcases the **(a)** triples associated with an IDS Connector and **(b)** triples associated with an IDS Data App.

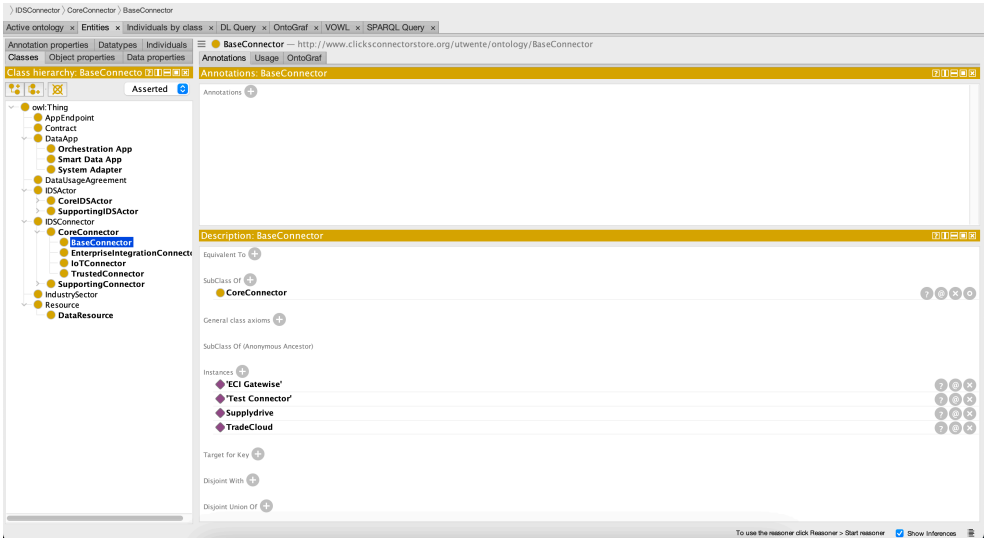


Figure 58 Protégé – Visualization of Axioms of the Connector Store Ontology

Next, the OWL model of the Connector Store ontology[52] is uploaded to a triple store database and made available for further querying. The triple store database here is responsible for persisting the metadata (regarding IDS Connectors, IDS Data Apps, data resources, data space participants, etc.) that are represented in RDF triples. This thesis uses Apache Jena Fuseki Server as the triple store, in contrast to an earlier work that uses TriplyDB due to the fact that the latter does not support INSERT query as opposed to the former (Firdausy et al., 2022a). This triple store database accepts an ontology graph represented in the Turtle format or its equivalent (i.e., N-

[52] https://raw.githubusercontent.com/danniarreza/connectorstoreontology/main/connectorstorev10.owl

triples, JSON-LD, or CSV, except the default OWL or RDF/XML formats), which allows exporting the output ontology to the target format.

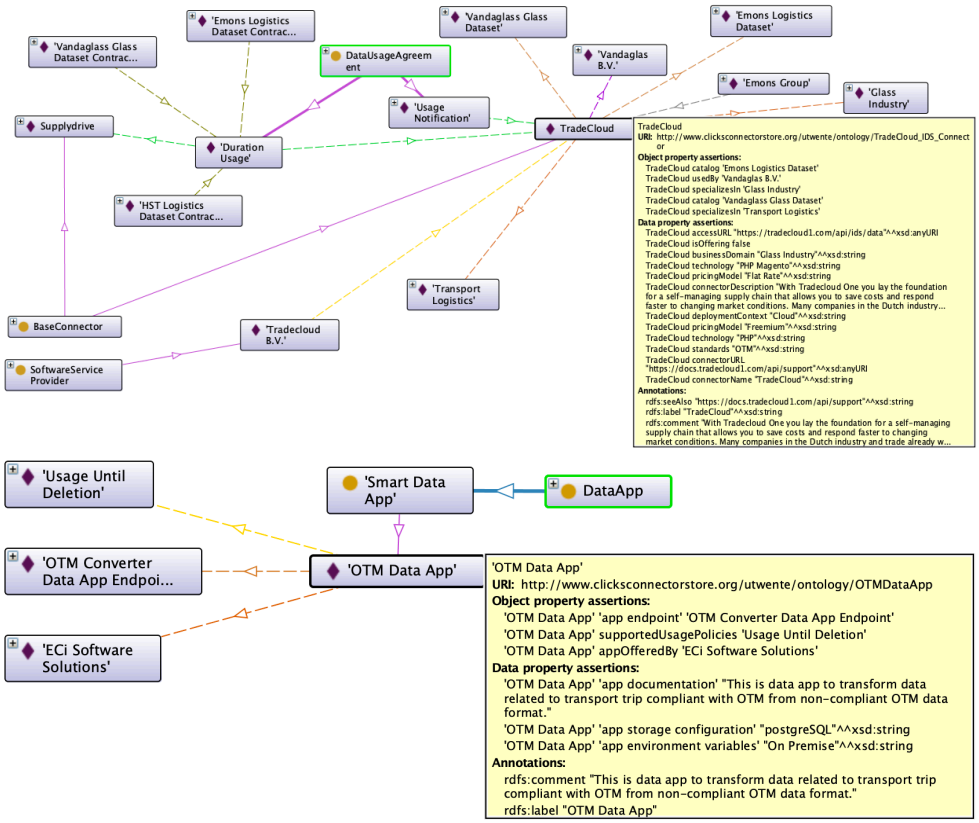


Figure 59 Protégé OntoGraf Plugin – (a) Visualization of IDS Connector Individual and (b) Visualization of the IDS Data App Individual

To streamline the deployment process of this triple store with the other prototypes, as well as to make it accessible online by the front-end application of the Connector Store, we deploy the Fuseki Server based on a Docker image[53] on the previously mentioned VPS. Next, we formulated some queries to retrieve metadata describing the data connectors offered by software and service providers and metadata relating to data sources provided by the data owners. **Figure 60** illustrates two SPARQL queries formulated to obtain a list of data resources and descriptions of an IDS Connector (irrespective of it being active or as an offering).

[53] https://hub.docker.com/r/stain/jena-fuseki

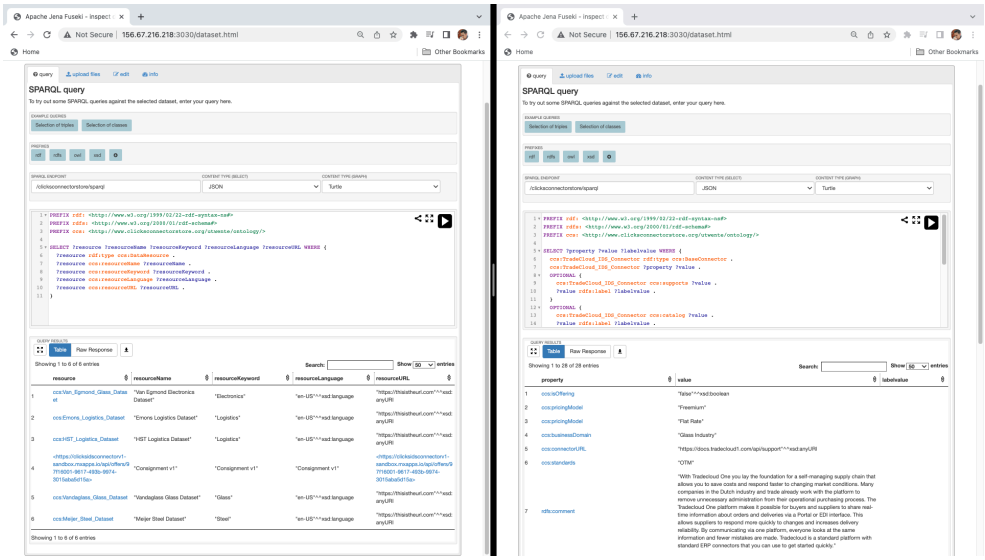


Figure 60 SPARQL Endpoint – Queries to Retrieve Metadata of a List of Data Resources and an IDS Connector

7.3.2. Developing Connector Store Front-End Web Application

The next phase focuses on the development of the front-end part of the Connector Store for interacting with the participants of an IDS ecosystem and their IDS Connectors. Similar to the IDS Connector prototype, the front-end web application was developed using Mendix[54]. **Figure 61** shows two visual interfaces supporting data space participants' discovery and onboarding process. The upper part of the figure illustrates how the Connector Store provides participants with a list of IDS Connectors currently active within a data space, along with their respective descriptions.

Whereas the bottom part shows how the Connector Store provides participants with a list of IDS Connectors offered by Software and Service Providers, along with the required information on how to request and/or deploy them to participate in a data space. The conceptual ontology proposed earlier characterizes the IDS Connectors according to their industry section, pricing model, offered data resources, supported standards, adopted technology, deployment context, and data usage policies (Firdausy et al., 2022b). This metadata could serve as filtering attributes to request the connectors once more instances are available.

[54] https://clicksconnectorstore-sandbox.mxapps.io/

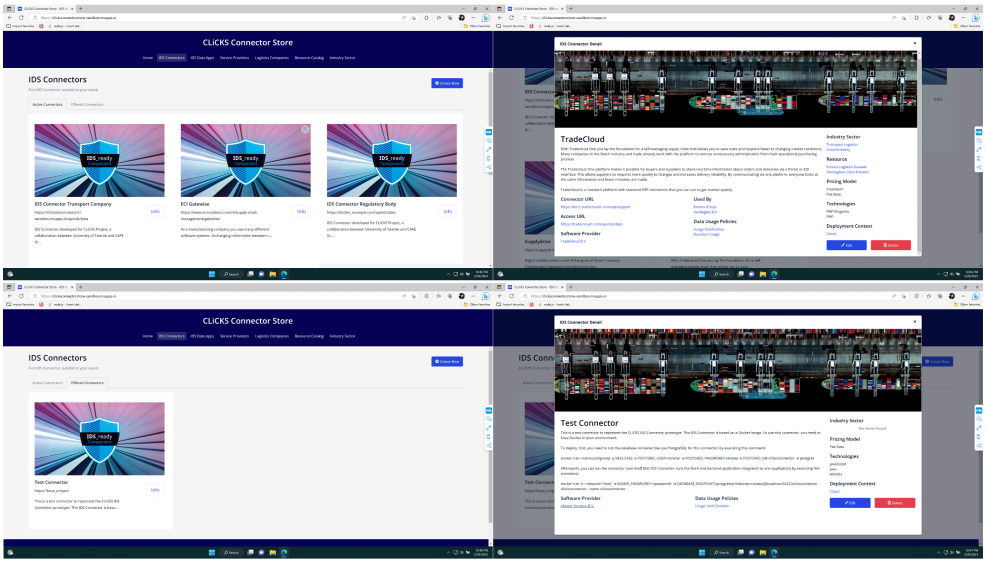


Figure 61 Connector Store Prototype – Active IDS Connectors and Offered IDS Connectors Metadata

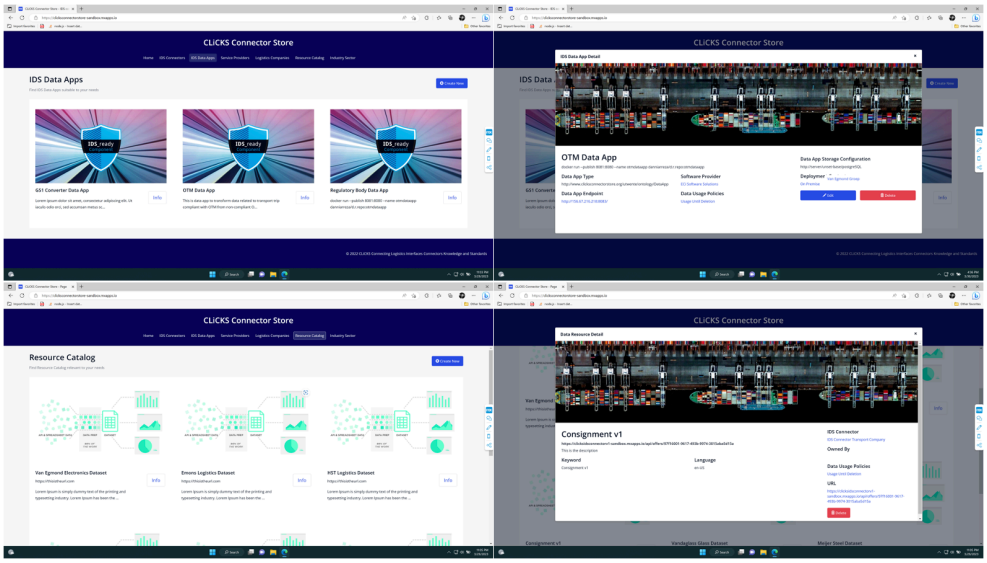


Figure 62 Connector Store Prototype – Offered IDS Data Apps and Data Resources Metadata

As shown in the upper part of **Figure 62**, there are 3 IDS Data Apps that are published. Taking one data app as an example, we describe an IDS Data App by its app documentation, app storage configuration, app environment variables, app endpoint, supported usage policies, and its software or service provider. Most of

these properties are taken from the IDS IM (IDSA, 2021f). Next to that, **Figure 62** also illustrates, on its bottom part, the menu where data space participants can browse through a list of data resources and the details of a particular data resource. Additional metadata is also made available to further describe data sources, for instance, the usage policies constraining the data usage, data representation language, and keywords related to the data content. Additionally, some properties describing the data resources, along with the IDS Connectors, were made available as hypertext reference (href) links. This is to maximize the advantage of following the Linked Data principles, which identify subjects and objects with HTTP URIs and enables associating with one another through their URIs to leverage resource discoverability on the users' side. When selecting one of these resources, a page will be presented to the user, listing several properties, such as the UUID of this resource that can be used by data users to initiate the resource request and contract negotiation with the data owner using their IDS Connectors as explained in **Sub-Section 7.1.4**.

7.3.3. Developing Connector Store Web Service API Integration

Finally, the APIs to facilitate metadata publication services are also implemented and made available for the participating IDS Connectors. **Figure 63** depicts the REST API documentation of the Connector Store. By the time this thesis is written, 3 essential endpoints are exposed. The first endpoint, indicated by the "/ids/data" endpoint, serves as the touchpoint for other IDS application components in the data space to get the self-description of the Connector Store as the broker. With this endpoint, IDS Connector can validate and retrieve the identity of the Connector Store before the connector registers itself to the store. In addition, this is also the endpoint where IDS Connectors can retrieve the metadata of resources that are managed by the Connector Store, as illustrated in **Figure 42**. Next to this, the "/ids/connector" endpoint is provided to receive IDS Connectors' self-description. In other words, this endpoint is intended for registering and publishing the participants' IDS Connectors. Lastly, the "/ids/connector" endpoint is reserved for registering IDS Connectors' resources to be offered to other participants in a data space. In this documentation, one can also observe the data structure used for this operation, which adheres to the IDS Connector data model shown in **Figure 25**. The REST API documentation of the Connector Store is publicly available for evaluation and testing in this link[55].

[55] https://clicksconnectorstore-sandbox.mxapps.io/rest-doc/api/

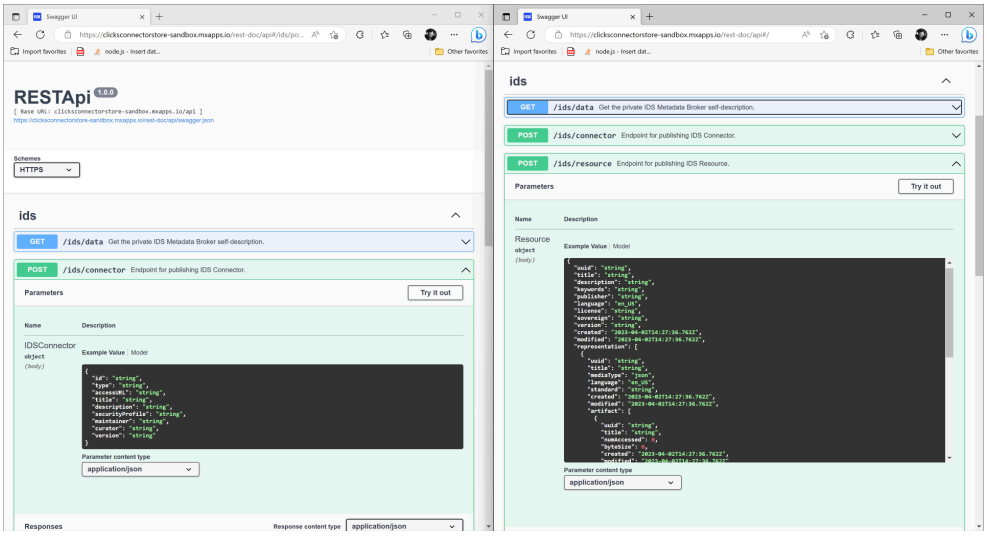


Figure 63 Connector Store Prototype – IDS Connector and Resources Metadata Publication Interface

7.4. Summary and Conclusion

This chapter covers the development of the prototypes based on the design and specifications discussed in the **Treatment Design** phase. The development focuses on three main application components for a logistics data space, namely the prototypes of IDS Connector, IDS Data App, and Connector Store. Along with describing the development process, the first section demonstrates how an IDS Connector can support data owners in managing data sovereignty and data interoperability through data usage policies enforcement and data app orchestration. The second section explains how IDS Data Apps can be built to handle data interoperability problems or even other data processing capabilities. Whereas the third section discusses how a Connector Store can be developed to improve participant onboarding and service discovery through semantic annotation of the offered data resources, IDS Connectors, and IDS Data Apps, along with how a Connector Store can be designed to connect with the already participating IDS Connectors. This requires validation with stakeholders involved in this project to demonstrate how these prototypes work, investigate what kind of scenario they might fit, and, assess how they can support data interoperability, data sovereignty, and service discovery for stakeholders in such a scenario. In the following chapter, we will discuss how these validation rounds are designed along with the data we collected from the panel of experts.

8 Validating Logistics Data Space Architecture and Demonstrator

This chapter aims to justify that the developed logistics data space demonstrator based on the proposed architecture can contribute to achieving the stakeholders' goals to manage data interoperability and sovereignty. To do this, **Section 8.1** discusses how the validation round is organized. Next, **Section 8.2** describes how the developed prototypes in the previous chapter work in practice to initiate a discussion for a use case. Then, **Section 8.3** elaborates on the application of the prototypes to address problems in these use cases, and its results will be discussed in detail in **Section 8.4**.

8.1. Validation Model and Research Methods

Wieringa (2014) specified that validation of such a treatment requires the definition of a validation model consisting of the model of the artifact interacting with a model of the problem context. In this thesis, the model of the artifact takes the form of the prototypes that were discussed in the previous chapter. Whereas the problem context is framed around data-sharing in the Dutch logistics sector, which considers data interoperability and sovereignty as the main requirements and stakeholder goals. Therefore, the next step is then to define (1) the mechanism to simulate the interaction between the artifact with its problem context and (2) the model of the problem context.

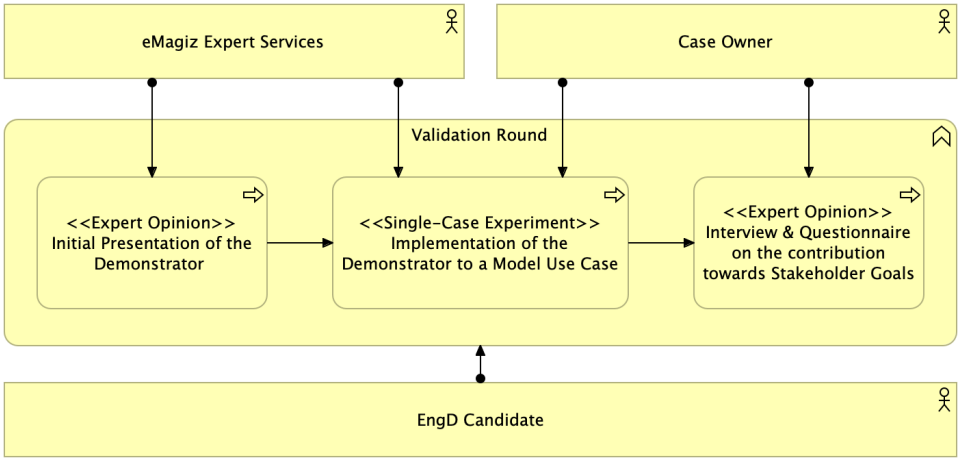


Figure 64 Validation Plan - Adopted Research Methods Viewpoint

In this thesis, we incorporate the expert opinion method, which is then followed by a single-case mechanism experiment. **Figure 64** illustrates the validation plan that

we carry out. The first round of the expert opinion works by presenting the demonstrator to a panel of experts from eMagiz Services B.V. Three experts from the Expert Services team (i.e., stakeholders listed in **Table 2**) were involved in the presentation, which aimed to (1) gain an understanding of how the demonstrator works and (2) come up with a relevant business case for the experiment. With this single-case experiment method, we aim to capture a reliable prediction of how the application prototypes can contribute (either positively or negatively) towards lowering barriers of data interoperability and improving data sovereignty aspects for the data space participants in this model of the problem context. To capture these results, again, we collect the opinion of the stakeholders involved in, or related to, the business case. **Table 37** presents the participating experts coming from mixed backgrounds spanning from industrial practitioners to academic researchers. A representative from SUTC, who also concurrently represents Transport en Logistiek Nederland (TLN) acting as the point of contact for experiences in the sector, also participated. Adhering to the validation plan presented in **Figure 64** earlier, in the next section, we will discuss the execution of the initial presentation of the logistics data space demonstrator to the expert panel.

Table 37 Composition of the Participating Expert Panel

ID	Role	Organization	Experience
E1	Expert Services Manager	eMagiz Services B.V.	- 12 years in Web Developer and Technical Consultant role. - 1 year in a Product Manager role, and 3 years in a Software Delivery Manager role. - 1 year in Expert Services Manager role.
E2	Product Manager	eMagiz Services B.V.	- 4 years in a Sales Consultant role. - 12 years in Product Consultant and Solution Delivery roles. - 6 years in Professional Services Manager and Product Manager roles.
E3	Case Owner	eMagiz Services B.V.	- 5 years in Technical Consultant and Solution Architect roles. - 3 years in Expert Services role.
E4	General Secretary – Policy Advisor	SUTC – TLN	- 15 years in Policy Advisor Innovation & Digitization and Secretary of Policy & Submarkets roles.
E5	Ph.D. Candidate	TU Delft	- 3 years of research on the topic of Business Models for Data Platforms. - 2 years in the Business & Technology Integration

8.2. Initial Presentation of the Logistics Data Space Demonstrator

For the experts to see how the proposed architecture of the logistics data space performs in a practical setting, first, its demonstrator has to be operationalized into a particular business case. For this purpose, we arranged a presentation session to demonstrate a mockup scenario regarding tax reporting of transport trip carbon emissions. Before the session, the experts were given an introduction document, explaining the background of this design project and a brief description of the relevant IDS components, as well as the business case scenario for the demonstration. The referred business case is explained as follows.

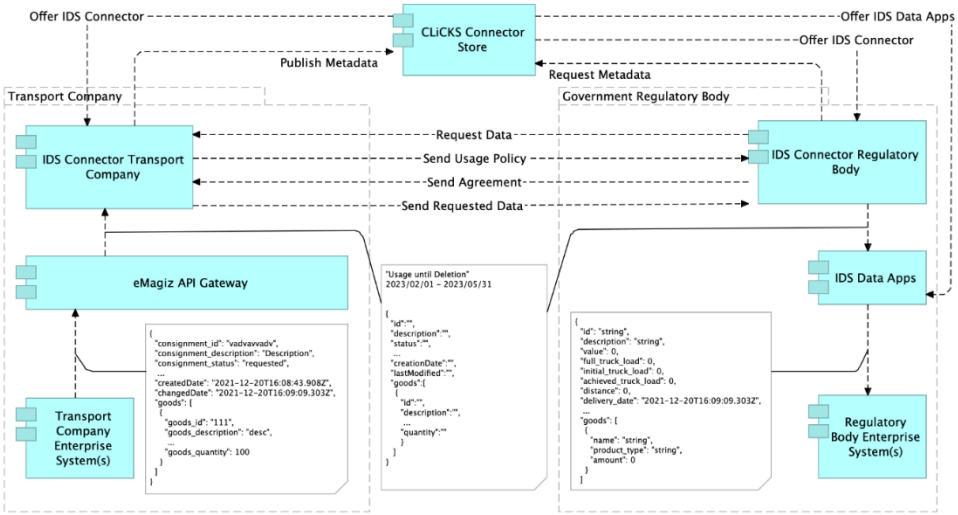


Figure 65 Demonstration Scenario – Mockup Transport Trip Carbon Emission Tax Reporting

Today's supply chain consists of collaborations and data-sharing among multiple parties. These collaborations, nowadays, aim to increase efficiency and environmental sustainability. Especially, in the transport logistics sector, reducing CO2 emissions becomes a prominent objective. National and international governmental bodies released a set of laws and requirements that should be fulfilled by actors in the supply chain. The Regulatory Body makes sure that all these laws and requirements are met. To enforce this, the Regulatory Body requires Transport Companies to report and share their transport trip data. Based on this data the Regulatory Body will calculate the tax that the company will have to pay. Sharing this data means opening up the possibility for the company to pay less tax due to the lower carbon footprint they produced, but also risks compromising their competitive advantage due to sensitive and confidentiality concerns. On the other hand, sharing them also requires data transformation and, possibly, data integration efforts from other data sources before they can be received by the Regulatory Body.

Therefore, in this demonstration, we try to solve this challenge by applying the IDS Connectors and IDS Data Apps for each of these actors that are brokered by a Connector Store. A preview of the system interaction and data flow landscape for this scenario is shown in **Figure 65**. This is the landscape that we will try to simulate in the demonstration. From there on we will brainstorm future opportunities and other possible business cases. From these validations, the thesis intends to collect expert opinions (either from eMagiz or other case owners) by means of questionnaires and interviews based on the questions presented in **Table 38**.

8.3. Implementation of the Demonstrator to a Model Business Case

The initial demonstration concludes that this mockup case resembles a project by one of eMagiz's clients. This client needs to share their transport trip data with the Dutch CBS for carbon footprint calculations, as well as with other partners for exchanging transport orders and proof of delivery. However, exchanging this data is complex due to the diverse definitions of data, processes, protocols, and systems adopted by every company. To alleviate this issue and make it possible to exchange all these data between many different systems via one uniform link, SUTC introduced the VESDI project[56]. This initiative promotes the use of OTM as the standardized data specifications for web service interfaces. Through a standardized specification, VESDI supports data types as follows: mobility (time, routes, numbers of vehicles), goods flow (shipments, origin, destination), and energy performance (CO2 emissions and energy consumption per kilometer).

For this project to work, the CBS adapts its own systems so that it can receive data for the transport survey via VESDI by means of a web service interface. As shown in **Figure 66**, transport companies have to fill in the road transport survey (i.e., enquête wegvervoer). One way to complete this is by manually filling out the online survey manually, but, this administrative work takes a lot of time and effort. Another way of doing this is by sharing the data directly from the companies' IT systems. Despite the less manual labor work, setting up the system integration between these diverse IT systems, and the supported message formats, are still a relatively expensive investment. To solve this, the VESDI project proposes the participating IT providers implement OTM, specifically the VESDI profile, in their systems. This way, these IT providers enable their clients (i.e., logistics companies) to share their data automatically with the CBS. As a bonus, their clients can then benefit from the possibility of exchanging data with other partners via OTM format.

[56] https://www.sutc.nl/en_US/het-vesdi-project

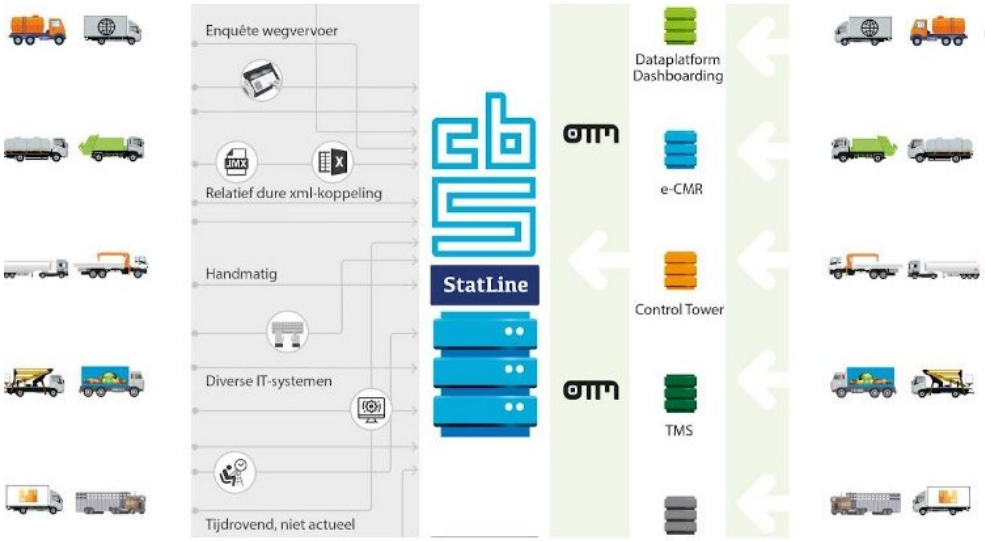


Figure 66 VESDI Project – Promotional Overview

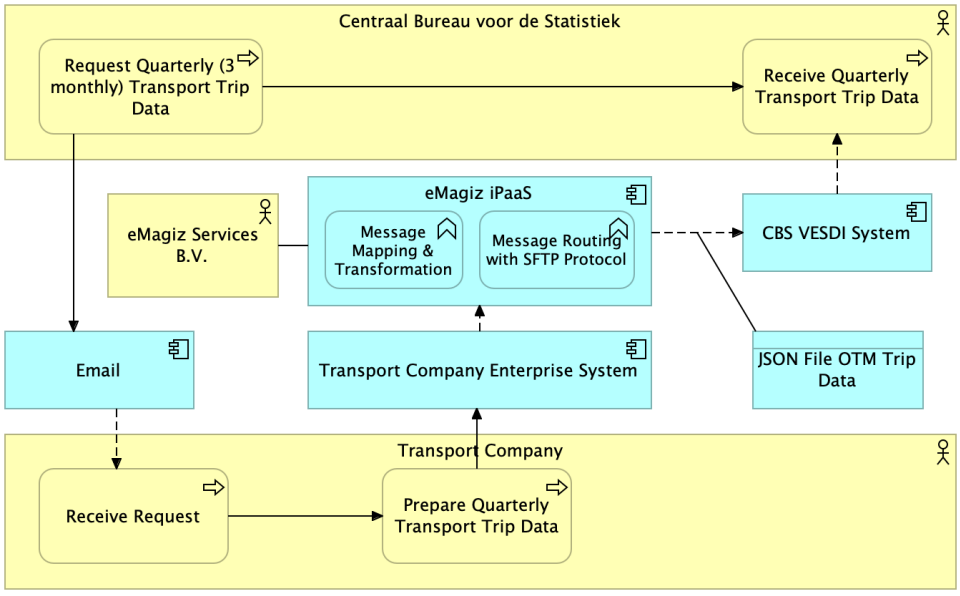


Figure 67 eMagiz VESDI Project – Baseline Architecture

From a short interview with an expert from eMagiz Services B.V., who acts as the case owner of this project, we can identify at least 3 involved stakeholders. First of all, there is the CBS which demands up-to-date data from the transport companies

for calculating the sector's overall performance and carbon footprint. Second of all, the transport companies themselves have to share their data with CBS, preferably, in the most efficient way and timely manner. Lastly, the IT providers provide the IT infrastructure and integration landscape to support this project. In this particular project, eMagiz Services B.V. acts as the IT provider. Relating it to the concept of a data space, such a scenario indicates that CBS is the data user and transport companies are the data owners. Meanwhile, the IT providers act as the service or software providers, who provide the necessary components or configurations for the data space to operate.

Figure 67 visualizes the business processes followed by the stakeholders, along with the current solution provided by eMagiz. First, every 3 months, CBS sends a request to transport companies (which is being done by email so far) for their quarterly transport trip data. One of the reasons for this is the data confidentiality aspect, specifically, because transport companies are not willing to fully share their competitively sensitive data. This tension calls for the application of the data usage policy enforcement offered by the IDS Connectors. Upon receiving this request, the transport company prepares its quarterly transport trip data from its enterprise system (i.e., select the period that they need to report and aggregate the raw data) to be sent to the CBS. In the current implementation, this is done by forwarding the selected data to an endpoint in eMagiz, where this data will be mapped to the VESDI-profile OTM format required by CBS while filling in some empty values (e.g., UUIDs and the company's office and warehouse coordinates). Next to this, for the VESDI project, CBS requires companies to send their data through the SFTP protocol as opposed to HTTP. Therefore, eMagiz also facilitates the use of SFTP protocol for this message in its routing to the CBS's system to be received by the CBS itself.

The nature of this scenario indicates two possible data exchange approaches. First, the fact that the CBS initiates the data request (via email), explains that such a scenario can be supported by the pull approach proposed by the sequence diagram in **Figure 26** and prototype implementation in **Figure 43** earlier. This approach can work, provided that the Transport Company as the data owner has prepared and uploaded their transport trip data beforehand and published its metadata on its IDS Connector, so then the CBS as the data user can make the request without the need of using the email. The caveat here is that, when the Transport Company has not uploaded their transport trip data, then the CBS can not make the request, forcing them to go back to making the request via email.

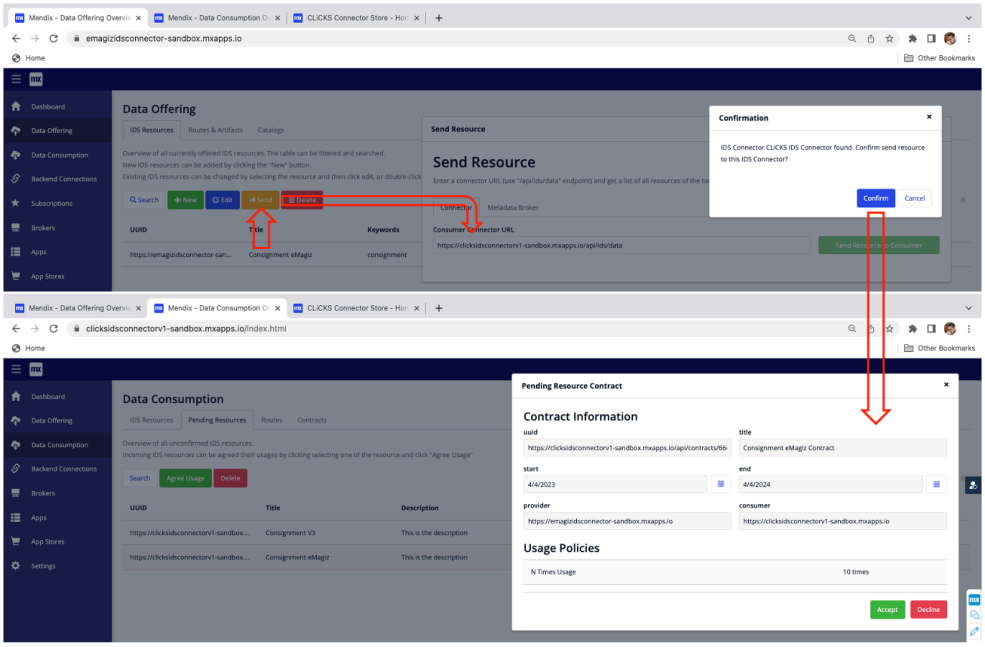


Figure 68 IDS Connector Prototype – Push Resource to Consumer IDS Connector

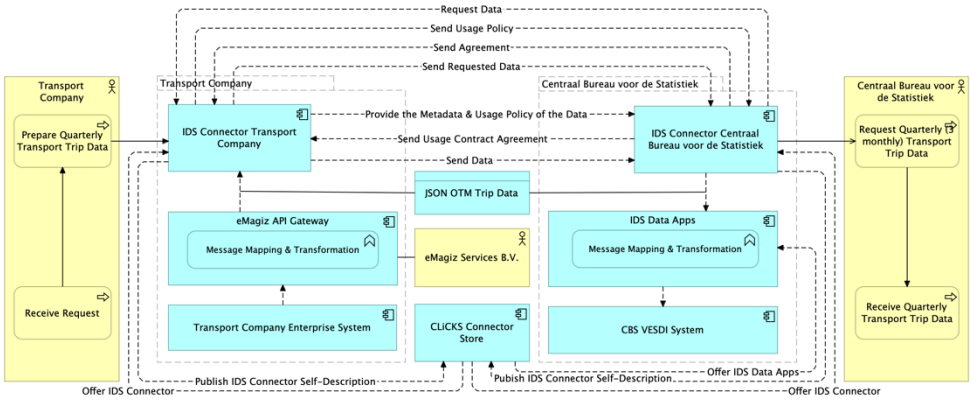


Figure 69 eMagiz VESDI Project – Target Architecture

Therefore, to overcome this limitation, we updated the IDS Connector prototype to also support push behavior. In the case that the CBS is making the request via email, then the Transport Company can act by pushing its data resource to the CBS's IDS Connector without needing the CBS to make a redundant request via its IDS Connector. For this use case, the top part of **Figure 68** illustrates how the data owner can now select and send an offered resource to a consumer connector. Upon confirming the send, this IDS Connector acting as the provider will send the

resource (along with its representations, artifacts, contracts, and usage rules) to the consumer connector. As shown in the bottom part of **Figure 68**, for the consumer connector to be able to consume this, the data user needs to provide his agreement to the usage contract of this resource. Thereupon, the consumer connector will send a contract agreement to the provider connector and then this resource, which was listed under the "Pending Resources" page, will be transferred to the list of "IDS Resources" accessible to the data user. From there on, we facilitated the two possible data exchange approaches for the VESDI project. As a result, reusing the data-sharing demonstrator described earlier in **Figure 65**, we realized a data-sharing landscape for the VESDI scenario as visualized in **Figure 69**.

8.4. Contribution of the Logistics Data Space Demonstrator towards Managing Data Interoperability and Data Sovereignty

To validate whether the proposed architecture, through the developed application prototypes, can treat data interoperability and sovereignty issues in this model business case, we capture the opinions and assessments from a panel of experts involved in or relevant to this project. After we demonstrated how the logistics data space demonstrator performs in front of the experts, we presented a questionnaire to them to capture their perceptions of the contributions of the architecture to satisfy stakeholder goals. The questions asked to the participants for this validation are based on the goals listed in the Motivation Viewpoint of IDS Adoption and Connector Store Implementation in **Figure 13**. One important note is that, with respect to the realization of the data sovereignty aspect, we omitted the last goal (i.e., acquire IDS-ready labels for organizational assets and software components) as it requires further investigation into a rigorous evaluation and certification process. Therefore, we focus on the more technical goals in this thesis, which are the first four goals in that viewpoint. We listed these goals along with the derived questions in **Table 38**.

Table 38 Validation Pointers based on the Goals in the Motivation Viewpoint of IDS Adoption and Connector Store Implementation

No.	Goals	Questions	Score	Opinions
1	Facilitate data transformation between standards and data formats	To what extent can such an architecture, through its demonstrator, positively contribute to facilitating data transformation between standards and data formats?		
2	Stimulate data findability, accessibility, interoperability, and reusability	To what extent can such an architecture, through its demonstrator, positively stimulate data findability, accessibility, interoperability, and reusability?		

3		To what extent can such stimuli motivate data space participants to devise new business models and value co-creation?		
4	Enable quick onboarding to a data-sharing ecosystem	To what extent can such an architecture, through its demonstrator, enable potential participants' quick onboarding to a data-sharing ecosystem?		
5	Enable data owners to be self-determined regarding the usage of their data assets by data users	To what extent can such an architecture, through its demonstrator, enable data owners to enforce usage control over their data assets by data users?		
6		To what extent can such an architecture, through its demonstrator, support data owners to maintain the confidentiality aspect of their data assets during and after the data exchange?		

Following the demonstration of the prototype with the expert panel, we asked their judgments through these 6 questions listed in the questionnaire. In **Table 39** we present the quantitative assessments perceived by the experts. These assessments are measured using a Likert Scale ranging from 1 to 5. From the obtained scores, we observed that the average and standard deviation for each question ranges from 3.2 to 4.2 and 0.45 to 1.14 respectively. This indicates that the respondents shared a relatively common assessment for questions 1, 2, 5, and 6. We can see that these five respondents agree that the architecture of the logistics data space, through its demonstrator, can (1) positively contribute to facilitating data transformation between standards and data formats, (2) positively stimulate FAIRness of data, (3) enable data owners to enforce data usage control, and (4) support data owners to maintain confidentiality aspect of their data assets. Such a consensus, on the other two questions, has not been reached. In other words, there is still a significant uncertainty among respondents that the architecture of the logistics data space, through its demonstrator, is able to (1) motivate data space participants to devise new business models and (2) enable participants' quick onboarding to the ecosystem. Despite that, some additional remarks in regards to validating the achievement of each of the stakeholder goals were also obtained in the interview that followed the demonstration sessions, which should be taken into consideration for the implementation of the proposed architecture in the real-world setting.

Table 39 Questionnaire Results on the Architecture's Contribution Towards Achieving Stakeholder Goals

No.	Questions	E1	E2	E3	E4	E5	AVG	S.DEV
1	To what extent can such an architecture, through its demonstrator, positively contribute to facilitating data transformation between standards and data formats?	4	3	3	5	4	3.8	0.84
2	To what extent can such an architecture, through its demonstrator, positively stimulate data findability, accessibility, interoperability, and reusability?	3	3	4	4	4	3.6	0.55
3	To what extent can such stimuli motivate data space participants to devise new business models and value co-creation?	3	4	3	2	5	3.4	1.14
4	To what extent can such an architecture, through its demonstrator, enable potential participants' quick onboarding to a data-sharing ecosystem?	5	2	3	3	3	3.2	1.10
5	To what extent can such an architecture, through its demonstrator, enable data owners to enforce usage control over their data assets by data users?	4	4	5	4	4	4.2	0.45
6	To what extent can such an architecture, through its demonstrator, support data owners to maintain the confidentiality aspect of their data assets during and after the data exchange?	3	4	4	4	3	3.6	0.55
		AVG		3.6	S.DEV			0.77

Feedback from the Manager of Expert Services of eMagiz Services B.V.

E1, as someone who is less knowledgeable of the business case but experienced in eMagiz's Software Delivery team, mentioned that the IDS Connector prototype, in several parts, has replicated the functionality of eMagiz API Gateway: it is capable of supporting message routing configuration and service orchestration. However, the development effort of the data processing services (i.e., IDS Data Apps), as demonstrated in this case study, still requires raw-code and manual deployment approaches as opposed to the low-code style offered by the eMagiz platform. This also replicates the real-world, situation where each party will have different IT developers and solution providers to a certain extent.

Feedback from the Product Manager of eMagiz Services B.V.

E2, responsible for the solution delivery and feedback acquisition on behalf of eMagiz's Product Manager role, commented that the prototypes work as stated in the presentation before the demonstration itself. Although, he noted that the preposition of the logistics data space demonstrator can facilitate participants' quick onboarding to a data-sharing ecosystem is still questionable. Its success will also depend on what kind of data, business cases, and business models the data-sharing ecosystem has to offer. This was also confirmed by a keynote presentation at the Data Spaces Symposium 2023 in The Hague[57] which mentioned that developing and scaling up a data ecosystem is indeed a chicken and egg problem.

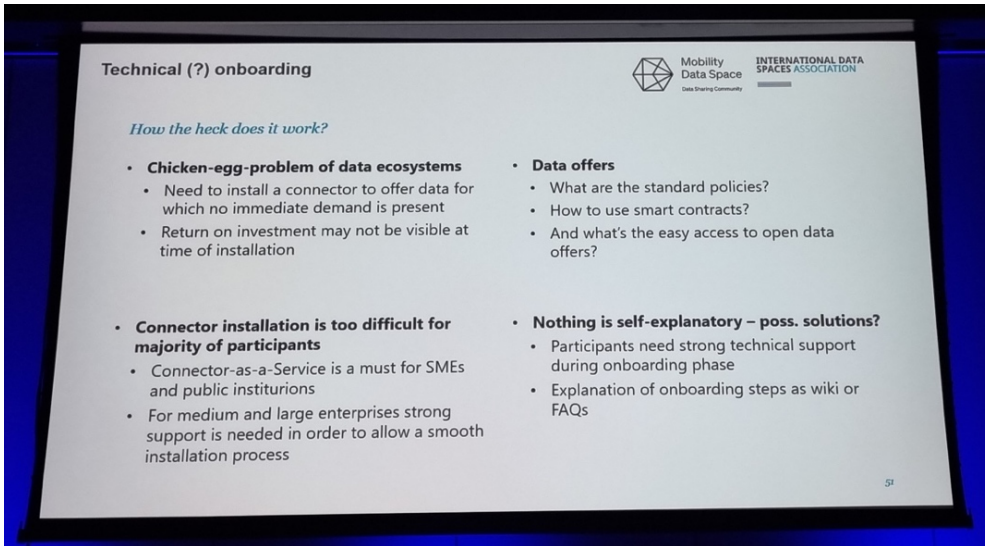


Figure 70 Data Spaces Symposium 2023 in The Hague - Technical Onboarding Dilemma

[57] https://internationaldataspaces.org/data-spaces-symposium/

As shown in **Figure 70**, on the one hand, participants need to install a connector to offer data in a data space for which no immediate demand might be present. On the other hand, if such demand is not visible, then the return on investment may not be visual as well. Nevertheless, the demand for an artefact or a mechanism to facilitate participants' onboarding is indeed present. Next to this, E2 also mentioned that the selling point of this kind of prototype demonstration also depends on how its designer presents it to the audience. Since the concept of data spaces prescribed by the IDS is relatively new to him, the results obtained from him could have been different if the demonstration was divided into parts (i.e., demonstrations specific for the IDS Connector, IDS Data Apps, and Connector Store), so that there can be a discussion for promising business models that might attract participants onboarding. However, this limitation occurred mainly because of the limited time slot available for the demonstration.

Feedback from the Case Owner of the VESDI Project in the eMagiz Services B.V.

E3, acting on behalf of the case owner of the VESDI project from eMagiz Services B.V., provided a more elaborated expert judgment. The first thing that comes into mind is the additional complexity brought into the data-sharing landscape due to the introduction of IDS Connectors for each party. Such a complexity issue points to the possible amount of message routings and data usage policy enforcements that the CBS will have to manage in the future if the CBS expects a huge amount of datasets coming from hundreds of transport companies, let alone multiplied with the amount of reporting rounds happening in a year. Next to that, getting a critical mass to adopt the IDS-compliant application processes for a single data exchange can be a huge challenge as it requires quite a lot of back-and-forth communication between systems and not many systems can facilitate this yet at the moment. Considering the complexities and uncertainties in the configuration of the integration solution that is already present in the real-world practice, it would be easier to stay close to the baseline architecture, where the involved parties can just configure eMagiz iPaaS to route the message from TC directly to the CBS system. However, the drawback of this approach is that the participants will lose the opportunity to enforce data usage policies, which defeats the purpose of adopting the IDS design principles in the first place.

Another foreseen solution to overcome the added complexity issue while maintaining the data sovereignty capability is to revamp the eMagiz iPaaS platform to act as an IDS Connector so that it supports the execution of such data usage policies. However, for this to work, it means that the eMagiz iPaaS will need to store (or have persistent storage for) the instances of the data usage policies defined by data owners. If realized, this will contradict the value proposition offered by eMagiz as an integration platform for their customers to not store any data nor metadata that binds the data being transported. Besides, as the requested data resource leaves the system, there is still a possibility in the future for a data breach that happens in the data user's internal environment that is already outside the reach of the data owner. This issue has also been confirmed by our discussion with a representative from Sovity who we encountered at a symposium. He, along with the paper by

Zrenner et al. (2019), testified that addressing such a loophole will require a policy enforcement point implemented in the data users' enterprise systems and operating systems environments and not only on the application level as demonstrated by this prototype and Sovity Dataspace Connector implementations.

The same concern also applies to the resource offering and consumption functionalities that are currently supported by the IDS Connector prototype. These functionalities will be challenging for the eMagiz platform to adopt if it has to act as an IDS Connector since it means that the platform has to store (at least temporarily) the metadata describing these resources. Therefore, to adopt such a data space design for this particular business case, considering the main requirements of data interoperability and sovereignty, this architecture (presented in **Figure 69**) is still perceived as the most suitable one. The final remark provided by E3 is that this proposed architecture, through its demonstrator, can support data owners to enforce data sovereignty. However, by the time this thesis is written, he has not encountered an incident (or a major one), which causes losses (e.g., financial, etc.) for the data owners that can justify the demands for the enforcement of usage policies.

Feedback from the Policy Advisor of TLN and General Secretary of SUTC

E4 shared his knowledge of the VESDI project from the perspective of the SUTC as well as his view on the demonstrator when applied to such a case. First of all, he clarified that the obligation of the Dutch transport companies to report their transport trip data to the CBS is, at the moment, mainly intended to contribute to the performance evaluation of the national logistics sector. Calculating and predicting the carbon footprint generated from land freight transport is indeed part of the plan, but the actual implementation is still yet to come in the future. With this regard, transport companies are obliged to file their report, and they are subject to penalty or otherwise.

Despite so, E4 testified that the logistics data space demonstrator, through the demonstration session, works well and the intention to solve data sovereignty and interoperability issues on runtime is easy to understand. With respect to the VESDI case, CBS is a governmental institution that gathers data for national statistical reasons means that it is a trusted entity. This is backed up by the fact that CBS is forbidden to make use of the data without any clear intentions and consent of the transport companies as the data owners. However, as the VESDI project sponsors the OTM as the data standard, it advertises the exchange of transport trip data among logistics partners to be more accessible and desired in the future. The growth of such a demand will then resonates with the concern of data sovereignty among participants of the ecosystem. Companies tend to not easily trust the competition. Most of the time, what holds companies to share and exchange data with other parties is the uncertainty of what their data will be used for the receiving end. To convince them, any intention for data sharing or exchange has to be backed up with a solid purpose and business model of why the transaction needs to be carried out.

Considering this, E5 saw that the data usage policy enforcement can help to provide more control to the data owners regarding the utilization of their data by other participants. Next to that, having the data offering definition and data request contract agreement process preceding the data exchange can help to technically specify and enforce the intention to use the data, thus, bringing more confidence for partners to share their data and form a partnership to optimize their collective performance.

Feedback from the Ph.D. Candidate of Digital Platforms and Data Marketplaces from TU Delft

E5, as a researcher in the field of digital platforms and data marketplaces, provided his inputs for this demonstrator from the perspective of the prospective business models. Through the demonstration, he testified that the infrastructure constituting this demonstrator shows a promising fit for the data-sharing infrastructure for European Data Market promoted by the European Commission for companies to securely trade, store, and access high-quality data assets. One of the goals of such an initiative is to provide an infrastructure for creating sales contracts, matching demand and supply, as well as supporting transactions for the transfer and payment of data assets as the sold products (Bergman et al., 2022). For such a goal, the Connector Store is perceived by E5 to be capable of providing information regarding offered data resources and services, participating IDS Connectors, as well as IDS Data Apps. While the demonstration itself may not have placed a significant focus on data reusability, the underlying architecture of the Connector Store indeed holds the potential for enhancing data reusability. In addition, this inherent reusability allows organizations to efficiently repurpose the offered IDS Data Apps for various data processing tasks, minimizing redundant development efforts, and maximizing the value derived from data. Thus, the store shows a promising future to enable demand and supply matching in a data-sharing environment, should it be further enriched with more supporting functionalities.

Next to that, he testified that the prototypes in this thesis have demonstrated their capabilities to support data users to quickly set up and establish ad-hoc usage-constrained data exchanges with data owners. In case of data transformation or other data processing capability is called upon, IDS Data Apps can be retrieved from the Connector Store to facilitate transformation between different standards and formats. Enabling this quick connection with unacquainted partners opens up the possibility for data users to explore new uncharted business opportunities. This, in the current practice, is proven to be challenging due to the fact that setting up a connection in a conventional data exchange takes both time and resources to configure the integration between partners' enterprise systems. Despite that, several consideration points regarding the IDS Data Apps emerged. The first one is the question of the effort needed to develop such mappings between formats and standards. Secondly, what will be the incentive models to develop such mapping, and who will have the responsibility and credits for developing it? These questions need to be examined in future research. Once this data interoperability issue is solved, there is still tension between partners participating in a data space regarding

what their data will be used for. For this, the data usage policy enforcement is aimed to provide the solution. E5 testified that the data usage policy prescribed by the IDS specification can become one of the prominent building blocks in the future for a secure and trusted data marketplace. One of the usage rules, the "Usage During Interval" pattern, is relevant and can be used for an exchange of time-sensitive data, in which the intrinsic value of the data fluctuates over the span of time.

E5 believes that the data space demonstrator can motivate participants to devise new business models and co-create value. The success of such a data-sharing ecosystem depends on (1) its adoption rate by partners, (2) the amount of participating members, and (3) its significance to the other members. Therefore, to accelerate the adoption rate by partners, there is a need to attract vocal partners, or partners with major significance in a supply chain, to adopt this ecosystem so that their smaller partners are willing to follow. Next to that, technology providers, acting as the operating company, are needed to establish the infrastructure for a data marketplace and enable data exchange-as-a-service. Through the Connector Store established as a shared service, data marketplace operators can then focus on their core value creation activities (e.g., curating high-quality datasets and providing data analytics services). From the perspective of data owners and data providers, the Connector Store can support them to explore secondary business models (e.g., data monetization) that was previously unexplored. An example scenario is a data provider with historical weather data can now offer tailored datasets or insights to data users in specific industries (e.g., agriculture or renewable energy) who can utilize this data to optimize their operations and reduce risks. Next to that, the Connector Store can enable data users to access and integrate data from various providers to generate valuable insights. For example, a retail company can develop targeted marketing campaigns, optimize store layouts, and improve customer satisfaction that leads to increased revenue and a more competitive business model by combining data from multiple sources (e.g., foot traffic data, sales data, and social media sentiment analysis). Third-party developers can also join the market as the Connector Store encourages the development of reusable IDS Data Apps that can be executed by the IDS Connector, thus, creating prospective incentive models.

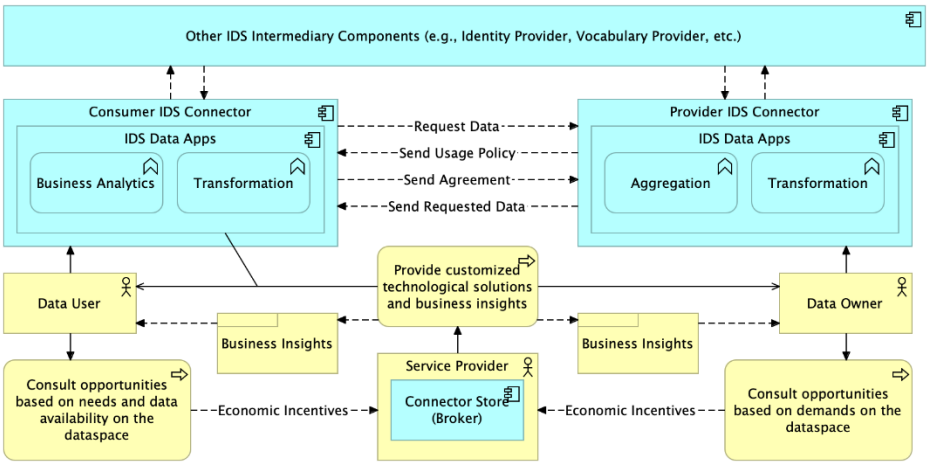


Figure 71 Connector Store – Prospective Business Model Based on Expert Opinion

Lastly, based on his recent study, E5 made a comment that data assets or other items offered by such a data broker can be difficult to price and trade in real-world market structures as business entities may not capture the value proposition of the data assets if offered as it is (Bergman et al., 2022). A recommended business model that can enhance the value of such a data marketplace is then to enrich the ecosystem with complementary customized brokering and consulting services (e.g., data aggregation, data quality assurance, personal consultation about supply and demand matching based on specific needs and data availability, etc.). An illustration of the realizable business model based on the discussion with E5 is captured and depicted in **Figure 71**. Therefore, core participants in the data space pay not only for the data they exchange but also for the additional services that the data marketplace operator (i.e., the service provider company that will operate and maintain the Connector Store) provides. This result serves as a recommendation for future development in case such a logistics data space supported by a Connector Store is to be scaled up and deployed to a real-world setting. Correspondingly, this marks the end of the **Treatment Validation** phase of the design science methodology followed by this EngD thesis.

8.5. Summary and Conclusion

In this chapter, we have demonstrated how we investigated the contribution of the proposed architecture through its logistics data space demonstrator to manage data interoperability, data sovereignty, and resource discovery. The validation work adopted the protocol prescribed by Wieringa (2014), which derives a validation model from the target environment. The validation model is comprised of the application prototypes as the artifact model of the architectural design of a data-sharing ecosystem brokered by a Connector Store. These are applied to the transport trip reporting the case as the model of the Dutch Logistics Sector context.

Five experts with relevant backgrounds participated in this validation effort. Based on the demonstration presented to them, one of the experts highlighted that such an architecture is a good fit for participants in a real-world scenario to handle data interoperability issues, who, in practice, might have partnered with different IT solution providers already. The adoption of the concept of IDS Data Apps, cataloged by a Connector Store and orchestrated by IDS Connectors, allows participants to select and assemble their own message transformation and routing solution as well as to avoid vendor lock-in. Two experts shared the vision and importance of the data usage policy enforcement by IDS Connectors to improve participants' data sovereignty. This is in tune with the need for a technically enforced trust for the ad-hoc data exchange that is predicted to be a trend of data sharing in the future. Despite that, another expert noted that the success of the logistics data space demonstrator brokered by a Connector Store to facilitate participants' quick onboarding will depend on the balance between the availability of offered data in the data space as well as the prospective business cases. In this regard, the discussion with the panel of experts resulted in a proposition to enrich the ecosystem with complementary customized brokering and consulting services to further promote the value for data space participants and service providers participating in the ecosystem. With these results in mind, in the following chapter, the overall conclusion, especially how the obtained results so far contribute to addressing the design questions posed by this thesis, will be discussed.

9 Final Remarks

This chapter concludes this EngD thesis by first discussing, in **Section 9.1**, how this project has addressed the design questions with the architecture of a Connector Store for a Logistics Data Space. Next, the shortcomings and other forms of barriers that were limiting the overall output of this design project will be discussed in **Section 9.2**. Finally, **Section 9.3** closes this thesis by giving pointers to future work following from this work, lessons learned, and limitations of this project.

9.1. Conclusion

In **Chapter 1**, we established that this EngD thesis is focusing on investigating a suitable design of a Connector Store to improve digital data-sharing between logistics companies participating in a Dutch Logistics Data Space. Accordingly, we refined this main question into 5 relevant sub-questions to help address it systematically. Therefore, based on the obtained results so far, we conclude this thesis by re-visiting and answering these sub-questions in the remainder of this section.

What is state-of-the-art data-sharing in the logistics industry?

In **Chapter 3**, we conducted an SLR to gain an understanding of the current pain points and developments of data-sharing in the context of the logistics sector. Through this study, we confirmed that the two most prominent barriers for companies to share data are data interoperability (i.e., syntactic interoperability, semantic interoperability, standards adoption, interoperability between standards, etc.) and data sovereignty (i.e., data confidentiality, control over the usage of the shared data, etc.). With regard to data sovereignty, there is also a question regarding the architecture pattern to be adopted: a centralized or a decentralized approach.

From the literature, we learned that there is a call for the development of an inclusive and sovereign data-sharing ecosystem that prioritizes the use of existing technologies and standards to foster adoptions. To solve data interoperability problems, such an ecosystem should incorporate schema mappings and data transformation functionalities, with several studies suggesting the adoption of the ESB that can be provided by an ISP. Despite the simplicity a centralized paradigm can offer to the ecosystem, the trends from recent studies tend to promote a decentralized approach considering the demand for data sovereignty and a level playing field for every stakeholder in the network. Several initiatives have been proposed that satisfy these prerequisites. Among them, the IDS is the most discussed standard in recent literature that claims the means to solve data interoperability and sovereignty issues. Blockchain technology came as an alternative, however, its cost and benefit are still hard to justify by the industry players, especially for SMEs. Therefore, this thesis adopted the IDS as the design

principle for the data-sharing ecosystem brokered by the Connector Store to solve data interoperability and data sovereignty issues.

What is a suitable architecture to establish an IDS-compliant data-sharing ecosystem?

Representatives from SUTC and eMagiz have reported that implementing a complete IDS ecosystem adhering to the IDS RAM model, with its organizational roles and technical mechanisms, can be considerably challenging for SMEs both from an organizational and technical perspective (Firdausy et al., 2022c). Therefore, this thesis tries to demonstrate the IDS vision by first identifying the core business roles and software components of an IDS-based ecosystem that are essential for instantiating a secure and interoperable "logistics data space". While adhering to TOGAF ADM, we captured the alignment of the stakeholders' goals and requirements with the underlying software applications and communication infrastructure into several architectural viewpoints as discussed in **Chapter 4**.

Starting with the motivation viewpoint, as shown in **Figure 13**, we identified five main requirements (that are derived from the goals, assessments, outcomes, and drivers of the corresponding stakeholders) and associated them with the 5 services to be realized in the business layer. These services are comprised of (1) the evaluation and (2) certification of the candidate participants, (3) the provisioning of IDS Connectors and (4) IDS Data Apps, as well as (5) the metadata publication to facilitate resources and services discovery for data space's participants. We then further detailed the realization of the IDS Evaluation and Certification Services in the next viewpoint shown in **Figure 14**, where we illustrate the onboarding process for interested companies to participate in a data space. In this viewpoint, we also show how the IDS Connectors developed and provided by Software Providers are evaluated and certified first before being offered to candidate participants through the Connector Store. Subsequent to being certified, the new participant can now be actively engaged with other data space participants using their IDS Connectors. The overall landscape of the interactions between participants as data users and owners in the space via their IDS Connectors brokered by a Connector Store is then captured in the Data and Metadata Exchange viewpoint shown in **Figure 15**. From this viewpoint, one can also observe how (1) the sequence of data exchange between participants as specified by the IDS, (2) data users can discover data owners and their data offerings via the Metadata Publication Service realized by the Connector Store, and (3) data users can use IDS Data Apps obtained from a Data App Store to execute data processing capabilities prior to consuming the data from data owners. Complementing these details, the last viewpoint depicted in **Figure 16** is presented to illustrate the functionalities the IDS application components highlighted in the previous viewpoint should support as well as the deployment environment that is required for them. These 4 high-level enterprise architecture viewpoints then guide the instantiation of the logistics data space demonstrator to be focused on designing and developing the (1) IDS Connector and IDS Data Apps to demonstrate secure and interoperable data exchanges prescribed by the IDS, and (2) Connector Store to

facilitate the discovery and selection of IDS Connectors, IDS Data Apps, and data resources for the participants in a data space.

How can the underlying application components of an IDS-compliant data-sharing ecosystem be designed to manage data interoperability and data sovereignty at runtime?

Through their reference architecture model and technical specification reports, the IDSA described that data interoperability and data sovereignty in an IDS environment are mainly managed by IDS Connectors and IDS Data Apps. To further investigate how they can be designed to manage these issues, **Chapter 5** reports the software requirements specification of the mentioned components based on the technical reports published by the IDSA. Next to that, software architectural viewpoints were also produced based on the listed requirements as well as the reference implementations by TNO and Fraunhofer-Sovity.

As discussed by requirements NIR_6, IR_1.6, OS_1, and APS_5 an IDS Connector attempts to solve data interoperability issues in the data exchange between data space partners by supporting the orchestration and routing of IDS Data Apps. The architectural viewpoint in **Figure 20** was presented to illustrate how such routing, which is orchestrated by the IDS Connectors when receiving or offering data, can look like. Interoperability issues such as syntactic mismatches or conflicts between standards can be treated by orchestrating different combinations of IDS Data Apps offering data processing capabilities (e.g., data transformation, data aggregation, schema matching, etc.). Following this, **Sub-Section 5.2.1** further described how these two components look internally and how can they be assembled. The internal system architecture of an IDS Connector shown in **Figure 23** was presented to provide clarity for these questions. This system architecture is generalized from the reference implementation of Dataspace Connector and TNO Security Gateway shown in **Figure 21** and **Figure 22**. Similarly, the internal system architecture of an IDS Data App was also presented in **Figure 24** to illustrate how this app can receive data or routing instructions from the IDS Connector by means of REST or SOAP API communication. While ensuring interoperability between connectors, the IDSA has defined a standardized data model shown in **Figure 25** to be used by IDS Connectors to describe, offer, and receive data resources along with their usage rules.

To manage data sovereignty, requirements IR_1.5, IR_1.7, APS_3, USC_1, USC_2, USC_3, and other related requirements were listed specifying that IDS Connectors should be designed to facilitate data owners in defining and enforcing data usage policies. **Table 17** in **Chapter 5** was presented to list the data usage policy patterns prescribed by the IDSA that should be supported by an IDS Connector by the time this thesis is written. Based on these policy patterns, **Sub-Section 5.2.3** further elaborated on the process required to (1) define the data usage policy when data owners create an offering of their data to the data space as well as (2) enforce it throughout the lifecycle of the data usage contract agreed by the data users. UML Sequence Diagrams in **Figure 26** and **Figure 27** visualized this process and indicated

who initiates the data exchange transaction. Following these diagrams, the technical processes (by means of JSON Snippet) to create offered data, discover data owner's metadata from Metadata Broker, initiate contract negotiation, and execute artifact consumption based on the reference implementation of Dataspace Connector were also discussed. These specifications, architectural designs, as well as technical instructions for key processes to enact data exchange transactions and enforce the defined data usage policy were then used to guide the development of the prototypes comprising the logistics data space demonstrator for the validation purpose.

How can a Connector Store be designed to support companies to discover and select the underlying application components?

A keynote presentation at the Data Spaces Symposium 2023 in The Hague confirmed that participants need to install an IDS Connector prior to participating in a data space, for which, no immediate demand might be present on the get-go. This matter calls for an instrument that is capable of alleviating such an onboarding issue as well as facilitating discovery to connect supply and demand. Thus, a Connector Store is proposed. In this thesis, we concluded that the business role and application component from the IDS specification that is in line with the goal and requirements the Connector Store must achieve and satisfy (refer to the Motivation Viewpoint discussed in **Figure 13**) is the Broker Service Provider role with its IDS Metadata Broker component. In **Section 6.1**, according to the IDS specification, we investigated (1) how a Metadata Broker can support participant and service discoveries in a data space through its metadata publication service and then (2) we captured its system architecture and interaction with its environment in an enterprise architecture viewpoint presented in **Figure 28**. Based on this investigation, this thesis proposed the Connector Store to act as an extension of the IDS Metadata Broker by providing additional functionality to support the semantic discovery and selection of IDS Connectors.

The IDS specification suggests that such an implementation of a broker should be combined with the adoption of the Semantic Web and Linked Data technology. Therefore, in **Section 6.2**, we discussed how the design of a Connector Store was initiated with the development of an ontology that describes IDS Connectors for potential participants of an IDS ecosystem. The goal for this is to annotate the IDS Connectors, along with other interrelated properties (e.g., its offered data resources, service providers, users, etc. with their contextual information), and explain them in the form of subject-predicate-object triples according to the RDF format. This resulted in an ontology requirement specification document and preliminary Connector Store Ontology conceptual model presented in **Table 35** and **Figure 29** respectively. From there on, **Figure 30** illustrated how the developed preliminary ontology can be operationalized as an operational ontology and deployed as a working application that can interact with other participants in a data space. **Figure 31** was also presented as an alternative viewpoint to **Figure 28** to show how such a Connector Store extends the functionality of a base IDS Metadata Broker with the IDS Connector Provisioning Discovery Service to help data owners and users find

and acquire the best IDS Connectors fitting to their requirements based on the connectors' contextual information. Additionally, this viewpoint also shows how software and service providers can submit the metadata describing the IDS Connectors that they develop and offer in the ecosystem. Having these two viewpoints guided the development of the Connector Store prototype as a validation instrument of the proposed architecture with a panel of experts.

To what extent the IDS-compliant data-sharing ecosystem architecture can support logistics companies to manage data interoperability and data sovereignty at runtime?

To investigate whether the proposed architecture of the logistics data space can support logistics companies to manage data interoperability and sovereignty at runtime, the architecture has to be operationalized into a demonstrator and then applied to a model case of the intended context. For this, first, in **Chapter 7**, we described how we instantiated the demonstrator of the data space by developing the comprising application prototypes in accordance with the design and specifications elaborated in the preceding chapters. The development started with the prototype of an IDS Connector. We discussed the implementation of (1) containerized deployment environment, (2) IDS Connector Data Model, (3) resource offering and metadata publication, (4) resource request and contract negotiation, (5) artifact consumption and route execution, and (7) IDS Data Apps management. Next to this, we also elaborated on the development of IDS Data Apps prototypes. To provide implementation alternatives, in this thesis, we demonstrated the development of data transformation and message routing apps using both conventional raw-code approach (i.e., Java Spring Boot) as well as low-code (i.e., eMagiz iPaaS platform). The instantiation of the demonstrator was then followed by the development of the Connector Store to support participants' onboarding and service and resource discovery. As discussed in **Sub-Section 7.3.1**, the development of this component started with the instantiation of the Connector Store operational ontology that is represented in OWL-Turtle format, then deployed to a Fuseki Server instance running on a VPS. Next, we developed an application running on top of this Fuseki Server to provide (1) visual interfaces for data space participants to support their onboarding and discovery process and (2) REST API endpoints to facilitate metadata publication services for the participating IDS Connectors.

Having these application components that comprise the data space developed and deployed, next we examined, by means of expert opinions, their contribution to managing data interoperability and data sovereignty problems in a use case from the Dutch Logistics Sector. We demonstrated to a panel of experts from eMagiz Services B.V, a representative from SUTC-TLN, as well as a Ph.D. researcher from TU Delft, how the logistics data space demonstrator works in realizing an interoperable and sovereign data exchanges for a transport trip reporting case called the VESDI project. From the demonstration with the panel, we obtained quantitative assessments of the proposed architecture of the logistics data space and its contributions to the management of data interoperability and data sovereignty

issues. Measured with a Likert Scale ranging from 1 to 5 and then analyzed based on their average and standard deviation, we can see that that the five respondents fairly agree that the architecture of the logistics data space, through its demonstrator, can (1) positively contribute to facilitating data transformation between standards and data formats, (2) positively stimulate FAIRness of data, (3) enable data owners to enforce data usage control, and (4) support data owners to maintain confidentiality aspect of their data assets. Meanwhile, there is still a significant uncertainty among respondents that the architecture of the logistics data space, through its demonstrator, is able to (1) motivate data space participants to devise new business models and (2) enable participants' quick onboarding to the ecosystem.

Additionally, through the discussion following the demonstration, we reached a preliminary consensus that the proposed architecture of the logistics data space (i.e., as shown in **Figure 69** when applied to the VESDI Project that is based on **Figure 15**) is suitable for participants in a real-world scenario to handle data interoperability and data sovereignty issues. Using the IDS Data Apps, requested from a Connector Store and orchestrated by IDS Connectors, participants are able to solve proprietary and conflicting data formats by assembling their own message transformation and routing solution while minimizing vendor lock-ins. The implementation of the data usage policy enforcement conforming to the IDS principles as accommodated by the proposed logistics data space architecture is also in tune with the need for a technically enforced trust for the ad-hoc data exchange that is predicted to be a trend of data sharing in the future. Despite the lack of alternative use cases to demonstrate the generalizability of the proposed logistics data space architecture, the panel discussion resulted in a proposition to inject the ecosystem with complementary customized brokering and consulting services, as illustrated in **Figure 71**, to further enhance the value for data space participants and service providers participating in the ecosystem.

9.2. Limitations

Despite the successful delivery of the architecture and demonstrator of the logistics data space brokered by a Connector Store, several shortcomings limiting the overall output of this design project were also identified. The first limitation revolves around the scoping of this thesis as well as the positioning of the Connector Store within a data-sharing ecosystem. This thesis aims to investigate a suitable design of a Connector Store to manage data interoperability and data sovereignty. However, the IDS specification prescribes that such issues in the data space are handled by the usage of IDS Data Apps that are orchestrated by the IDS Connector. This principle then calls for the Connector Store to facilitate the discovery and selection of suitable IDS Connectors and IDS Data Apps for the data space's participants. This would imply that this thesis also has to investigate the design of these three application components for validation purposes. This thesis could have used the reference implementations of the IDS Connectors and IDS Data Apps by Fraunhofer-Sovity and TNO and focused on the design and development of the Connector Store. However, as we tested it during the period of this project, the Dataspace Connector

published by Fraunhofer was not able to fully execute two key functionalities as promoted, namely (1) IDS Data Apps orchestration and routing, and (2) data usage policy enforcement, especially the "Usage Until Deletion" data usage rule. Relying on this reference implementation would lead to the risk of hindering the development of the Connector Store as well as the evaluation of the IDS principles in managing data interoperability and data sovereignty. Therefore, to maximize the flexibility of customizing the functionalities as well as investigating their internal mechanism, we decided that we would develop replicas of these two application components from scratch. This development cost us a big portion of time in the end and reduces the time available for the validation effort.

This decision then led to the second limitation of the implemented functionalities of the developed prototypes. Due to the limited duration of the project, we could only implement key functionalities prescribed by the IDS that are directly contributing to solving data interoperability and sovereignty issues within a data transaction. The developed IDS Connector prototype could have supported more data usage policies such as "Usage Logging" to a Clearing House. Enabling this could have addressed one more requirement regarding event logging that was mentioned in the original proposal of this design project. Therefore, we delegated the investigation and development of such a Clearing House to a master thesis. The data usage patterns supported by the IDS Connector prototype were also implemented with simple custom algorithms as opposed to the specification prescribed by the IDS that uses standardized technologies, such as the MYDATA Control Technologies[58]. IDS Data App prototypes elaborated in this thesis were also developed pragmatically and not fully compliant yet with the technical specifications and protocols advocated by the IDSA. In addition, the Connector Store itself could have been enriched with more advanced functionalities, such as a search function to find suitable resources based on their contextual information. Furthermore, the Connector Store could also have implemented a recommendation feature based on the already available metadata to further enhance data space participants' discoverability and onboarding process. However, there was limited time available to formulate the most optimized SPARQL queries for this purpose. Because of this, the development of these functionalities was suspended and the development effort was redirected to the development of the IDS Connector and IDS Data App prototypes.

Lastly, there is also a limitation on the validation effort conducted in this thesis. The validation work discussed in **Chapter 8** was focused on collecting the experts' opinions on the logistics data space's contribution to solving data interoperability and sovereignty issues. However, the validation could still be enriched with the measurements of the performance of the developed prototypes. An example of this is to measure the time, complexity, effort, and cost needed to perform a data exchange in the baseline architecture against the same measures in the target architecture. Next to this, there is still a question of how to systematically measure

[58] https://internationaldataspaces.org/wp-content/uploads/dlm_uploads/IDSA-Position-Paper-Usage-Control-in-the-IDS-V3..pdf

the data sovereignty in the proposed target architecture that is perceivable by data space participants. Measuring the improvements in the discoveries and onboarding process as compared to the existing situation will also be beneficial. These shortcomings, which mainly occurred due to the limited time available for this project, if addressed attentively can provide more insights for data space developers and providers on the promised benefits and improvements over the existing data-sharing solutions.

9.3. Future Work

Based on the presented architecture as well as the demonstrator of the logistics data space brokered by the Connector Store, we could observe that there are several trajectories for improvements. First of all, due to the intrinsic value and sensitivity of the exchanged data in the data space, there might be scenarios where data owners call for a mechanism that facilitates logging and reporting with respect to the usage of the data by the data users. In addition, there might also be scenarios where inconsistencies or other sorts of failures in data transactions occur at some point in time within an operational data space. The IDSA introduced a business role as well as an application component for this purpose known as the Clearing House which is responsible for logging and clearing functions within a particular data space. As mentioned earlier, due to the limitations in the development of the prototype, the proposition of the demonstrated logistics data space can be complemented with the development of the Clearing House component. Establishing the connection between IDS Connectors and a Clearing House could further enhance data sovereignty and trust aspects in the data-sharing ecosystem through its event logging, clearing, and conflict resolution propositions. Though the IDSA has provided some guidelines regarding its functions and responsibilities, by the time this thesis is written, little to no research works have been found that describe and investigate the design of the Clearing House, especially to be implemented in the logistics sector. Moreover, the impact (e.g., cost and benefit analysis) of implementing such a solution for this industry sector is also still an unchartered research area. Therefore, an investigation into the design and implementation of this component is needed.

Secondly, to treat data interoperability and sovereignty issues, earlier work proposed a high-level architecture of a data-sharing ecosystem that comprises an Interoperability Simulator to complement the Connector Store investigated in this thesis. This simulator is aimed to simulate collaboration opportunities between participants prior to implementation and might combine the Digital Twin and IDS principles for simulating interoperability within a network. In connection with the artifacts developed in this thesis, users of the IDS Connector and Connector Store could benefit from the value proposition offered by such a simulator. By taking advantage of the metadata stored and managed by the Connector Store, data space participants could use the Interoperability Simulator to assemble, configure, and simulate data exchange scenarios. Particularly, to investigate (1) which data space participants should be interacting in such a scenario, (2) which application components (i.e., IDS Connectors, IDS Data Apps, and participants' enterprise

systems) should be involved to realize such a scenario, (3) which protocols and data formats are used for such a scenario, (4) what will be the anticipated costs and benefits perceived by involved parties in realizing such a scenario, etc. Therefore, a future trajectory following this thesis is to come up with a suitable design of such a simulator and investigate a concrete (preferably, enterprise) architecture that illustrates the interplay between the Interoperability Scenario and the Connector Store in supporting data space participants to configure and assess their data exchange scenarios prior to implementation and deployment.

Lastly, based on the results we obtained from the validation efforts, we learned that the success of the logistics data space demonstrator, brokered by a Connector Store to facilitate participants' quick onboarding, connects well with the balance of the offered data' supply and demand in the data space as well as the prospective business cases. In this regard, the discussion with the panel of experts resulted in a proposition to enrich the ecosystem with complementary customized brokering and consulting services to further promote the value for data space participants and service providers participating in the ecosystem. Such a recommendation then calls for a further study that investigates the proposition from several perspectives. To start, we argue that there should be an investigation into different data should be offered in the data space to serve or support what kind of business models, including how these data should be maintained and managed throughout their usage lifecycle. Having these business models set in place, the next step is to investigate how customized brokering and consulting services should be managed by service providers. Consequently, one should also investigate (1) the capabilities that service providers must have for this purpose, as well as (2) the benefits or revenue models that involved parties can perceive from these services. Thus, we would like to close this discussion with a conclusion that this design project has (1) demonstrated the technical feasibility of developing the proposed IDS-compliant logistics data space brokered by a Connector Store and (2) provided the lesson learned from the development and validation that serve as a basis for the future research endeavor.

References

Abebe, E., Behl, D., Govindarajan, C., Hu, Y., Karunamoorthy, D., Novotny, P., . . . Vecchiola, C. (2019). *Enabling Enterprise Blockchain Interoperability with Trusted Data Transfer (industry track)*.

Abecker, A., Brauer, T., Magoutas, B., Mentzas, G., Papageorgiou, N., & Quenzer, M. (2014). *Standards and semantics to support interoperable software solutions in the water distribution chain*.

Abu-elezz, I., Hassan, A., Nazeemudeen, A., Househ, M., & Abd-alrazaq, A. (2020). The benefits and threats of blockchain technology in healthcare: A scoping review. *International Journal of Medical Informatics, 142*. doi:10.1016/j.ijmedinf.2020.104246

Andreeva, E., Poletaeva, T., Abdulrab, H., & Babkin, E. (2016). *A semantic solution for seamless data exchange in supply networks*.

Bader, S., Bruckner, F., Böge, G., Kubitza, D. O., Langkau, J., Murthy, D., & Nagel, R. (2020). *Specification: IDS Meta Data Broker*. Retrieved from

Bader, S., Pullmann, J., Mader, C., Tramp, S., Quix, C., Müller, A. W., . . . Lipp, J. (2020). *The International Data Spaces Information Model–An Ontology for Sovereign Exchange of Digital Content.* Paper presented at the International Semantic Web Conference.

Banek, M., Juric, D., Pintar, D., Skocir, Z., Vranic, M., & Vrdoljak, B. (2008, 10-12 Sept. 2008). *E-business infrastructure for supporting the integration of tourist services.* Paper presented at the 2008 50th International Symposium ELMAR.

Bastiaansen, H. J. M., Kollenstart, M., Dalmolen, S., & van Engers, T. M. (2020). *User-centric network-model for data control with interoperable legal data sharing artefacts: Improved data sovereignty, trust and security for enhanced adoption in interorganizational and supply chain is applications*.

Bergman, R., Abbas, A. E., Jung, S., Werker, C., & de Reuver, M. (2022). Business model archetypes for data marketplaces in the automotive industry. *Electronic Markets, 32*(2), 747-765. doi:10.1007/s12525-022-00547-x

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific american, 284*(5), 34-43.

Bhatt, T., & Zhang, J. R. (2013). Food Product Tracing Technology Capabilities and Interoperability. *Journal of Food Science, 78*, B28-B33. doi:10.1111/1750-3841.12299

Bicocchi, N., Cabri, G., Mandreoli, F., & Mecella, M. (2019). Dynamic digital factories for agile supply chains: An architectural approach. *Journal of Industrial Information Integration, 15*, 111-121. doi:10.1016/j.jii.2019.02.001

Bol Raap, W., Iacob, M.-E., Sinderen, M. v., & Piest, S. (2016). *An architecture and common data model for open data-based cargo-tracking in synchromodal logistics.* Paper presented at the OTM Confederated International Conferences" On the Move to Meaningful Internet Systems".

Bonnema, G. M., Veenvliet, K. T., & Broenink, J. F. (2016). *Systems design and engineering: facilitating multidisciplinary development projects*: CRC press.

Bouter, C., Biagioni, G., van Gessel, T., Korteling, W., de Graaf, E., & Hofman, W. (2022). *Towards a Modular Ontology for Event-Based Data Sharing in the Logistics Domain*.

Campos, J. G., Martin, R. M., Lopez, J. S., & Quiroga, J. I. A. (2016). *E-traceability for ISO STEP CAD/CAM/CNC supply chains*.

Carvalho, A., Melo, P., Oliveira, M. A., & Barros, R. (2020). *The 4-corner model as a synchromodal and digital twin enabler in the transportation sector*.

Cicchetti, A., Ciccozzi, F., & Pierantonio, A. (2019). Multi-view approaches for software and system modelling: a systematic literature review. *Software and Systems Modeling, 18*(6), 3207-3233.

Cirullies, J., & Schwede, C. (2021). *On-demand shared digital twins - An information architectural model to create transparency in collaborative supply networks*.

Dalmolen, S., Bastiaansen, H. J. M., Kollenstart, M., & Punter, M. (2019). *Infrastructural sovereignty over agreement and transaction data ('metadata') in an open network-model for multilateral sharing of sensitive data*.

Dalmolen, S., Bastiaansen, H. J. M., Somers, E. J. J., Djafari, S., Kollenstart, M., & Punter, M. (2019). Maintaining control over sensitive data in the Physical Internet: Towards an open, service oriented, network-model for infrastructural data sovereignty.

Das, M., Cheng, J. C. P., & Law, K. H. (2015). An ontology-based web service framework for construction supply chain collaboration and management. *Engineering, Construction and Architectural Management, 22*(5), 551-572. doi:10.1108/ECAM-07-2014-0089

Debicki, T., & Kolinski, A. (2019, Oct 10-11). *Integration Platform In Global Supply Chains - Who Is Beneficiary And Who Is Not.* Paper

presented at the 19th International Scientific Conference on Business Logistics in Modern Management, Osijek, CROATIA.
Dinalog, T. (2020). The Logistics Data Sharing Infrastructure (2020). In.
Ferreira, J., Agostinho, C., Ilie-Zudor, E., Jardim-Goncalves, R., & Asme. (2012, Nov 09-15). *Monitor For Information Alignment And Sustainability In Logistics Networks.* Paper presented at the ASME International Mechanical Engineering Congress and Exposition, Houston, TX.
Firdausy, D. R., de Alencar Silva, P., van Sinderen, M., & Iacob, M.-E. (2022a). *A Data Connector Store for International Data Spaces.* Paper presented at the International Conference on Cooperative Information Systems.
Firdausy, D. R., de Alencar Silva, P., van Sinderen, M., & Iacob, M. E. (2022b). Semantic discovery and selection of data connectors in international data spaces. *Proceedings http://ceur-ws. org ISSN, 1613*, 0073.
Firdausy, D. R., de Alencar Silva, P., van Sinderen, M., & Iacob, M. E. (2022c, 15-17 June 2022). *Towards a Reference Enterprise Architecture to enforce Digital Sovereignty in International Data Spaces.* Paper presented at the 2022 IEEE 24th Conference on Business Informatics (CBI).
Främling, K., Parmar, S., Hinkka, V., Tätilä, J., & Rodgers, D. (2013) Assessment of EPCIS standard for interoperable tracking in the supply chain. In*: Vol. 472. Studies in Computational Intelligence* (pp. 119-134).
Frey, C., Hertweck, P., Richter, L., & Warweg, O. (2022). Bauhaus.MobilityLab: A Living Lab for the Development and Evaluation of AI-Assisted Services. *Smart Cities, 5*(1), 133-145. doi:10.3390/smartcities5010009
Gnimpieba Zanfack, D. R., Nait-Sidi-Moh, A., Durand, D., & Fortin, J. (2015) Publish and subscribe pattern for designing demand driven supply networks. In*: Vol. 467* (pp. 45-55).
Gómez-Pérez, A., & Suárez-Figueroa, M. C. (2009). NeOn methodology for building ontology networks: a scenario-based methodology.
GS1. (2021). GS1 Transport & Logistics. Retrieved from https://www.gs1.org/industries/transport-and-logistics
Guizzardi, G. (2005). Ontological foundations for structural conceptual models.
Heinbach, C., Meier, P., & Thomas, O. (2022). Designing a shared freight service intelligence platform for transport stakeholders using mobile

telematics. *Information Systems and e-Business Management*. doi:10.1007/s10257-022-00572-5

Hofman, W. (2016). *Data sharing requirements of supply - And logistics innovations - Towards a maturity model*.

Hofman, W. (2019). Toward large-scale logistics interoperability based on an analysis of available open standards. *Proceedings of the I-ESA Conferences, 9*, 249-261. doi:10.1007/978-3-030-13693-2_21

Hofman, W. J. (2019). *A Methodological Approach for Development and Deployment of Data Sharing in Complex Organizational Supply and Logistics Networks with Blockchain Technology*.

Iacob, M.-E., Charismadiptya, G., van Sinderen, M., & Piest, J. P. S. (2019). *An architecture for situation-aware smart logistics.* Paper presented at the 2019 IEEE 23rd International Enterprise Distributed Object Computing Workshop (EDOCW).

IDSA. (2019). *IDS Reference Architecture Model Version 3.0*. Retrieved from Berlin:

IDSA. (2020a). IDS is Officially a Standard: DIN SPEC 27070 is Published. Retrieved from https://internationaldataspaces.org/ids-is-officially-a-standard-din-spec-27070-is-published/

IDSA. (2020b, 06-04-2020). IDS Metadata Broker API Reference Implementation. Retrieved from https://app.swaggerhub.com/apis/idsa/IDS-Broker/1.3.1

IDSA. (2020c). *IDSA Rule Book Whitepaper* (S. Steinbuss Ed.). Berlin: International Data Spaces Association,.

IDSA. (2021a). IDS Connector Architecture. Retrieved from https://international-data-spaces-association.github.io/DataspaceConnector/Documentation/v6/Architecture

IDSA. (2021b). IDS Connector Data Model. Retrieved from https://international-data-spaces-association.github.io/DataspaceConnector/Documentation/v6/DataModel

IDSA. (2021c). *IDS Data Apps App Types and Profiles*. Retrieved from https://industrialdataspace.jiveon.com/docs/DOC-3882

IDSA. (2021d). IDS Deployment Examples: Provider-Consumer Example. Retrieved from https://github.com/International-Data-Spaces-Association/IDS-Deployment-Examples/tree/main/dataspace-connector/provider-consumer

IDSA. (2021e). IDS Usage Control Policies. Retrieved from https://international-data-spaces-

association.github.io/DataspaceConnector/Documentation/v6/UsageControl
IDSA. (2021f). The International Data Spaces (IDS) Information Model. Retrieved from https://github.com/International-Data-Spaces-Association/InformationModel
IDSA. (2021g). Using Apache Camel. Retrieved from https://international-data-spaces-association.github.io/DataspaceConnector/CommunicationGuide/v6/Camel
IDSA. (2022). Dataspace Connector. Retrieved from https://github.com/International-Data-Spaces-Association/DataspaceConnector
IDSA Certification Working Group. (2021). Component Certification Criteria Catalog V2.1.2. In: International Data Spaces Association.
INAD Industrie Software B.V. (2022). EDI4STEEL. Retrieved from https://www.edi4steel.eu/about/
Janowicz, K., Van Harmelen, F., Hendler, J. A., & Hitzler, P. (2015). Why the data train needs semantic rails. *AI Magazine, 36*(1), 5-14.
Karatas, C., & Gultekin, M. (2021). *EDI Based Secure Desing Pattern for Logistic and Supply Chain*.
Kazantsev, N., Pishchulov, G., Mehandjiev, N., Sampaio, P., & Zolkiewski, J. (2022). Investigating barriers to demand-driven SME collaboration in low-volume high-variability manufacturing. *Supply Chain Management, 27*(2), 265-282. doi:10.1108/SCM-10-2021-0486
Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering.
Lankhorst, M. (2009). *Enterprise architecture at work* (Vol. 352): Springer.
Lopes-Martínez, I., Paradela-Fournier, L., Rodríguez-Acosta, J., Castillo-Feu, J. L., Gómez-Acosta, M. I., & Cruz-Ruiz, A. (2018). The use of GS1 standards to improve the drugs traceability system in a 3PL Logistic Service Provider. *Dyna, 85*(206), 39-48.
Markus, M. L., & Bui, Q. N. (2012). Going concerns: The governance of interorganizational coordination hubs. *Journal of Management Information Systems*, 163-198.
McGuigan, E., van den Bremen, J., McKillips, B., Roy, P., & Mukherjee, S. (2022). *Interoperability: Value untangled - Accelerating radical growth through interoperability.* Retrieved from https://www.accenture.com/content/dam/accenture/final/capabilities/technology/software-engineering/document/Accenture-Report-ITL-IPS.pdf

OpenTripModel. (2021). What is the Open Trip Model? Retrieved from https://www.opentripmodel.org/page/about

Otto, B. (2019). Interview with Reinhold Achatz on "Data Sovereignty and Data Ecosystems". *Business & Information Systems Engineering, 61*(5), 635-636. doi:10.1007/s12599-019-00609-z

Otto, B., Hompel, M. t., & Wrobel, S. (2019). International data spaces. In *Digital Transformation* (pp. 109-128): Springer.

Otto, B., & Jarke, M. (2019). Designing a multi-sided data platform: findings from the International Data Spaces case. *Electronic Markets, 29*(4), 561-580.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems, 24*(3), 45-77. doi:10.2753/MIS0742-1222240302

Pérez, J., Arenas, M., & Gutierrez, C. (2006). *Semantics and Complexity of SPARQL.* Paper presented at the International semantic web conference.

Pham, H. C., Nguyen, T.-T., McDonald, S., & Tran-Kieu, N. Q. (2019). Information Sharing in Logistics Firms: An Exploratory Study of the Vietnamese Logistics Sector. *The Asian Journal of Shipping and Logistics, 35*(2), 87-95. doi:https://doi.org/10.1016/j.ajsl.2019.06.001

Piest, J. P. S., Iacob, M. E., & van Sinderen, M. (2020). *A federated interoperability approach for data driven logistic support in SMEs.* Paper presented at the 10th International Conference on Interoperability for Enterprise Systems and Applications, I-ESA 2020: Interoperability in the era of artificial intelligence.

Piest, J. P. S., Meertens, L. O., Buis, J., Iacob, M. E., & van Sinderen, M. J. (2020). *Smarter interoperability based on automatic schema matching and intelligence amplification.* Paper presented at the 10th I-ESA SIFAI Workshop. URL http://ceur-ws. org.

Rouhani, B. D., Mahrin, M. N. r., Nikpay, F., Ahmad, R. B., & Nikfard, P. (2015). A systematic literature review on Enterprise Architecture Implementation Methodologies. *information and Software Technology, 62*, 1-20.

Salma, S., Bouneffa, M., & Habiba, C. (2019, 24-25 Oct. 2019). *Ontology and Semantic Web in Logistic Applications: State of the Art.* Paper presented at the 2019 7th Mediterranean Congress of Telecommunications (CMT).

Schöggl, J. P., Fritz, M. M. C., & Baumgartner, R. J. (2016). Toward supply chain-wide sustainability assessment: A conceptual framework and

an aggregation method to assess supply chain performance. *Journal of Cleaner Production, 131*, 822-835. doi:10.1016/j.jclepro.2016.04.035

Scholz, J., De Meyer, A., Marques, A. S., Pinho, T. M., Boaventura-Cunha, J., Van Orshoven, J., . . . Nummila, K. (2018). Digital Technologies for Forest Supply Chain Optimization: Existing Solutions and Future Trends. *Environmental Management, 62*(6), 1108-1133. doi:10.1007/s00267-018-1095-5

Slavova, S. (2021). *Predicting the occupancy rates of truck parking locations: a machine learning approach.* University of Twente,

Sommerville, I. (2015). Software engineering 10th Edition. *ISBN-10, 137035152*, 18.

Soylu, A., Corcho, O., Elvesæter, B., Badenes-Olmedo, C., Martínez, F. Y., Kovacic, M., . . . Simperl, E. (2020). *Enhancing public procurement in the European Union through constructing and exploiting an integrated knowledge graph.* Paper presented at the International Semantic Web Conference.

Stolwijk, C., & Berkers, F. (2020). *Scalability and agility of the Smart Connected Supplier Network approach*. Retrieved from

Suárez-Figueroa, M. C., Gómez-Pérez, A., & Villazón-Terrazas, B. (2009). *How to write and use the ontology requirements specification document.* Paper presented at the OTM Confederated International Conferences" On the Move to Meaningful Internet Systems".

Tan, B. Q., Wang, F., Liu, J., Kang, K., & Costa, F. (2020). A blockchain-based framework for green logistics in supply chains. *Sustainability (Switzerland), 12*(11). doi:10.3390/su12114656

TNO. (2020). Smart-Connected Supplier Network (SCSN) Addressbook. Retrieved from https://broker.ids.smart-connected.nl/#home

Top, J., Janssen, S., Boogaard, H., Knapen, R., & Şimşek-Şenel, G. (2022). Cultivating FAIR principles for agri-food data. *Computers and Electronics in Agriculture, 196*. doi:10.1016/j.compag.2022.106909

Tran, T. T. H., Childerhouse, P., & Deakins, E. (2016). Supply chain information sharing: Challenges and risk mitigation strategies. *Journal of Manufacturing Technology Management, 27*(8), 1102-1126. doi:10.1108/JMTM-03-2016-0033

Veenstra, A. W. (2018). Digitalization: a supply chain revolution? *Journal of the Hitachi Research Insitute, 13*(3), 28-33.

Verhoosel, J., Van Bekkum, M., & Verwaart, T. (2018). Semantic interoperability for data analysis in the food supply chain. *International Journal on Food System Dynamics, 9*(1), 101-111. doi:10.18461/ijfsd.v9i1.917

Voswinckel, T., Hardjosuwito, D., Gehring, T., Siruet, R., & Fuessler, A. (2020) Impact Analysis of Industrial Standards on Blockchains for Food Supply Chains. In*: Vol. 598* (pp. 524-533).

Wang, H., Liu, Z., & Liang, Y. (2019). *Research on the Three-in-One Model of Agricultural Products E-commerce Logistics under the Combination of Resource Saving and Blockchain Technology*.

Wang, X., Zhang, C., Jin, Y., & Zhao, X. (2018, 9-11 May 2018). *CPSP: A Cloud-based Production Service Platform Supporting Co-Manufacturing of Cross-Enterprise.* Paper presented at the 2018 IEEE 22nd International Conference on Computer Supported Cooperative Work in Design ((CSCWD)).

Wang, X. X., Liu, X. Y., & Li, Z. Q. (2018). A social collaborative urban distribution integration platform. *Journal of Interdisciplinary Mathematics, 21*(5), 1109-1113. doi:10.1080/09720502.2018.1493038

Wieringa, R. J. (2014). *Design science methodology for information systems and software engineering*. London: Springer.

Zhao, H. J., & Liang, Y. J. (2015). *Maritime Information Integration Based on RFID Middleware*.

Zrenner, J., Möller, F. O., Jung, C., Eitel, A., & Otto, B. (2019). Usage control architecture options for data sovereignty in business ecosystems. *Journal of Enterprise Information Management*.

Appendices

Appendix A. Related Publications and Technological Artefacts

This appendix contains an overview of published scientific papers and links to the repository for conceptual and technological artefacts related to this EngD thesis.

Scientific papers directly related to this EngD thesis

Firdausy, D.R., de Alencar Silva, P., van Sinderen, M., Iacob, ME. (2022). Towards a Reference Enterprise Architecture to enforce Digital Sovereignty in International Data Spaces. In: 2022 IEEE 24th Conference on Business Informatics (CBI), 15-17 June 2022 2022, vol. 01, pp. 117-125. https://doi.org/10.1109/CBI54897.2022.00020

Firdausy, D.R., de Alencar Silva, P., van Sinderen, M., Iacob, ME. (2022). Semantic discovery and selection of data connectors in international data spaces. Proceedings http://ceur-ws.org ISSN, 1613, 0073.

Firdausy, D.R., de Alencar Silva, P., van Sinderen, M., Iacob, ME. (2022). A Data Connector Store for International Data Spaces. In: Sellami, M., Ceravolo, P., Reijers, H.A., Gaaloul, W., Panetto, H. (eds) Cooperative Information Systems. CoopIS 2022. Lecture Notes in Computer Science, vol 13591. Springer, Cham. https://doi.org/10.1007/978-3-031-17834-4_14

Links for technological artefacts related to this EngD thesis

- CLiCKS Connector Store Mendix Application (version 9.19.0)

  https://clicksconnectorstore-sandbox.mxapps.io/

  https://github.com/danniarreza/CLiCKS-Connector-Store-main

- CLiCKS Connector Store RDF Turtle Dataset

  https://github.com/danniarreza/connectorstoreontology/blob/main/connectorstoreinferredv10.ttl

- CLICKS IDS Connector Mendix Application (version 9.19.0)

  https://clicksidsconnectorv1-sandbox.mxapps.io/

  https://github.com/danniarreza/CLiCKS-IDS-Connector-v1-main

- CLiCKS IDS Connector Mendix Application Docker Image

danniarreza/d.r.repo:clicksconnector

- OTM Data App and Regulatory Body Data App Docker Image

  danniarreza/d.r.repo:otmdataapp

  danniarreza/d.r.repo:regulatorybodydataapp

- eMagiz API Gateway SwaggerUI and Sample Consignment Endpoint (with additional headers)

  https://api-cloud0108.emagizcloud.com/swaggerui/index.html

  https://api-cloud0108.emagizcloud.com/cnsmtfra/04TRN1002

  x-api-key : s6_*@V-k+0W8MC+OUXS7JtRj5@Y+-!~QlQmxI_(q_Euc

  Accept-Encoding : application/json

- Mockup System of Transport Company A Mendix Application

  https://atransportco-sandbox.mxapps.io/

- Mockup System of Transport Company B Mendix Application

  https://btransportcompany-sandbox.mxapps.io/

- Mockup System of Regulatory Body Mendix Application

  https://regulatorycomplianceproject-sandbox.mxapps.io/

Appendix B. SLR Extracted Data

No.	References	Research Goal	Challenges	Treatment
P1	(Ferreira et al., 2012)	Suggest aligning the different enterprises operating within the network with their different systems and information structures, using morphisms to relate and develop transformations between their information models and a common data exchange standard.	Seamless network interoperability is hard to maintain due to different requirements, policies information systems, and data formats (unique IDs, data model). Confidentiality issues surface in a supply-chain network as participating members may be each other's competitors.	Implementation of standards (to integrate Quotation, Order Confirmation, Delivery Note, etc.) Introduced model morphism (data mapping and transformation mechanism) approach between IS in the network. Proposed an intermediary hub and mapping knowledge base for each involved IS.
P2	(Bhatt & Zhang, 2013)	Determine the interoperability of product tracing technology systems: Assess interoperability when full supply-chain data are provided (in pieces) Assess the ability of the network of software systems to include/exclude potential sources or recipients of contaminated product	Concerns about the data confidentiality, protection of formulation information, and potential loss of competitive advantage create corporate resistance to more actively engaging in whole-chain traceability, some small companies in the food industry still resort to paper-based systems for track & trace in their logistics. A centralized database system approach is argued to have	Call for minimum data sets/common data model/standards needed to enable connectivity. Call for a design for an interoperable framework to accommodate a wide variety of platforms, technologies, and business practices and to be inclusive for anyone to follow. Call for an approach to protect data and maintain privacy while sharing data with partners and regulatory agencies.

			serious performance issues for interoperating complex systems and processing large data sets.	
P3	(Främling et al., 2013)	Analyze why GS1's EPCIS has not been universally adopted as a global data-exchange standard for track and trace by analyzing three application areas despite its promised benefits.	Proprietary tracking or serial numbers are the most used identifiers, despite their unsuitable nature for inter-organizational data exchange. The adoption of standards is hindered by (1) the concern over securities and privacies on data exchange, and (2) the presence of the existing/own ISs in the IT landscape EPCIS can be perceived to be too complex or costly to implement compared to the benefits.	Introduce GS1's EPCIS as standards for RFID-enabled shipment and product individuals' identification based on a common XML schema. Call for a software integrator company that can provide a ready "EPCIS product", or mapping towards EPCIS standards.
P4	(Abecker et al., 2014)	Introduces the WatERP project to achieve a higher degree of interoperability by developing a flexible and extensible communication architecture for different kinds of tools in use, based on SOA complemented with Multi-Agent System for service identification and orchestration.	Comprehensive and real-time data and software interoperability throughout the whole water-supply chain is not given enough attention in practice.	Introduced an architecture of a data-sharing platform that adopts SOA for inter-systems communication paradigm and MAS for services discovery and orchestration. Highlighted a central, OWL-based data and knowledge base stored in a triple store that aims to increase interoperability and unlock the opportunity towards the Linked Open Data paradigm.

P5	(Das et al., 2015)	Present a framework for integrating the construction supply chain to resolve the data heterogeneity and data sharing problems in the construction industry.	Due to the multi-party nature of the construction projects, data integration and exchange from various domains that may vary in data structure and type are needed. A centralized collaboration system is perceived to be inefficient and impractical for the multi-party and temporary nature of the construction industry as participants often hesitate to store and share their data in a 3rd party central location.	Proposed the creation of an ontology as a common data model (using Protégé) and data transfer and mapping based on Semantic Web technology (using SAWSDL) for translating one data schema to another, mapping from web service message to domain ontology and back (XML to OWL, OWL to XML). Data stay on their owner, are exchangeable, and do not have to be stored in a 3rd party central repository.
P6	(Gnimpieba Zanfack et al., 2015)	Present the development of a cloud-based and service-oriented bus for business interoperability and integration of multiple technologies in logistics flows.	How supply chain partners will share data with all partners while (1) dealing with confidentiality, security, access right, and service level agreement, and (2) merging proprietary protocols to a common uniform protocol among partners. Most solutions for system integration in the supply chain sector led to the use of ESB, but the proposed architectures that involve it rarely mention the notion of notification and real-time event processing.	Call for such ESB to enable data and event sharing and notification between supply chain actors that can be based on Publish/Subscribe pattern arching from Event-Driven Architecture. Call for such ESB to incorporate Protocol Adapter for transforming the protocol of a message sent by a client to another unique format.

P7	(Zhao & Liang, 2015)	Discuss the enterprise information exchange within the marine sector based on XML to achieve a significant performance improvement over the state-of-the-art information exchange model.	Low level of information exchange technology between logistics nodes and the big difference between departmental data representation and service processes. Non-uniform marine data standards, a wide variety of sources, and heterogeneous data sources.	Propose information integration and sharing mechanism based on RFID middleware. Establish an ontology describing terms (concepts, properties, and instances) in the marine sector and their relationships, then achieve information exchange based on XML between RFID middleware and enterprise by data adapters.
P8	(Andreeva et al., 2016)	Propose the semantic solution for data exchange in dynamically changing supply networks.	Despite plenty of data standards (UBL, UN/CEFACT, GS1 Logistics Interoperability Model), seamless information exchange between information systems of supply network parties is hindered by poor semantic interoperability.	Introduce a new ontology-based data metamodel of supply networks by ontologizing standards in the logistics domain. Present the ontology in OWL format and formulate a set of queries to the data model based on competency questions that are relevant to the concerned parties.
P9	(Campos et al., 2016)	Facilitate a traceability view of data exchange between partners' heterogeneous systems in a supply chain.	Geographically dispersed enterprises rise common traceability data sharing problems, i.e., disconnected product information, lack of data integrity, inability to inter-relate product data, and visibility across the supply chain.	Highlight that traceability information (data provenance) needs to be integrated, standardized, and linked with different electronic views of the product data to address common traceability data sharing problems and technical challenges.
P10	(Hofman, 2016)	Based on an IT typology, this paper investigates	Multiple protocols and data-sharing architectures (Service-	Messaging, SOA, and EDA are suitable for Transaction

		suitable technical protocols for data sharing supporting supply and logistics innovations.	Oriented, Event-Driven, and Messaging Architectures) exist, leading to the question of which one could best meet data-sharing requirements.	Management IT applications, but Visibility applications are suggested to follow EDA. Complex-Event Processing IT Applications are suggested to follow the EDA approach, which is normally provided by most ESBs.
P11	(Schöggl et al., 2016)	Provides key industrial requirements that necessitate a software solution to enable data exchange for supply chain sustainability.	Data collation is both company-specific and a bilateral affair between a specific company and its direct suppliers, due to the lack of trust, standardization, and mechanism to maintain confidentiality.	Call for a framework or a set of software solutions that can facilitate efficient data exchange and can alleviate confidentiality issues.
P12	(Tran et al., 2016)	Investigate how managers perceive risks associated with sharing information with trading partners, and how they attempt to mitigate them.	Information exchange is seen as a trade-off between efficiency and the responsiveness of information resources when complex information systems are involved. Organizations are often reluctant to share complete information due to (1) security risks and sensitive information leakage, and (2) the adverse competitive implications.	Call for a solution that can (technically) enforce agreements on confidential information (e.g., pricing information, customer details) and legal contracts. Call for a solution that enables information sharing and ensures that data are only shared among intended/trusted parties with secure services.
P13	(Scholz et al., 2018)	Deal with data exchange and multi-entity collaboration aspects in combination with interoperability challenges	When talking about collaboration in a supply chain, several questions arise related to the confidentiality of data and	Call for a platform that is applicable to any supply chain with similar characteristics and based on a bottom-up approach of gathering

		related to the integration among multiple process data collection tools and advanced planning systems.	agreements on cost allocations between partners. Implementing technical solutions and standards to apply a collaborative approach through data sharing can be a big investment burden for SMEs.	existing solutions for different pieces of a supply chain. Call for a flexible data structure that allows format changes on a case-by-case basis rather than a rigid data specification.
P14	(Verhoosel et al., 2018)	Describe the design and engineering of the semantic approach (ontology) to enable interoperability between data sources when different sources are combined.	Visibility in the supply chain requires sharing of data across the entire supply chain and data sources that are accessible online for continuous real-time usage.	Proposed a platform that enables access to a variety of data sources via linked data web-based mechanisms (based on RDF, stored in Apache Jena Fuseki, and accessible through SPARQL interface) and incorporates security mechanisms to ensure that each data producer remains in control of who gets access to the stakeholders' data.
P15	(Wang, X. X. et al., 2018)	Propose the prototype of a social collaborative integration platform for urban distribution to facilitate coordination between actors featured in the crowdsourcing and sharing economy.	The increasing availability of and access to traffic and logistics data from the vehicle, freight, etc., calls for better collaborative decision support systems for urban distribution areas.	Proposed a collaborative information portal that connects stakeholders and offers co-created and open-source apps utilizing open R packages for data analysis, data mining, and other machine learning techniques for LSPs, which their usage might come with some pricing scheme.
P16	(Abebe et al., 2019)	Lay the foundation for an approach to enabling	A conservative approach to interoperability based on	Propose an architecture and a proof-of-concept for trusted data-

		trusted data exchange between two distinct blockchain networks.	traditional point-to-point integration is insufficient for preserving the underlying trust decentralized networks provide. Applications addressing enterprise use cases impose requirements, i.e., scalability, privacy, confidentiality, and audibility.	sharing between two blockchain networks, which demonstrate a trusted cross-network data transfer accompanied by a proof that represents the consensus view of the network.
P17	(Bicocchi et al., 2019)	Propose an architectural framework and key requirements conjugating features of both service-oriented and data-sharing architectures for access to services, aggregation of data, and orchestration of processes.	The main difficulty of data sharing lies in the lack of agreement on the adopted data models and languages, the vocabularies and schema to describe the data, and the semantics of data values. Relationships between data exposed by ISs are usually expressed through mappings between target data instances and more than one data source.	Devise an interoperability platform to support agile and global multi-tier supply chains, e.g., dataspaces, peer data management systems, and polystores. Data is organized in a dataspace of data sources that can exchange data through mappings and support dynamic configurations. Provide rich semantic descriptions to (web) services to support discovery and execution that include keywords or synonyms.
P18	(Dalmolen, Bastiaansen, Kollenstart, et al., 2019)	Elaborate on an open network-model approach for maintaining sovereignty over metadata.	Greater awareness of organizations' control over data sovereignty in supply chains, embedded within the digital data-sharing infrastructure itself, is needed. Existing data-sharing architectures lack data sovereignty functionalities, with	Adopt a service-oriented business architecture that incorporates data brokering and clearing house roles and infrastructures to support sovereignty over (meta)data, such as the IDS initiative. Incorporate metadata creation and utilization processes for data providers and consumers to

			security mainly focused on encrypted data transactions, user authentication, and authorizations.	manage data-sharing transactions and agreements.
P19	(Debicki & Kolinski, 2019)	Outline what future possible solutions could be to overcome inconveniences and that all parties could benefit from business models involving an integration platform in global supply chains.	Most enterprises do not see the need to introduce or develop a new integration platform, but rather to integrate the platforms already in use. A common transmission protocol and API format structure like JSON exists, but there also exist different schema and semantic model being used by partners. Most enterprises are afraid of losing sensitive/critical data for the company.	A call for IT integration tools that support quick and easy integration that only take days instead of months to set up and switch or migrate the integration landscape.
P20	(Hofman, W., 2019)	Analyze the implementation of open standards by providing an overview of the available ones based on different implementation strategies for B2B and B2G in international trade and logistics.	Both JSON (de facto standard in software development) and RDF (de facto standard for representing linked data) are not yet applied in the supply and logistics sector. Pragmatic standards for modalities in logistics are either based on XML or EDI, e.g., ISO[59] or GS1 standards based on XML	Propose a development for data-sharing in supply and logistics with key elements of the IDS. Call for a solution to capture the choreography (as BPM) of business interactions. Call for the implementation of open standards by an organization and the representation of open

[59] www.tln.nl

			definitions (XSDs) for road modality and SMDG[60] based on EDI for sea modality.	standards, preferably in OWL representation.
P21	(Hofman, W. J., 2019)	Propose a methodological approach for the specification of data that can be shared with rapid deployment by a blockchain-based or peer-to-peer infrastructure.	Inter-organizational process synchronizations for supply-chain optimizations require data-sharing rules. The adoption of blockchain-based infrastructure, Distributed Ledger Technology (DLT), presents several obstacles, e.g., energy consumption and low transaction rate.	Capture business transactions and movements of goods in the real world with a Digital Twin, which represents concepts with properties and associations based on an ontology. Present a data-sharing reference model for supply and logistics deployed with a blockchain-based infrastructure, applied to a use case in commodity trading.
P22	(Wang et al., 2019)	Analyze the relationship between blockchain technology and the sustainable development of a resource-conserving society.	The poor communication of information, information leakage, and lack of trust leads to the inability of supply chain partners to fully share and interact with information, which slows down supply chain efficiency.	Propose the implementation of blockchain technology as the bottom layer of the supply chain logistics information ecosystem to prevent privacy leaks and enforce privacies.
P23	(Abu-elezz et al., 2020)	Explore and categorize the benefits and threats of blockchain technology application in a healthcare system.	Little is known about the benefits and threats of blockchain technology, especially in healthcare	Eight threats of blockchain application: installation costs, interoperability issues, lack of technical skills (maintenance), regulation issues, scalability,

[60] https://smdg.org/

				energy consumption, and slow processing speed.
P24	(Bastiaansen et al., 2020)	Contribute to the development of the network-model data-sharing ecosystem by identifying architectural options for realizing interoperability on the legal concepts for controlled data sharing.	Organizations increasingly require improved data control capabilities that prevent their shared data from being misused. Data sovereignty capabilities might impose end-users with potential customer lock-in and major integration efforts to manage data sovereignty over multiple data-sharing relationships.	Adoption of network-model approach as an alternative to the traditional hub-model approach, e.g., the IDS, that provides a single-entry point for the end-user with peer-to-peer data sharing and agreed upon protocols for defining and enforcing data control capabilities across multiple data sharing environments.
P25	(Carvalho et al., 2020)	Propose collaboration networks between logistics stakeholders that provide interoperable, low-cost, reliable, and secure data exchange without requiring significant IT developments.	The demand for (1) interoperable data exchange mechanisms that enhance secure connectivity between stakeholders and facilitate the integration of existing legacy systems that lack data interoperability methods, and (2) digital process and assets representation for the physical reality in the logistics sector.	Propose the 4-corner topology model, in which two distinct (back-end) systems can securely and reliably exchange data between them through the access points (eDelivery nodes) that are interfacing them.
P26	(Piest, Iacob, et al., 2020)	Propose an approach revolving around the application of the IDS concepts that aim at lowering the barriers of logistics SMEs to use,	SMEs in logistics experience practical barriers to using, sharing, and exploiting data in their operational processes, including the abundance of proprietary data schemas, the	Design of the federated logistics data space architecture based on the principles of IDS, comprising the Connector Store and the Interoperability Simulator.

		exchange, share, and exploit real-time data.	cost of implementing standards, the lack of trust among partners, and the uncertainty of benefits and return on investments.	Incorporate existing standards for logistics (e.g., eCMR[61] and OTM), data services offered by the NLIP[62], and standardized agreements such as iSHARE [63].
P27	(Tan et al., 2020)	Present a reference framework for green logistics based on blockchain to reach the sustainable operations of logistics, with the integration of the IoT and big data.	Many packages carry a large number of customers' sensitive information, leading to mistrust among stakeholders for building a cooperative relationship. The implementation of blockchain for green logistics can be counterintuitive as it requires large data storage capacity, large network overhead, and high operational, electricity, training, and maintenance costs. The incentive mechanism in the logistics industry to validate every transaction is also still questionable.	Adopt blockchain technology to enable data sharing among stakeholders, while improving transparency among stakeholders and establishing trust (through the consensus mechanism, fulfillment of contracts, and trusted payment process)
P28	(Voswinckel et al., 2020)	Analyze and conceptually compare the existing traceability processes for the food supply chain (e.g., GS1 EPCIS) with	The use of blockchain technology in industrial applications still lacks standardizations, in which, existing implementations of blockchain technologies for increasing transparency in	Propose a decentralized storage network architecture in which data sharing takes place between the company's node within the blockchain network and each node

[61] https://cargonaut.nl/
[62] https://www.nt.nl/glossary/nlip-neutraal-logistiek-informatie-platform
[63] https://ishare.eu/

		blockchain-based processes.	supply chains have not yet been analyzed sufficiently.	is connected to the company's integration platform. Privacy and data protection are ensured through role and access rights management that govern who may read what kind of data from which network partner.
P29	(Cirullies & Schwede, 2021)	Apply the DSRM to (1) state the requirements for shared digital twins based on five industrial use cases and (2) present a concept for a shared digital twin providing data on demand.	Supply chain partners seem to be willing to share critical data if they know that it will not be used against them. The necessity to adapt local data formats to a global standard poses a major problem for companies to share data over their company boundaries.	Propose requirements for an on-demand shared digital twin with a focus on data decentralization, data standardization, and data sovereignty through the adoption of RDF format to comply with global ontology and IDS Connector to share data and define data usage policies.
P30	(Karatas & Gultekin, 2021)	Create a design pattern that supports all kinds of data exchange and offers a holistic security solution for problems inherited from heterogeneous data.	Due to the use of different data types, systems can become complex, which can lead to problems related to data security.	Enforce security in data exchange through the adoption of AS2 encoding, digital signing, and message encryption.
P31	(Bouter et al., 2022)	Present the development of a modular ontology to support event-sharing in the logistics domain across modalities (e.g., road and	How to reconcile (ontological) models for the logistics sector, which has been constituted by, e.g., the existing SmartRail[64] ontology for the rail modality and	Propose the modular ontology that facilitates the reuse of existing models of the various transport modes by (1) defining the functional requirements (CQs) including the

[64] https://ontology.tno.nl/smart-rail/

		rail) as envisioned by the DTLF.	the ITM for the road modality, etc.	SPARQL queries to verify the CQs and (2) integrating the semantic model with a data space architecture to validate the transformation of various data sources (e.g., OTM or enterprise data) using an RML-based mapper,
P32	(Frey et al., 2022)	Describe the implementation of the technical platform supporting the Bauhaus.MobilityLab, which focuses on data sharing based on the concepts of the IDS and the integration of AI algorithms.	The output quality of AI algorithms is determined by the quality of the input data, hence the call to set up a cross-domain living lab.	Propose the Bauhaus.MobilityLab, a secure and easy-to-use platform to develop new services for smart city applications with basic requirements as follows: Support the creation of innovative smart city services. Provide access to relevant data sources and external platforms. Isolate data exchange to the defined gateways that can enforce data sources with usage policies. Support integration of external systems and AI algorithms.
P33	(Heinbach et al., 2022)	Establish a design science research project to conceptualize a shared Freight Service Intelligence Platform (FSIP) and introduce freight service intelligence as an interdisciplinary research field.	An interview with experts reported that such a platform must support real-time monitoring of transport progress and conditional state. Data exchange based on electronic documents requires a trustworthy environment, in	Blockchain technologies are suggested for such a data-sharing platform to allow secure data exchange and communication of participating transport stakeholders for freight documents, transaction logs, and events from freight transport assets.

			which a blockchain-based environment is suggested.	
P34	(Kazantsev et al., 2022)	Explore challenges of small and medium-sized enterprises (SMEs) in collaborative manufacturing amid the emergence of a dedicated B2B platform.	Smaller suppliers are unwilling to give information/share (forecasting) data with them due to the potential intention of espionage for competition. Global integration of suppliers can be challenging due to several proprietary IT systems and the lack of standards and interfaces for data transfer.	Listed factors and barriers that are impeding SMEs' transparency and willingness to share information and collaborate.
P35	(Top et al., 2022)	Explores the further operationalization of the FAIR principles in agriculture and food by investigating the prerequisites before data can be effectively shared and reused.	Data cannot be shared only by an open-by-default policy. Data collected by research projects are often incidental and not well-structured, as opposed to data collected in the industry which are coming from their own operations and often are not shared due to competitive concerns.	Introduce the FAIRification process to data in the supply chain to facilitate potential users to: Find a dataset based on the metadata describing it through a single access point on the web. Decide from the metadata whether the dataset is suitable for their task and whether it is allowed to be used for that purpose. Import the dataset into their data processing environment, including mapping the variables into their tool. Enrich the original data with metadata that links to the updated data.

Appendix C. Transcript of Validation Scenario at eMagiz

ENGD BUSINESS & IT

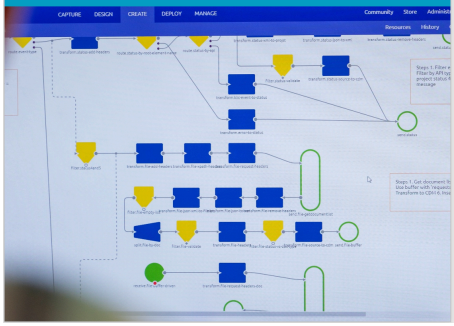
CLICKS CONNECTOR STORE

PROTOTYPE DEMONSTRATION & VALIDATION SCENARIO AT EMAGIZ

DANNIAR REZA FIRDAUSY

21 MARCH 2023

UNIVERSITY OF TWENTE.



UNIVERSITY OF TWENTE.

1. BACKGROUND AND MOTIVATION

This EngD research project, entitled "Designing a Connector Store for a Logistics Data Space", is part of the work package defined within the CLICKS Project that stands for Connecting Logistics interfaces, Converters, Knowledge, and Standards. One of the focuses of this project is to design and propose an approach to make the exchange and sharing of real-time data more accessible and secure to logistics SMEs. Two main problems within the data-sharing context to be addressed in this research are regarding data interoperability[1] and data sovereignty[2].

To achieve this, we try to adopt a design proposed by the initiative called the International Data Spaces (IDS) currently developing in Europe, which introduces the so-called IDS Connector and IDS Data Apps application components. An IDS Connector is described as a gateway at a company's logical border that defines web-service interfaces for external parties to exchange data. An IDS Connector supports data owners' data sovereignty by allowing them to attach and technically enforce usage policies of their data. For this to work, this will require every participant to have and use their own IDS Connector. To manage data interoperability between participants' enterprise systems, the IDS Connector supports the execution and orchestration of IDS Data Apps, which are described as independent and reusable software applications that provide additional data processing and transformation functionalities for the IDS Connector.

In this EngD project, the Connector Store acts as a metadata repository that stores and provides data-sharing participants with information regarding (1) offered data resources or services, (2) participating IDS Connectors to request and exchange data, (3) offered IDS Connectors for candidate participants to join the data space, (4) reusable IDS Data Apps to facilitate participants' data processing tasks, and (5) software or service providers offering the IDS Connectors or Data Apps, among others. In addition, due to its identical functionalities with the eMagiz platform, in this research, we also consider eMagiz's API Gateway solution as a representation of these IDS Data Apps to perform data transformation and message routing and then try to integrate the deployed integration landscape with the IDS Connector prototype for solving data structure or message format issues.

2. DEMONSTRATION SCENARIO

Today's supply chain consists of collaborations and data-sharing among multiple parties. These collaborations, nowadays, aim to increase efficiency and environmental sustainability. Especially, in the transport logistics sector, reducing CO2 emissions becomes a prominent objective. National and international governmental bodies released a set of laws and requirements that should be fulfilled by actors in the supply chain. The Regulatory Body makes sure that all these laws and requirements are met. To enforce this, the Regulatory Body requires Transport Companies to report and share their transport trip data. Based on this data the Regulatory Body will calculate the tax that the company will pay accordingly. Sharing this data means that it opens up the possibility for the company to pay less tax due to the lower carbon footprint they produced, but also risks compromising their competitive advantage due to sensitive and confidentiality concerns. On the other hand, sharing such data also requires data transformation and, possibly, data integration efforts from other data sources before it can be received by the Regulatory Body.

[1] https://www.data4stats.com/initiatives/data-interoperability-collaborative
[2] https://internationaldataspaces.org/why/data-sovereignty/

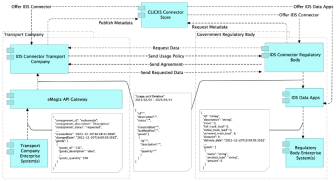


Figure 1 Demonstration Scenario - Mockup Transport Carbon Emission Tax Reporting

Therefore, in this demonstration, we try to solve this challenge by applying the IDS Connectors and IDS Data Apps for each of these actors that are brokered by the Connector Store. A preview of the system interaction and data flow landscape for this scenario is shown in **Figure 1** above. This is the landscape that we will try to simulate in the demonstration. From there on we will brainstorm future opportunities and other possible business cases. From these validations, the thesis intends to collect expert opinions (either from eMagiz or other case owners) by means of questionnaires and interviews based on the questions presented in **Table 1** below.

Table 1 Validation Questionnaire and Interview Questions for IDS Adoption and Connector Store Implementation

No.	Goals	Questions	Score	Opinions
1	Facilitate data transformation between standards and data formats	To what extent can such an architecture, through its demonstrator, positively contribute to facilitating data transformation between standards and data formats?	3	The data transformation is kept out of the IDS connectors for a purpose as that would complicate. It does mean that a solution needs to be found for (temporary) storing of information
2	Stimulate data findability, accessibility, interoperability, and reusability	To what extent can such an architecture, through its demonstrator, positively stimulate data findability, accessibility, interoperability, and reusability?	4	Having a shared and standardized catalog on top of the transformation capabilities of other parts (i.e. eMagiz) can help
3		To what extent can such stimuli motivate data space participants to devise new business models and value co-creation?	3	Though to tell for me. The challenge with all standarization is getting a critical mass to adopt the standard. As it requires quite a lot of back and forth communication between systems adoption might be low as not all current systems can facilitate this (yet).
4	Enable quick onboarding to a data-sharing ecosystem	To what extent can such an architecture, through its demonstrator, enable potential participants' quick onboarding to a data-sharing ecosystem?	3	Depending on the architecture of the client it can be quick or very slow.
5	Enable data owners to be self-determined regarding the usage of their data assets by data users	To what extent can such an architecture, through its demonstrator, enable data owners to enforce usage control over their data assets by data users?	5	When implemented according to the specifications it allows complete control of the data by the data owner.
6		To what extent can such an architecture, through its demonstrator, support data owners to maintain the confidentiality aspect of their data assets during and after the data exchange?	4	As data does leave the system it can still happen that in the future a data breach happens at the client side outside your influence but that does compromises the confidentiality of the data.

Appendix D. Transcript of Validation Scenario at SUTC

ENGD BUSINESS & IT

CLICKS CONNECTOR STORE

PROTOTYPE DEMONSTRATION & VALIDATION SCENARIO AT SUTC

DANNIAR REZA FIRDAUSY

21 MARCH 2023

UNIVERSITY OF TWENTE.

1. BACKGROUND AND MOTIVATION

This EngD research project, entitled "Designing a Connector Store for a Logistics Data Space", is part of the work package defined within the CLICKS Project that stands for Connecting Logistics interfaces, Converters, Knowledge, and Standards. One of the focuses of this project is to design and propose an approach to make the exchange and sharing of real-time data more accessible and secure to logistics SMEs. Two main problems within the data-sharing context to be addressed in this research are regarding data interoperability[1] and data sovereignty[2].

To achieve this, we try to adopt a design proposed by the initiative called the International Data Spaces (IDS) currently developing in Europe, which introduces the so-called IDS Connector and IDS Data Apps application components. An IDS Connector is described as a gateway at a company's logical border that defines web-service interfaces for external parties to exchange data. An IDS Connector supports data owners' data sovereignty by allowing them to attach and technically enforce usage policies of their data. For this to work, this will require every participant to have and use their own IDS Connector. To manage data interoperability between participants' enterprise systems, the IDS Connector supports the execution and orchestration of IDS Data Apps, which are described as independent and reusable software applications that provide additional data processing and transformation functionalities for the IDS Connector.

In this EngD project, the Connector Store acts as a metadata repository that stores and provides data-sharing participants with information regarding (1) offered data resources or services, (2) participating IDS Connectors to request and exchange data, (3) offered IDS Connectors for candidate participants to join the data space, (4) reusable IDS Data Apps to facilitate participants' data processing tasks, and (5) software or service providers offering the IDS Connectors or Data Apps, among others. Due to its identical functionalities with the eMagiz platform (a Dutch Integration Platform as a Service vendor that supports companies in the logistics and construction industries), in this research, we consider eMagiz's API Gateway[3] solution as one of the representations of these IDS Data Apps to perform data transformation and message routing and then try to integrate the deployed integration landscape with the IDS Connector prototype for solving issues of data structures and message formats.

2. DEMONSTRATION SCENARIO

Today's supply chain consists of collaborations and data-sharing among multiple parties. These collaborations, nowadays, aim to increase efficiency and environmental sustainability. Especially, in the transport logistics sector, reducing CO2 emissions becomes a prominent objective set by the (inter)national governmental bodies, who are governing the industrial actors that are active in the supply chain. These bodies make sure that all these laws and requirements are met. To enforce this, in The Netherlands, the Centraal Bureau voor de Statistiek (CBS) requires transport companies to report and share their transport trip data using a standardized Open Trip Model (OTM) data model[4]. Based on this data CBS will calculate the logistics sector's overall performance and carbon footprint. Sharing this data in a standardized format means that it opens up the possibility for the logistics partners to exchange transport orders and proof of delivery seamlessly. However, this also means they risk compromising their competitive advantage due to sensitive and confidentiality concerns. On the other hand, sharing such data also

[1] https://www.data4dgs.org/initiatives/data-interoperability-collaborative
[2] https://internationaldataspaces.org/why/data-sovereignty/
[3] https://emagiz.com/api-management/
[4] https://otm6.opentripmodel.org/RawSon/Open-Trip-Model

requires data transformation and data integration efforts from diverse data sources before it can be received by either the system used by the CBS (i.e., the VESDI interface[5]) or enterprise systems of other logistics partners.

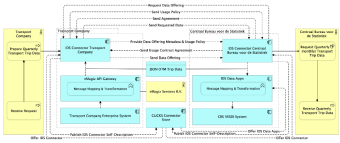


Figure 1 Demonstration Scenario – VESDI Project Transport Trip Reporting for Carbon Emission Monitoring

Therefore, in this demonstration, we try to solve this challenge by applying the IDS Connectors and IDS Data Apps for each of these actors that are brokered by the Connector Store. A preview of the system interaction and data flow landscape for this scenario is shown in **Figure 1** above. This is the landscape that we will try to simulate in the demonstration. From there on we will brainstorm on the contribution of these application prototypes in addressing data interoperability and sovereignty challenges. From these validations, the thesis intends to collect expert opinions by means of questionnaires and interviews based on the questions presented in **Table 1** below.

Table 1 Validation Questionnaire and Interview Questions for IDS Adoption and Connector Store Implementation

No.	Goals	Questions	Score	Opinions
1	Facilitate data transformation between standards and data formats	To what extent can such an architecture, through its demonstrator, positively contribute to facilitating data transformation between standards and data formats?	5	Absolutely necessary and with high added value
2	Stimulate data findability, accessibility, interoperability, and reusability	To what extent can such an architecture, through its demonstrator, positively stimulate data findability, accessibility, interoperability, and reusability?	4	Valuable but complicate due to high variety of data elements and systems used. This make findability challeging
3		To what extent can such stimuli motivate data space participants to devise new business models and value co-creation?	2	New business models are influenced by many more elements than data alone. E.g. market circumstances, investment needed, position in value chain
4	Enable data owners to be self-determined regarding the usage of their data assets by data users	To what extent can such an architecture, through its demonstrator, enable potential participants' quick onboarding to a data-sharing ecosystem?	3	See point 2. Quick onboarding on a large scale is complex due to high variety of data elements en systems used
5	Enable data owners to be self-determined regarding the usage	To what extent can such an architecture, through its demonstrator, enable data owners	4	

[5] https://www.sutc.nl/en_US/het-vesdi-project

	of their data assets by data users	to enforce usage control over their data assets by data users?		
6		To what extent can such an architecture, through its demonstrator, support data owners to maintain the confidentiality aspect of their data assets during and after the data exchange?	4	

Appendix E. Transcript of Validation Scenario at TU Delft

ENGD BUSINESS & IT

CLICKS CONNECTOR STORE

PROTOTYPE DEMONSTRATION & VALIDATION SCENARIO AT TU DELFT

DANNIAR REZA FIRDAUSY

21 MARCH 2023

UNIVERSITY OF TWENTE.

1. BACKGROUND AND MOTIVATION

This EngD research project, entitled "Designing a Connector Store for a Logistics Data Space", is part of the work package defined within the CLICKS Project that stands for Connecting Logistics interfaces, Converters, Knowledge, and Standards. One of the focuses of this project is to design and propose an approach to make the exchange and sharing of real-time data more accessible and secure to logistics SMEs. Two main problems within the data-sharing context to be addressed in this research are regarding data interoperability[1] and data sovereignty[2].

To achieve this, we try to adopt a design proposed by the initiative called the International Data Spaces (IDS) currently developing in Europe, which introduces the so-called IDS Connector and IDS Data Apps application components. An IDS Connector is described as a gateway at a company's logical border that defines web-service interfaces for external parties to exchange data. An IDS Connector supports data owners' data sovereignty by allowing them to attach and technically enforce usage policies of their data. For this to work, this will require every participant to have and use their own IDS Connector. To manage data interoperability between participants' enterprise systems, the IDS Connector supports the execution and orchestration of IDS Data Apps, which are described as independent and reusable software applications that provide additional data processing and transformation functionalities for the IDS Connector.

In this EngD project, the Connector Store acts as a metadata repository that stores and provides data-sharing participants with information regarding (1) offered data resources or services, (2) participating IDS Connectors to request and exchange data, (3) offered IDS Connectors for candidate participants to join the data space, (4) reusable IDS Data Apps to facilitate participants' data processing tasks, and (5) software or service providers offering the IDS Connectors or Data Apps, among others. Due to its identical functionalities with the eMagiz platform (a Dutch Integration Platform as a Service vendor that supports companies in the logistics and construction industries), in this research, we consider eMagiz's API Gateway[3] solution as one of the representations of these IDS Data Apps to perform data transformation and message routing and then try to integrate the deployed integration landscape with the IDS Connector prototype for solving issues of data structures and message formats.

2. DEMONSTRATION SCENARIO

Today's supply chain consists of collaborations and data-sharing among multiple parties. These collaborations, nowadays, aim to increase efficiency and environmental sustainability. Especially, in the transport logistics sector, reducing CO2 emissions becomes a prominent objective set by the (inter)national governmental bodies, who are governing the industrial actors that are active in the supply chain. These bodies make sure that all these laws and requirements are met. To enforce this, in The Netherlands, the Centraal Bureau voor de Statistiek (CBS) requires transport companies to report and share their transport trip data using a standardized Open Trip Model (OTM) data model[4]. Based on this data CBS will calculate the logistics sector's overall performance and carbon footprint. Sharing this data in a standardized format means that it opens up the possibility for the logistics partners to exchange transport orders and proof of delivery seamlessly. However, this also means they risk compromising their competitive advantage due to sensitive and confidentiality concerns. On the other hand, sharing such data also

[1] https://www.data4dps.org/initiative/data-interoperability-collaborative
[2] https://internationaldataspaces.org/what/data-sovereignty/
[3] https://emagiz.com/api-management/
[4] https://otm9.opentripmodel.org/#/section/Open-Trip-Model

requires data transformation and data integration efforts from diverse data sources before it can be received by either the system used by the CBS (i.e., the VESDI interface[5]) or enterprise systems of other logistics partners.

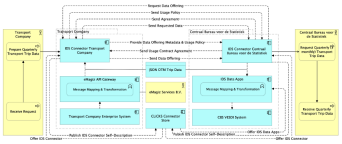


Figure 1 Demonstration Scenario – VESDI Project Transport Trip Reporting for Carbon Emission Monitoring

Therefore, in this demonstration, we try to solve this challenge by applying the IDS Connectors and IDS Data Apps for each of these actors that are brokered by the Connector Store. A preview of the system interaction and data flow landscape for this scenario is shown in **Figure 1** above. This is the landscape that we will try to simulate in the demonstration. From there on we will brainstorm on the contribution of these application prototypes in addressing data interoperability and sovereignty challenges. From these validations, the thesis intends to collect expert opinions by means of questionnaires and interviews based on the questions presented in **Table 1** below.

Table 1 Validation Questionnaire and Interview Questions for IDS Adoption and Connector Store Implementation

No.	Goals	Questions	Score	Opinions
1	Facilitate data transformation between standards and data formats	To what extent can such an architecture, through its demonstrator, positively contribute to facilitating data transformation between standards and data formats?	4	The Connector Store architecture effectively supports the execution and orchestration of IDS Data-Apps: independent and reusable software applications that provide data processing and transformation functionalities for the IDS Connector. This enables efficient and seamless data transformation between different standards and formats, making data exchange more accessible to organizations. Concrete examples include 1) converting Non-OTM formatted message to OTM-compliant or 2) OTM message to Regulatory Body format. One point of improvement is to examine the trade-off between the specificity and generalizability of the scenario. Beyond the OTM formatted scenario, how hard is it to the development such mapping? What are incentive models to develop such mapping? This needs to be examined in further research.
2	Stimulate data findability, accessibility, interoperability, and reusability	To what extent can such an architecture, through its demonstrator, positively stimulate data findability, accessibility, interoperability, and reusability?	4	1. Data Findability: The Connector Store is a metadata repository that stores and provides data-sharing participants with information regarding offered data resources or services, participating IDS Connectors, and more. This approach enables users to quickly discover and locate the required data, improving data findability. 2. Data Accessibility: By leveraging IDS Connectors as gateways for data exchange and offering a user-friendly platform, the architecture ensures that authorized parties can securely and easily access data. 3. Data Interoperability: The architecture addresses data interoperability challenges by facilitating data transformation between different standards and formats and improving communication between participants' enterprise systems (e.g., refer to the opinion in the #1 goal). While the demonstration itself may not have placed a significant focus on data reusability, the underlying architecture of the Connector Store indeed holds the potential for enhancing data reusability. The IDS Data Apps are designed as independent and reusable software applications that provide data processing and transformation functionalities for the IDS Connector. This inherent reusability allows organizations to efficiently repurpose these

[5] https://www.sutc.nl/en_US/het-vesdi-project

				applications for various data processing tasks, minimizing redundant development efforts and maximizing the value derived from data.
3		To what extent can such stimuli motivate data space participants to devise new business models and value co-creation?	5	The Connector Store architecture can motivate data space participants to devise new business models and co-create value. Examples: 1. Data space/data marketplace operators: By adopting the Connector Store as a shared service, data marketplace operators can focus on their core value creation activities, such as curating high-quality datasets and providing data analytics services. 2. Data Providers: With the Connector Store architecture in place, data providers can explore secondary business models, such as data monetization, that was previously unexplored. For example, a provider with historical weather data can now offer tailored datasets or insights to specific industries, such as agriculture or renewable energy, which can utilize this data to optimize their operations and reduce risks. 3. Data Consumers: The Connector Store can enable data consumers, such as retail companies, to access and integrate data from various providers to generate valuable insights. For example, by combining transaction data, location data, and social media sentiment analysis, the retail company can develop targeted marketing campaigns, optimize store layouts, and improve customer satisfaction, leading to increased revenue and a more competitive business model. Third-Party Developers: As previously mentioned, the Connector Store encourages the development of reusable IDS Data Apps that can be integrated with the IDS Connector. This creates opportunities for third-party developers to innovate and monetize specialized data processing and transformation applications.
4	Enable quick onboarding to a data-sharing ecosystem	To what extent can such an architecture, through its demonstrator, enable potential participants' quick onboarding to a data-sharing ecosystem?	3	The Connector Store architecture, through its demonstrator, provides a sufficient level of quick onboarding for potential participants, including SMEs. However, SMEs often face challenges such as a lack of awareness and limited technical capabilities, which could hinder their onboarding process. A more streamlined and low-code setup process could be implemented to address these challenges and further enhance the onboarding experience for SMEs. Additionally, there is an opportunity to develop consulting services or establish a data space support center specifically tailored to the needs of SMEs.
5	Enable data owners to be self-determined regarding the usage of their data assets by data users?	To what extent can such an architecture, through its demonstrator, enable data owners to enforce usage control over their data assets by data users?	4	The Connector Store architecture, as demonstrated, offers the substantial capability for data owners to exert usage control over their data assets. Recently, the system has enabled data owners to establish and enforce usage policies, safeguarding their data assets and ensuring adherence to their specified terms. Improvements include more detailed/consumable data usage policies beyond the N use or date validity. Additionally, mechanisms for data withdrawal in case of disputes or unforeseen circumstances could also be incorporated.
6		To what extent can such an architecture, through its demonstrator, support data owners to maintain the confidentiality aspect of their data assets during and after the data exchange?	3	Although the demonstration may not explicitly showcase the confidentiality aspect (it focused on data interoperability and sovereignty), the underlying principles and mechanisms within the Connector Store architecture still hold potential for addressing this concern. The focus on data sovereignty data owners to establish a secure environment for data exchange, which implicitly contributes to maintaining confidentiality. It is essential to recognize that the demonstration is just one example of how architecture can be employed. In future iterations or use cases, confidentiality could be more prominently showcased, highlighting the potential for securely sharing sensitive information within the data space. P.S. Notes that to some degree, data sovereignty also relates to security (or confidentiality) (Hummel et al., 2021). By discussing data sovereignty, we also "implicitly" discussed the security aspect.

DESIGNING ESSENTIAL COMPONENTS FOR LOGISTICS DATA SPACES

Connecting Logistics interfaces, Converters, Knowledge, and Standards

Danniar Reza Firdausy

Data-sharing, especially in the logistics sector, is vital for unlocking innovative business models and empowering supply chain partners to enhance their operations. However, establishing a data-sharing ecosystem requires addressing data interoperability barriers, such as conflicting formats and schema. The lack of technical enforcement for data sovereignty further hinders companies' willingness to share data. Additionally, uncertainties surrounding participating partners and technology adoption pose challenges for upfront investments in data space. This EngD thesis investigates a suitable design for a Connector Store and other essential components in the logistics data space, focusing on streamlining onboarding, reducing data interoperability barriers, and fostering data sovereignty. Through the application prototypes showcased to a panel of experts, this thesis demonstrates the technical feasibility of the proposed logistics data space to effectively address interoperability, sovereignty, and discovery issues. Ultimately, valuable insights derived from the thesis can serve as a foundation for future research in this field.

UNIVERSITY OF TWENTE.

ISBN (print): 978-90-365-5716-0
ISBN (digital): 978-90-365-5717-7
DOI: 10.3990/1.9789036557177