FINAL PROJECT

AUDITABLE DATA SHARING IN LOGISTIC DATA SPACE: DESIGN AND IMPLEMENTATION OF IDS CLEARING HOUSE FOR LOGISTIC DATA SPACE

Yesaya Galatia Maranatha

SUPERVISORS:
dr. ir. M. Van Sinderen (Committee Chair)
dr. J.L. Rebelo Moreira
dr.ing. E. Moritz Hahn
D.R. Firdausy, MSc

JULY 2023

UNIVERSITY OF TWENTE.

ACKNOWLEDGEMENT

As soon as I finished writing my thesis, I purposely left this page for last, allowing myself some time to think and reflect on the journey I have undertaken over the past six to seven months. During this contemplation, I came across a beautiful Bible verse, taken from the epistle to the Galatians, which my middle name refers to. In chapter 6, verse 2, it is written, "**Bear one another's burdens, and so fulfill the law of Christ**." As a follower of Christ, I have always believed in the virtue of both giving and receiving help from others, a practice that demonstrates the love of Christ in achieving certain goals. Thus, I wish to dedicate this page specifically to all those who have provided invaluable help, becoming extensions of Christ's blessings and love within the context of this thesis. This support has not only lightened the load but has also exemplified the essence of the verse itself.

Firstly, I would like to express my deepest sincere gratitude toward dr. ir. Marten J. van Sinderen, dr. Joao L.R. Moreira, dr. Ing. E. Moritz Hahn, and Danniar R. Firdausy for their guidance, support, and insightful direction throughout this research. I am thankful for their expertise and valuable feedback which has not only enriched my understanding of the subject matter but also inspired me to push my boundaries and strive for excellence. In addition, I would also like to thank the University of Twente and the Faculty of Electrical Engineering, Mathematics, and Computer Science (EEMCS) for providing me with scholarships to pursue my Master's study. Also, to the Change Leader and Management Honors Programme for helping me shape my personality, which also contributes to helping me manage and complete this research.

Special thanks to my dad and mom, Wibisono and Nurul, as well as my siblings, Mbak Rena, Mas Putra, and Yesita, for their unending love, unwavering support, countless nights of praying, and boundless encouragement throughout my journey. I also dedicate this thesis and degree to my beloved nephew and niece, Daniel and Eunike. I hope this can be an inspiring journey for them and their growth. Special thanks to my best friend, brother, and roommate, Sam Raka, for sharing laughter, tears, emotional and spiritual support, food, hundred kilometers of biking journey, and most importantly, a friendship that I have never realised and expected I needed. Also, special thanks go to my soul sisters, Kak Archi, and Kak Ju, for always supplying me with anything fun, delicious, or sometimes thrilling while capturing each moment in pictures and videos.

I am also grateful for my friends and communities in Enschede. Firstly to Vriendenkring and Karang Taruna Enschede, specifically to Silvi, Irine, Raka, Aldi, Michael, and Afif – thank you so much for the joy, motivation, and for helping me explore Enschede and the Twente region. I will never forget how we won the EEMCS Autumn Challenge. Secondly, my brothers and sisters in Christ gathered in the International Christian Fellowship Enschede (ICF-E) and the Indonesians Christian Fellowship in Enschede for every gathering, praise, worship, and prayer that keep me on God's path. Lastly, I would like to thank my band ASAF – Michael, Aldi, Raka, Daniel, and Afif – for keeping me entertained while also helping me develop my music skills. May Jesus bless you all!

Enschede, 31 August 2023
Yesaya Galatia Maranatha

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

In the logistics sector, collaboration between stakeholders through information or data sharing is an important key to ensuring smooth and efficient logistic services. The International Data Spaces (IDS), as one of the alternatives, enables a decentralised data-sharing platform in which partners from different sizes can exchange data securely while still granted the capability of being entirely self-determined with regard to their data.

However, the sensitivity and value of the data, as well as the risk of fraud or violation in the transactions due to lack of trust may hinder these enterprises from participating. Therefore, a data space is needed to ensure the trust and auditability of a data exchange transaction, specifically for the logistics sector. This thesis aims to explore the design and implementation of an IDS Clearing House.

Based on the systematic literature review carried out and analysing the IDSA's documentation, four main functionalities and components comprising such clearing house are extracted. Then, a set of architectures and designs of a clearing house based on the findings will be instantiated into a prototype. Later, the prototype will be used to demonstrate and simulate various scenarios of ensuring the auditability in data exchange through implementing an IDS Clearing House.

Keywords: Auditable data sharing, international data space, logistic data space, clearing house, trust intermediary, reference architecture, enterprise architecture

1. INTRODUCTION

1.1 BACKGROUND

The fourth industrial revolution or Industry 4.0 has modernised today's businesses and enterprises in most sectors. The modernisation includes digitalisation and automation of their business processes to increase their productivity, efficiency, and growth (Jafari et al., 2022). The logistics sector specifically has a great potential to maximise those benefits due to the vitality of the logistics sector in today's economy (Barenji & Montreuil, 2022). In the Netherlands, the logistics sector is one of the top sectors contributing to the country's economy as well as in Europe and worldwide. Despite the country's population being only a quarter of a percent worldwide, the logistics sector has contributed nearly four percent to worldwide trade[1]. Furthermore, the logistics sector in the Netherlands is projected to be sustainable, safe, and competitive by 2050. Fortunately, this can be achieved with the technological advancement during the fourth industrial revolution. For instance, the implementation of Internet of Things (IoT) devices, development of Artificial Intelligence (AI) systems, and specifically open data sharing or data exchange networks promises enterprises the means to increase their efficiencies and performances, specifically for those in the logistics sector(Jafari et al., 2022).

Data exchange is one of the technological advancements that facilitate enterprises to share or consume each other's data. While data was considered as a private asset of an organisation, data exchange has shifted this paradigm into foreseeing the potentiality of the data to unveil new value through collaboration with other partners (Bajoudah et al., 2019; Chen et al., 2021). In the logistics sector, collaboration between stakeholders through information sharing is an important key to ensuring smooth and efficient logistic services. However, mostly these happen through conventional methods such as telephone calls, e-mails, or physical documents that are prone to human error and lead to inefficiency (Imeri & Khadraoui, 2018). Therefore, data exchange should be able to increase the efficiency and effectiveness of logistics processes within the collaborating enterprises in this sector (Iacob et al., 2019). In conjunction with collaboration between partners, data exchange could offer greater benefits to the participating logistics enterprises. For instance, minimising and mitigating risk, minimising losses, ensuring high-quality and reliable services, and also ensuring sustainability (Barenji & Montreuil, 2022).

Despite the benefits aforementioned, logistics enterprises might be hindered from participating in exchanging their data for several various reasons. The nature of logistics data and processes is highly sensitive and requires special attention (Imeri & Khadraoui, 2018). As a result, the lack of trust between participants is one of the biggest challenges for these enterprises to participate in exchanging or sharing their data. The idea and possibility in data exchange to share and consume the data with new and unknown participants may be perceived as an unsafe and risky transaction. For example, the consumer of the data may question the quality of the data being shared. On the other hand, the provider or those who share the data may be concerned about the privacy and security of the shared data (Bargh et al., 2014; Luo et al., 2022). As a result, the data provider is also concerned about their ability to monitor and regain control over their data (Imeri & Khadraoui, 2018; Piest et al., 2021). Considering both concerns from the consumer and provider, it is necessary to consider the auditability and traceability aspect of data exchange in order to nurture trust between these collaborating enterprises.

There are various kinds of initiatives that enable data exchange in various cases and domains. In this study, the International Data Space (IDS) will be the focus. International Data Space (IDS) offers a

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

decentralised approach to data sharing which ensures secure and trusted data sharing to each participating enterprise. Moreover, IDS also ensures trust and the sovereignty of the data of the participant who shared the data (Pettenpohl et al., 2022a). These initiatives now have been adopted into several projects based on specific cases or sectors[2]. In response to the future development of data spaces, IDS has published a set of reference architectures[3], specifications, and standards to adopt IDS in specific use cases. For example, in the logistics sector, there are the Connecting Logistics interfaces, Converters, Knowledge, and Standards or the CLiCKS project which aims to facilitate logistics companies to share and exchange their data safely and securely in the data space[4]. The CLiCKS project is ensuring visibility and efficiency in participating enterprises to perform the data exchange by developing a connector store for logistics enterprises (Firdausy et al., 2022).

Typically, a data space is composed of several kinds of components which each has its own functionality and roles in order to ensure trust and sovereignty in the data exchange (Otto, B. et al., 2019). For example, the IDS Clearing House is one of the intermediary components in the data space to enhance participants' trust in exchanging their data due to its main function to clear and settle the transaction (Bastiaansen et al., 2020; Otto, B. et al., 2019; Pettenpohl et al., 2022a). These main services are somewhat similar to clearing and settlement in the financial domain, however, in IDS Clearing House it is not only limited to financial or commercial uses but also for data exchange purposes. The clearing service, for example, ensures the participants satisfy the required obligation in finance, technical, and legal requirements (Bastiaansen et al., 2020). After the transaction is cleared, then the settlement process may begin which includes documenting each transaction in the form of logging (Bastiaansen et al., 2020). Therefore, IDS Clearing House is able to ensure the traceability of the transaction. The traceability functionality will be useful for resolving conflicts or auditing purposes in case of fraud or violation occurs during the data exchange process (Bastiaansen et al., 2020; Otto, B. et al., 2019).

Despite the reference architecture model and specification of IDS Clearing House being made available, there is still limited research regarding the design or recommendation regarding the implementation of an IDS Clearing House in a data space. Consequently, the sense of urgency in having such component in a data transaction is missing, while IDS Clearing House is considered as an important component in data exchange (Otto, B. et al., 2019). In addition, the missing presence of the standardisation also implies to the various flavours of clearing house component implementation. Take an example from the Mobility Data Space (MobiDS) concept which originally included a standalone IDS Clearing House component, while in practice the clearing house function is minimally included in the Data Marketplace component instead (Drees et al., 2021; Pretzsch et al., 2022). In addition, the development of IDS Clearing House is remained open[5] which has a lot of potential to be implemented according to the need of a specific case or data space.

Therefore, this study will focus on exploring the significance, design, and architecture of an IDS Clearing House to ensure auditability and traceability in the data space, specifically within enterprises in the logistics sector. In regard to logistics enterprises' concerns about the lack of trust and security of the data exchange, this study will focus on the auditability ensured by implementing IDS Clearing House in Logistic Data Space. Later the implementation will take form as a prototype and will be tested and integrated with the CLiCKS project mentioned above. This implementation will be utilised to simulate and demonstrate the auditability and traceability functionalities of an IDS Clearing House in several data exchange scenarios.

[2] https://internationaldataspaces.org/make/use-cases-overview/
[3] https://docs.internationaldataspaces.org/ids-ram-4/
[4] https://www.nwo.nl/en/projects/43919633-0
[5] https://gitlab.com/tno-tsg/core-container/-/tree/master/ids-clearing

1.2 PROBLEM STATEMENT

The problem addressed in this study is the auditability of a data exchange due to the lack of trust between logistics enterprises. While this can be ensured by the IDS Clearing House which is a trusted intermediary in a data space, there is limited research on the urgency, standard, and design of such components. Moreover, there is an inconsistency between the concept and implementation of an IDS Clearing House.

1.3 RESEARCH GOAL

The goal aimed in this study is to understand the urgency of an IDS Clearing House in a data exchange focusing on enterprises in logistic sectors. In addition, the design and recommendation of IDS Clearing House implementation is also the focus. These designs and recommendations are necessary to understand the functionality and uses of a clearing house, specifically in the logistics data space under the CLiCKS project. Later, a set of simulations of IDS Clearing House in various scenarios of data exchange within the logistics dataspace.

1.4 RESEARCH QUESTIONS

In accordance with the research goal aforementioned, a main research question and its relevance to sub-research questions are defined. These questions are necessary to determine the scope and focus of the research. The main research question is formalised as:

"How to enable trusted and auditable data exchange in Logistic Data Space by designing and implementing a Clearing House?"

Based on the main research question above, the following sub-research questions are determined:

1. *What is the state of the art of Clearing Houses in data exchange?*

   The first sub-question aims to explore and understand the current development and implementation of a clearing house or a component that enable auditability and traceability in various data exchange implementation. Later, this understanding will be the main foundation for designing and developing a clearing house in order to ensure auditability in data exchange.

2. *How can a clearing house be designed to operate in a Logistic Data Space?*

   The second sub-question aims to define and recommend a complete set of designs and architectures of a clearing house according to the findings regarding the state of the art of clearing houses. In this study, several designs will be focused on and adjusted according to the requirements and specifications of a logistic data space.

3. *How can the designed clearing house be used with other components in the Logistics Data Space to achieve auditable data exchange?*

   The third sub-question aims to have the set of architectures and designs applied in Logistic Data Space. The functionality and features of the proposed clearing house applied must be able to demonstrate and show the auditability ensured by the clearing house in various cases during data exchange.

4. *How usable, useful, and general is the proposed clearing house?*

   Finally, this sub-question reflects on validating the defined architectures, designs, and implementation of a clearing house in a data exchange. Specifically, in the Logistic Data Space. Moreover, this sub-question also aims to ensure the proposed clearing house is also applicable in other data spaces and business cases.

1.5 RESEARCH APPROACH

Wieringa, R. J. (2014) defines "Design Science" methodology as a research approach for research in software engineering and information system. The methodology consists of five main cycles: Problem Investigation, Treatment design, Treatment validation, Treatment implementation, and Implementation evaluation. These cycles are illustrated in figure 1-1 below. In this study, the first three cycles are adopted due to time constraints and the scope of the study. The first cycle focuses on defining what can be improved through the research. In this research, this will be achieved through understanding the state of the art of the IDS Clearing House by conducting a Systematic Literature Review.

After the first cycle is completed, the second cycle is started by extending the findings and requirements gathered from the literature review into the architecture designs of an IDS Clearing House. Later, these designs are brought into a prototype that demonstrates and simulates the IDS Clearing House and its functionalities within logistic data space in various scenarios. Finally, a validation with experts will be conducted in order to ensure the proposed IDS Clearing House is sufficient and suitable to ensure auditability in data exchange, both in Logistic Data Space or other data spaces. Figure 1-1 illustrates the research approach of this study.

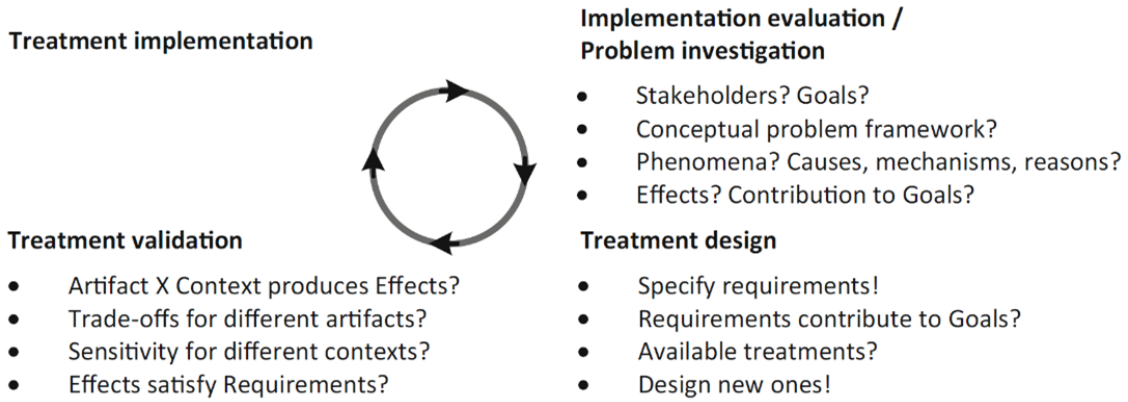


Figure 1-1 Design Science Methodology

1.6 RESEARCH CONTRIBUTIONS

The contributions expected from this study are state-of-the-art clearing houses, a set of architectures of an IDS clearing house, specifically in a logistic dataspace, and an instantiation of the architecture in the form of a prototype. Then, these architectures will be used to develop a prototype for simulating and demonstrating the functionality of a clearing house in several use cases. Finally, this study will provide general recommendations to adapt the IDS Clearing House architecture in this study to another study case.

1.7 REPORT STRUCTURE

The sections of the report are structured as follows: Chapter 1 discusses the background and motivation of this study. Chapter 2 explores the state of the art of a clearing house through a systematic literature review. Chapter 3 illustrates the design and architecture of an IDS Clearing House in a software requirement and specification, also in ArchiMate diagrams. Chapter 4 discusses the implementation and development of the IDS Clearing House based on the design and architecture described in Chapter 3. In this chapter, the simulation of the prototype toward several scenarios and use cases is also discussed. Chapter 5 discusses the result and validation of the implementation. Finally, the conclusion and future works will be discussed in Chapter 6.

2. STATE OF THE ART OF CLEARING HOUSES

In this chapter, terminologies relevant to clearing houses are discussed in order to understand the context of clearing houses in this study. Then, earlier studies that are relevant to this study are also addressed. In this section, a systematic literature review is conducted to explore the latest development or practice of clearing house implementation in a data exchange environment. The findings from the systematic literature review will be further used to design the architecture of a clearing house.

2.1 EXPLORATORY STUDY

i. Terminology

There are several terminologies relevant to the study that will be referred to and used in this document. In this sub-section, each terminology is explained and described.

ii. Data Sharing

Data sharing is a course of action where an enterprise or organisation enables access to other enterprises or organisations (Pettenpohl et al., 2022b). Data sharing enables beneficiaries for both participants. Those who are sharing their data, may commercialise their data and gain revenue from any access to those data. Then, those who are consuming the data may be able to generate new values from the consumed data.

i. Transaction

Data exchange or data sharing transactions are referred to as a transaction in this study. The transaction enables the sharing of data between participants. The action includes data offering, data requesting and data responding of each participant involved in the data exchange environment.

ii. Clearing

Clearing in this study refers to the data-sharing clearing process. The clearing process ensures the integrity and validity of each participant before a transaction can happen (Bastiaansen et al., 2020). For example, by validating the identity of the participant and whether is authorised to participate according to the contract and agreement made.

iii. Settlement

Settlement in this study refers to the data-sharing settlement process. The settlement process is a set of processes following the clearance of a transaction in order to that a cleared transaction delivered successfully to the participant. The settlement also ensures all data are received in good condition. Furthermore, Settlement also ensures a transaction is discharged whenever it is complete or no longer valid (Bastiaansen et al., 2020).

iv. Trace

Every record of transaction history is referred to as a trace in this study. Trace takes the form of a log or record of information regarding a transaction occurring (Bargh et al., 2014). Later, the trace will be beneficial to enable reporting and auditing purposes over a data-sharing transaction.

v. Audit

An audit or auditing process is an action following the settlement of a transaction. Each participant has their right according to the contract and agreement made prior to the transaction. For example, participants who share their data have a right to have their data accessed according to the contract. On the other hand, participants who consume the shared data have a right to receive the correct and good-quality data as specified in the contract. Therefore, each participant has a chance to audit or review each transaction they made (Bastiaansen et al., 2020).

In case of fraud or violation is discovered through the auditing phase, each participant may file a claim for justification. Justification can be delivered in several ways or forms. For instance by paying a compensation fee for the fraud or violation made.

vi. Sovereignty

Sovereignty in a data-sharing transaction ensures trust and participants' confidence in sharing their data by enabling participants to gain control over the data being shared[6]. Data sovereignty is enabled by enforcing a set of usage policies and rules agreed upon by the participants beforehand.

vii. Trusted Data Sharing

Trusted data sharing enables participants who lack trust in each other due to limited knowledge or experience in sharing data to share their data safely, securely, and reliably. Trusted data sharing is realised by ensuring several aspects such as sovereignty, traceability, auditability, privacy preservation, and other security aspects (Bargh et al., 2014).

viii. International Data Space

International Data Space (IDS) is one of many initiatives in delivering secure and sovereign data exchange and collaboration between enterprises or organisations[7]. IDS Association or IDSA has established a set of reference architecture and frameworks to enable a secure and trusted environment for participants to share their data.

2.1.1 Clearing House

i. Context

Clearing House in the financial domain has been known for years as a neutral trusted third party between participants in a financial and valuable transaction to ensure safe and fault-free transaction (Herbst-Murphy, 2013). In the data-sharing context, IDS specifically, Clearing House is identified as an intermediary component whose main responsibility is to ensure trusted data sharing between participants

[6] https://internationaldataspaces.org/wp-content/uploads/IDSA-Brochure-IDS-Standard-for-Data-Sovereignty-Indispensible-Element-for-Data-Ecosystems.pdf
[7] https://internationaldataspaces.org/

(Pettenpohl et al., 2022b). Trust enabled by clearing house happens through the mediation processes performed which ensure both participants to oblige their contract and argument in exchanging their data (Bastiaansen et al., 2020). Therefore, a clearing house is also considered one of the components to ensure data sovereignty and privacy preservation in data exchange. The class diagram below illustrates the general concept of a clearing house in data exchange derived from the specification of an IDS clearing house (Bastiaansen et al., 2020).

In general, a clearing house in a data exchange must ensure that every transaction is authorised, authenticated, and complete. Authorisation and authentication are supported by the role of a Clearing module or service. Meanwhile, to ensure that a transaction is completed is supported by the Settlement module. Each transaction that passes through these services must be recorded by another component, like a Logger or Logging Mechanism for instance. The clearing house must be able to interact with both Data Provider and Data Consumer.

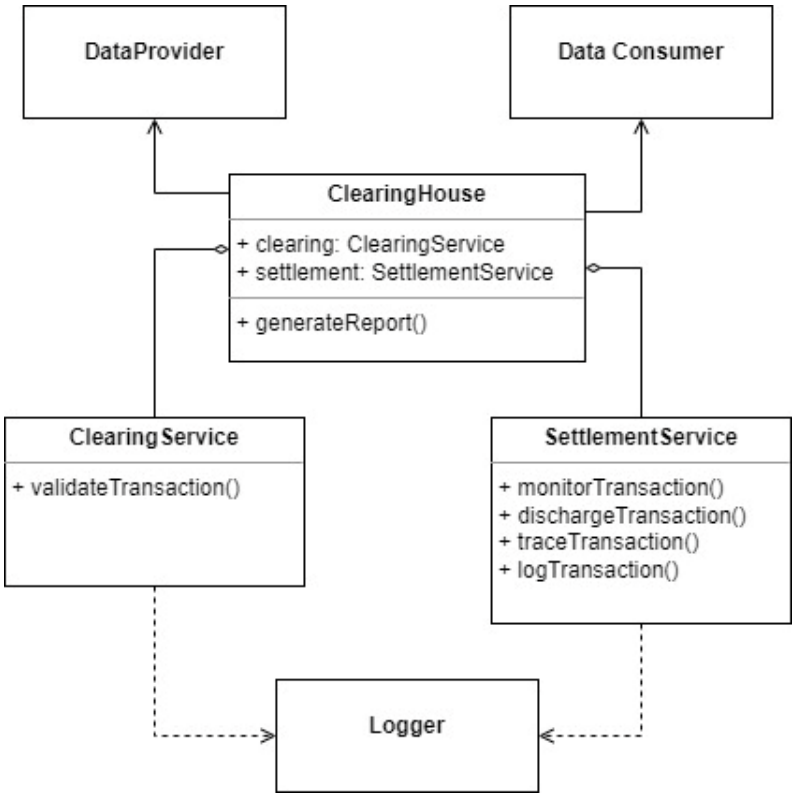


Figure 2-1 Clearing House Concept Class Diagram

ii. Clearing and Settlement Concepts in Data Exchange

The clearing process is initiated prior to the transaction process where the clearing house ensures the integrity of each participant to perform the transaction (Caytas, 2016). Clearing in IDS is based on the enforcement of the usage contract (*3.5.5 Clearing House*, n.d.). In addition, the clearing house also needs to verify whether each participant fulfils the legal and financial requirements to exchange the data. If a participant fails to satisfy these requirements and contracts, the clearing house may cancel and block the transactions. (Bastiaansen et al., 2020).

Then, after a transaction is cleared, the settlement process begins. For example, if Data Consumer payments are completed and valid, the clearing house may settle the transaction. In addition, the

settlement also includes monitoring the data transaction performed which later can be used in resolving conflict between participants concerning contract violations or other unwanted situations during and after the transaction (Bastiaansen et al., 2020). Moreover, IDS clearing house functionalities are not limited to clearing and settlement only but are also extended to support auditability and conflict resolution by providing traceability of a data exchange transaction (Pettenpohl et al., 2022b). Therefore, IDS Clearing House mainly logs all meta-data of the transaction which later these are used to support all functionalities offered above (Bastiaansen et al., 2020).

Based on the context of clearing house described above and as there are little to no literature specifically addressing clearing house in data sharing context, this study focuses on defining a clearing house as an application, middleware, or service that provide several functionalities such as clearing, settlement, and logging to ensure trust, auditability, reconciliation, and transparency in data exchange.

iii. Clearing and Settlement Concepts in the Financial Domain

In this section, the concepts of clearing and settlement of a transaction between issuing and acquiring banks are elaborated. The first concept is clearing and settling the transaction asynchronously, familiarly known as the Dual-Message Transaction system. In this system, clearing and settlement are performed separately from the authorisation process (Herbst-Murphy, 2013). As a result, the Merchant or acquirer only requires authorising the transaction, then later clears and settle all transaction in batches. Normally, clearing in batch happens daily or could be more or less depending on the merchants' number of transactions. The concept results in a faster settlement process, and effective reconciliation scheme and minimises the risk of loses due to fraud (Herbst-Murphy, 2013). Figure 2-2 below illustrates the asynchronous clearing concept.

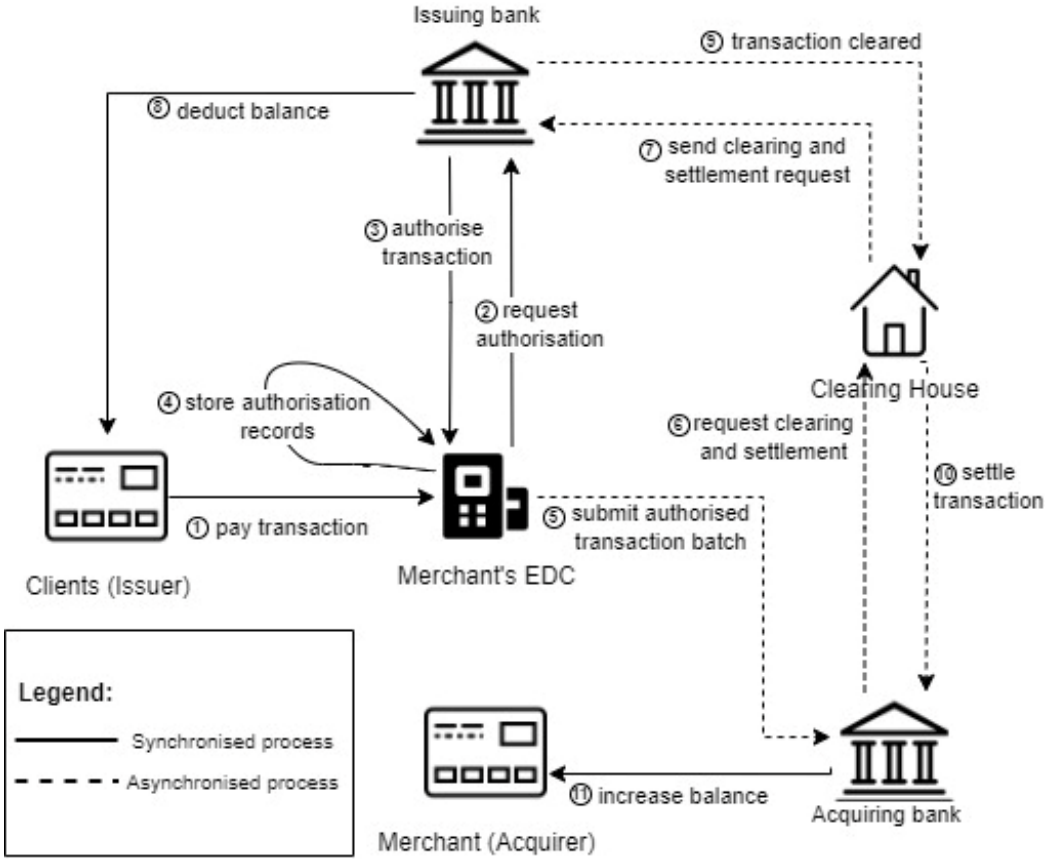


Figure 2-2 Dual-Message Transaction System

In contrast, the synchronous clearing process is known as the Single-Message Clearing System. This concept is not suitable for the merchant which requires fast clearing and settlement processing due to multiple transactions that may occur within a small time window while it is highly applied for individual transactions in Automated Teller Machine (ATM) for example (Herbst-Murphy, 2013). Unlike the asynchronous process, the synchronous process enables authorisation, clearing, and settlement requested and processed immediately at once. Figure 2-3 below depicts the synchronous clearing concept.

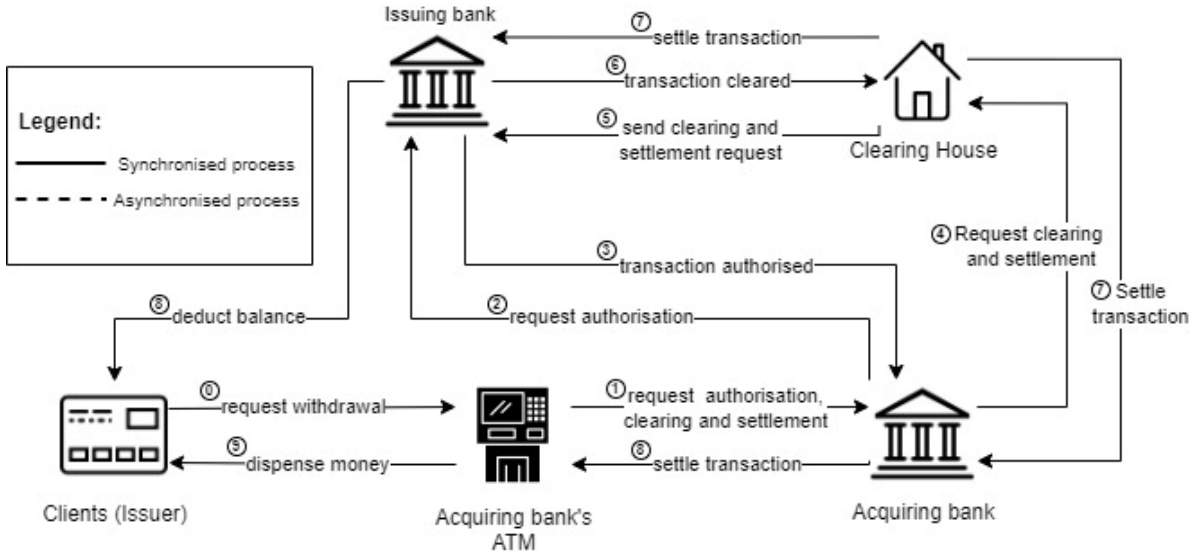


Figure 2-3 Single-Message Transaction System

iv. Data Exchange Clearing and Settlement Concepts Comparison with Financial Domain

Clearing and Settlement in data exchange and financial domain are used interchangeably, while one is derived from the financial domain, both have their characteristics and differences from one another. Referring to the previous section, the clearing and settlement concept in the data exchange domain is responsible for handling data-sharing transactions only. For instance, by ensuring the integrity and reliability of each participant to ensure secure and trusted data sharing. Clearing in data exchange ensures participants in the transaction are verified legally, financially, and technically. Thus, such a clearing function will require a contract, usage policy, financial or payment proof, and other technical aspect to enable the transaction. On the other hand, financial clearing is only responsible for financial transactions or activity. As a result, the data used by financial clearing is different from data sharing clearing. For example, account holder verification and authorisation and the balance amount in that account.

Then, settlement in data exchange ensures the data being shared is shared successfully and received completely by the data consumer. In this case, the data consumer will not be able to deny if they have received the data nor the data producer will not be able to deny if they have sent the data. In this case, settlement in data exchange utilises the contract, usage policy, and description of the transaction occurring. In contrast, financial settlement is the process of executing a transfer or transaction of some fund between accounts. After a transaction is financially cleared, the financial settlement will be able to increase or decrease the participating accounts in the financial transaction.

v. Data Exchange Clearing and Settlement Concepts Comparison with Two-Phase Commit Protocol

The clearing and settlement process often is akin to a two-phase commit protocol due to each main purpose to ensure consistency and reliability in exchanging data. However, both also have differences. The two-phase commit protocol is a protocol to ensure the transaction's atomicity and consistency between distributed multiple nodes by enforcing the decision between parties to commit or abort the transaction (Lechtenbörger, 2009). The two-phase commit starts with the preparation phase which assures each participating node can commit the transaction. If all participants are confirmed able to commit the transaction, then the second phase, the commit phase will start. Otherwise, the transaction will be aborted (Lechtenbörger, 2009; *Two Phase Commit*, n.d.). On the other hand, clearing and settlement guarantee the consistency and reliability of the transaction by ensuring all participants comply with the agreement and policy of the transaction (Pettenpohl et al., 2022b).

2.1.2 Previous Works

IDS Clearing House is one of the crucial components of the IDS (International Data Spaces) Ecosystem. International Data Spaces Association (IDSA) has published at least two documents used in this study as the fundamental information regarding the clearing house. The IDS Reference Architecture Model (RAM) version 4.0 provides a generalised overview of the concepts, functionality, and processes to ensure sovereignty and trust in data sharing. The basic architecture of IDS Clearing House is also provided along with a summary of the functionality such clearing house must offer. However, the information regarding how the clearing and settlement process should occur is limited. Then, Bastiaansen et al. (2020) described the minimum requirements and specifications of an IDS Clearing House in a white paper to complement the IDS RAM. This document also provides more details of the clearing and settlement functionality, along with several case scenarios to illustrate the functionality of the clearing house. Moreover, there is also a relation between IDS Clearing House and layers in IDS RAM which provide more information on how a clearing house should be implemented. However, some information such as the necessity of a clearing house or recommendation of technology and standards architecture was missing.

Besides the two documents above, there is almost no study or literature specifically researching clearing houses for data exchange. Therefore, the studies considered for this research are within the topic of enabling a component or middleware that serves functions and features similar to a clearing house. Bargh et al (2017) present the feedback mechanism to preserve the privacy of the information in accordance with increased trust between participants. There are four concepts or practices to provide feedback between a data provider and data consumer based on the completeness (partial or complete) and the timing (real-time and delayed) of the feedback which depends on the requirements of the transaction. However, there is no clear architecture and technology recommendation provided. In contrast, Reniers et al (2020) conducted a follow-up research on their previous works on ensuring trust in data exchange by introducing smart contracts in their previously researched blockchain technology. The smart contract is used to enable the auditability features of shared data in order to resolve conflict in the future. There are also more approaches within blockchain technology or also other technologies that will be discussed in the next section.

2.2 SYSTEMATIC LITERATURE REVIEW

The Systematic Literature Review (SLR) in this study is intended to use the method introduced by Kitchenham & Charters (2007). The SLR consists of three phases, including Planning, Conducting the Review, and Reporting. Each phase is elaborated in the following table below:

Table 2-1 Kitchenham & Charters Systematic Literature Review Phases

Planning	1. Specify main and sub-research question(s)
	2. Specify scientific database(s)
	3. Formulate research queries based on the research question
	4. Specify inclusion and exclusion criteria
Conducting the reviews	5. Selection i. Gather literature in scientific databases using the defined search query ii. Apply inclusion criteria to the results iii. Eliminate the duplicate results iv. Exclude irrelevant results according to the title and abstract relevancy to the research questions v. Exclude results based on the full-text availability and relevancy to the research questions
	6. Quality Assessment
	7. Data Extraction
	8. Data Synthesis
Reporting	9. Report the synthesis result

The structure of this chapter is also derived from the table above. However, the Data Synthesis and Reporting phase will be carried out in the upcoming chapters.

2.2.1 Planning

The Planning phase is the initial phase of the SLR which is important in order to determine the urgency and objective to perform the SLR. This phase begins with specifying the research questions, and then specifying the scientific databases which then be used in formulating the search queries. Finally, this phase concluded by specifying the inclusion and exclusion criteria.

i. Research Questions

Specifying the research question is a crucial step in shaping the focus and clarifying the objective aimed in the study (Kitchenham & Charters, 2007). As the study aims to understand the implementation and

development of a Clearing House or component that behaves similarly to a Clearing House in data exchange, the research question is specified as follows:

Main Research Question:
What is the state of the art of Clearing House implementation in Data Exchange

Sub-Research Questions:
1. What and how is the importance of Clearing House in data exchange? and what functionalities does it support?
2. What type of transaction in Data Exchange requires the existence of a Clearing House?
3. What effect can such a Clearing House bring to data exchange?
4. What is the preliminary architecture of a Clearing House based on the findings, factors, and alternatives discovered in the literature?

ii. Scientific Databases

Three scientific databases are selected due to their relevancy for software engineering or software technology development. In addition, this may prevent the study from getting broader by selecting kinds of literature out of the specified domain and objectives. The selected databases are defined as follows:

1. Association for Computing Machinery (ACM) Digital Library (https://dl.acm.org/search/advanced)
2. IEEE (https://ieeexplore.ieee.org/)
3. Scopus (https://scopus.com)

While the first two databases are specifically for engineering purposes, the third database was chosen due to its reputation of having the largest collections of literature of several databases that are not included in this study.

iii. Search Query Formulation

The search query is composed of keywords or terminologies that are relevant and related to the main and sub-research questions. There are 4 main groups of queries in this study, each group contains the main query and its relevant synonym to expand the selection. The following table 2 lists all keywords related to the main query.

Table 2-2 Search Query Keywords

Clearing House	Features	Data Exchange	Architecture
Clearing House	Clearance	Data Exchange	Architecture
Broker	Settlement	Data Sharing	Reference
Middleware	Conflict Handling	Data Transaction	Pattern
Intermediary	Conflict Resolution	Data Sharing Environment	Technology
Mediator	Audit	Data Interchange	
Reconciler	Trust	Data Clearing	
	Monitoring	Trusted Data Exchange	
	Integrity		

Based on the keywords above, the search query was then formulated. Each keyword and its synonym will have an OR relationship, while between keywords group will have an AND relationship in the query. The query is also formulated according to each database format, and if applicable, the search will focus on the title, abstract, and keyword.

ACM Digital Library (advanced):
AllField:("clearing house" OR broker OR middleware OR intermediary OR mediator OR reconciler)
AND
AllField:(clearance OR settlement OR "conflict handling" OR "conflict resolution" OR "audit" OR trust OR "monitoring" OR “integrity”)
AND
AllField:("data exchange" OR "data sharing" OR "data transaction" OR "data sharing environment" OR "data interchange" OR "trusted data exchange")
AND
AllField:(architecture OR reference OR pattern OR technology)

IEEE:
(("clearing house" OR broker OR middleware OR intermediary OR mediator OR reconciler)
AND
(clearance OR settlement OR "conflict handling" OR "conflict resolution" OR "audit" OR trust OR "monitoring" OR “integrity”)
AND
("data exchange" OR "data sharing" OR "data transaction" OR "data sharing environment" OR "data interchange" OR "trusted data exchange")
AND
(architecture OR reference OR pattern OR technology))

Scopus (advanced):
TITLE-ABS-KEY (("clearing house" OR broker OR middleware OR intermediary OR mediator OR reconciler)
AND
(clearance OR settlement OR "conflict handling" OR "conflict resolution" OR "audit" OR trust OR "monitoring" OR “integrity”)
AND
("data exchange" OR "data sharing” OR "data transaction" OR "data sharing environment" OR "data interchange" OR "trusted data exchange")
AND
(architecture OR reference OR pattern OR technology))

iv. Inclusion and Exclusion Criteria

In order to reduce the likelihood of bias Kitchenham & Charters (2007) recommends a review protocol, one of which is by defining the selection criteria. The selection criteria, further defined as inclusion and exclusion criteria is a set of protocols to refine the search results. Table 3 below describes the inclusion and exclusion criteria defined in this study.

Table 2-3 Inclusion and Exclusion Criteria

Inclusion	Exclusion
Studies written in English and peer-reviewed	Duplicated collections based on the title and content
Studies published in conferences and research articles	Irrelevant studies based on the title, abstract, and keyword relevancy toward the main and sub-research questions defined
Studies focused on subject areas of Computer Science, Information Technology, Information Science, Communication Technology, Engineering	Full-text unavailability

Based on the table above, a study or article will be considered if written in English due to internationalisation reasons (e.g., reviews are done internationally, from peers with various backgrounds and nationalities). In addition to the peer-reviewed criteria, the collected studies are also required to be published as a proceeding from a conference or a research article from a specific journal in order to maintain the quality of this study. Then, focusing the collection for studies in specific subjects (e.g., computer science, information technology) should prevent broadening the collection that may be irrelevant to the topic of this study. Finally, there is no restriction in regard to the publication years of the study.

However, there is still a chance where irrelevant studies are gathered in the selection, thus the exclusion criteria apply. This study will eliminate studies that are duplicated based on the title and the content. Later, studies that are irrelevant to the aims of this study will be eliminated as well to streamline the collection. There is also a likelihood that a study may not be freely accessed by the public, is incomplete (e.g., only containing the abstract), or is too short, therefore these studies will not be considered as well.

2.2.2 Conducting The Review

In this chapter, the process of article selection and extracting data from the selected articles are described.

i. Selection

The selection process begins with executing the search query in the chosen databases. As there may be irrelevant studies presented in the result and collection, applying the inclusion and exclusion criteria as discussed above is necessary. Not limited to ensuring the quality of the research and streamlining the collection, this also streamlined the extraction process later where only relevant, or the most relevant studies are included and considered. Therefore, after retrieving results from the database through exporting and then importing the metadata to the Zotero, the reference manager utilised in this study, the inclusion criteria should be applied. Then, every duplication of the remaining collection will be removed. Next, the remaining articles will be assessed based on each title, abstract, and keywords' relevance to this study's goals and objectives. Then, the articles' files or text will be retrieved, if the article is incomplete or not available in full text, the article will not be considered for a full read and assessment based on the content. Finally, 23 articles are left in the collection and will be considered in this study. The following table 4 captures the selection process.

Table 2-4 Literature Selection Process

Selection Phase	Numbers of Articles Left	
Gather literature in scientific databases using the defined search query	Scientific Databases: ACM: 240 IEEE: 84 Scopus: 109	Combined: 433
Apply inclusion criteria to the results	Scientific Databases: ACM: 171 IEEE: 80 Scopus: 78	Combined: 329
Eliminate the duplicate results	273	
Exclude irrelevant results according to the title and abstract relevancy to the research questions	114	
Exclude results based on the full-text availability	106	
Exclude results based on the relevance to the research questions	23	

ii. Study Quality Assessment

Kitchenham & Charters (2007) emphasises the importance of ensuring the quality of selected studies in addition to the inclusion and exclusion criterion defined in order to minimise the bias in the study while increasing the internal and external validity of the study. Table 5 below describes the quality assessment of the study which includes the goal or purpose, the method, and the outcome of the selected studies. Kitchenham and Charters defined four categories of the method of the study, namely, Quantitative Empirical Studies (QE), Correlation or Observational Studies (O), Surveys (S), and Experiments (E).

Each study has distinguished goals and purposes and so does the outcome of the study. Most of the selected literature produces a theoretical overview or explanation of Clearing House concepts or applications, these literatures are marked with T. Several studies also present a conceptual model (CM), detailed architecture (A), or implementation (IA) which in line to this study research question in understanding the architecture and technology used in Clearing Houses development.

Table 2-5 Study Quality Assessment

No.	Reference	Research Purpose	Method				Outcome			
			QE	O	S	E	T	CM	A	IA
P1	(Pincheira et al., 2020)	"This paper proposes an architecture based on the blockchain to build a network to share and validate data acquired by untrusted sources"	✔			✔	✔		✔	✔
P2	(Sober et al., 2022)	"Blockchains provide properties such as public verifiability, transparency, integrity, and redundancy … we investigate if the utilization of (P2P) and blockchain technology can provide	✔				✔	✔		✔

No.	Reference	Research Purpose	Method				Outcome			
			QE	O	S	E	T	CM	A	IA
		the necessary technological functionalities for IoT data marketplaces"								
P3	(Huang et al., 2020)	"We design a secure data sharing scheme among several entities who interact with smart contracts, … to achieve the privacy preserving of medical data"	✔				✔	✔		✔
P4	(Belhi et al., 2022)	"We investigate the integration of the blockchain technology through Hyperledger Fabric Smart Contracts with industrial applications … that the tracking of a transaction is more robust and also all transactions are final"	✔					✔		✔
P5	(Si et al., 2022)	"We propose a secure CDMD sharing method in a decentralized way based on blockchain and cryptography technology to realize secure data sharing among entities … Hence, it can make distributed storage, data with traceability and immutability"	✔					✔		
P6	(Huang et al., 2017)	"This paper utilizes the feature of data is not tampered and completely transparent, combines with the time stamp and the transaction details in the process of storage and trading, so that it can be trusted by many parties."	✔					✔	✔	✔
P7	(Rodriguez-Garcia et al., 2021)	"The contributions of the design described in this paper are the following: (1) providing an auditable mechanism for patients to give consent on the use of their data,"	✔					✔		✔
P8	(Reniers et al., 2019)	we systematically investigate several architectural approaches to implement auditable blockchain-based private data sharing	✔						✔	
P9	(Reniers et al., 2020)	"We motivate our proposed solution for auditable data sharing in the context of an industrial case"		✔			✔	✔		✔
P10	(A. Khan & Anjum, 2022)	"We have proposed a blockchain-based architecture of a healthcare ecosystem for accountable medical data sharing … The core of our proposed architecture	✔				✔		✔	

No.	Reference	Research Purpose	Method				Outcome			
			QE	O	S	E	T	CM	A	IA
		consists of a blockchain-based shared infrastructure comprising smart contracts and distributed ledger to log users record, devices record, permissions record and access log"								
P11	(Anderson & Edwards, 2010)	"We describe the need for technically leveraged policy models and governance strategies to support data sharing between a range of disparate stakeholders where trust is not easily established or maintained."	✔				✔			
P12	(Pasquier et al., 2017)	"In this paper we present CamFlow (Cambridge Flow Control Architecture) … Log records can be made efficiently of all attempted flows, whether permitted or rejected, and this log provides a possible basis for audit, data provenance and compliance checking. By this means it can be checked whether application-level policy has been enforced and whether cloud service provision has complied with contractual obligations"	✔				✔		✔	✔
P13	(Kohli & Suarez, 2016)	"The paper emphasizes on securing the payment data generated from one or more than one ERP applications and formalizing treasury approvals guidelines and constructing secured infrastructure on best practices."	✔				✔		✔	
P14	(Xia et al., 2022)	"We introduce Data Station, an intermediary data escrow, the computational and data management infrastructure designed to enable the formation of data-sharing consortia … all computation that takes place on the platform must be transparent so third-party auditors and compliance officers can audit the consortia."	✔				✔		✔	
P15	(Esteves et al., 2022)	"This is the result of the PROTECT project, in particular of its Work	✔				✔	✔		

No.	Reference	Research Purpose	Method				Outcome			
			QE	O	S	E	T	CM	A	IA
		Package 1.4 As such, this contribution seeks to, relying on existing legal and ethical requirements, propose an ODRL profile capable of facilitating interactions between different entities upon transparent information on a certain data processing activity."								
P16	(Serrano et al., 2014)	"This paper addresses this important issue by presenting the main problems affecting cyber security information sharing and proposing some solutions that would enable the development of an information sharing system that would meet the cyber security domain's specific requirements."	✔				✔	✔		
P17	(Bargh et al., 2014)	"We consider privacy protection in data sharing settings … we study various feedback mechanisms that can in a way be exploited to protect privacy and to prevent establishment of inadequate relationships."	✔				✔	✔	✔	✔
P18	(Oh et al., 2014)	"The goal of this paper is to explore the design space for privacy-preserving audit with automation supported by formal logical representation of audit policies."	✔				✔		✔	✔
P19	(Wang et al., 2018)	"In this paper, we propose a systematic compliance monitoring framework for regulatory supervision in supply chains by analysing operational processes focusing on import process and manufacturing activities"	✔				✔		✔	✔
P20	(Choi et al., 2013)	"We present a security framework for citizen data sharing for Government services (Secure-Gov) that enables secure sharing of citizens' documents and data among government agencies in delivering citizen services"	✔						✔	
P21	(Chen et al., 2021)	"In this paper, we resolve the data and value exchange without TTP. Based on data exchange contract, we mainly protect the privacy and	✔					✔		

No.	Reference	Research Purpose	Method				Outcome			
			QE	O	S	E	T	CM	A	IA
		integrity of privacy using TEE, and design algorithms to verify data and allocate assets once dispute."								
P22	(Bajoudah et al., 2019)	"This work follows on from our earlier proposal for a IoT data marketplace, … The approach suggested is based on the idea that each participant would periodically report to a Smart Contract on the data sent to and received from other participants, and the Contract would then be able to use such reports to settle any disputes."	✔				✔		✔	✔
P23	(Luo et al., 2022)	"This paper puts forward a three-layer technology architecture for blockchain-based secure data sharing by surveying the approaches of data storage, data sharing and privacy protection, with a view to providing useful inspiration and thoughts for future research."	✔			✔	✔		✔	✔

iii. Data Extraction

In this chapter, all data and information extracted from the selected studies are recorded in Table 6 below. The extracted data later be used to answer and address the research question of this study (Kitchenham & Charters, 2007) as described in Chapter 3.1.1 above. Based on the research question and objective of this study, the data extracted from the literature are categorised into four categories: Transaction Type (TT), Motivation and Objective (MO), Strategy and Process (SP), and Architecture and Technology (AT). Transaction Type as well as Motivation and Objective will be used to distinguish the urgency of Clearing House applications based on the type of data being transacted in data exchange and the reason why a Clearing House is needed in the first place. In addition, each type and motivation of the transaction should have a distinguished Strategy and Process (SP) in order to achieve the goal of applying Clearing House. Finally, each application's architecture and main technology building block are recorded in the AP column.

Table 2-6 Data Extraction

No	Reference	TT	MO	SP	AT
P1	(Pincheira et al., 2020)	IoT devices measurement on a crop field Photos captured by	Data acquisition from an untrusted source Ensure secure, immutable, and transparent	Metadata collection, validation, and tracking	Blockchain Smart contracts

No	Reference	TT	MO	SP	AT
		personal drone or cell phone	distributed infrastructure		
P2	(Sober et al., 2022)	IoT devices measurement	Valuable IoT data mostly shared in centralised approaches Data marketplace with public verifiability, transparency, integrity, and redundancy properties	Data encryption with private key exchanged through Broker Settlement process before the transaction Rating system after exchange occurs	Blockchain Smart contracts
P3	(Huang et al., 2020)	Electronic Health Records (EHR)	Privacy preservation in sharing medical data	Storing participants' behaviour in blockchain	Blockchain Smart contract
P4	(Belhi et al., 2022)	Manufacturing ERP	The complexity of transactions and collaboration between actors in Supply Chain Management	Data integrity checking through a synchronisation process	API Integration Broker
P5	(Si et al., 2022)	Chronic Disease Medication Data (CDMD)	Data centralisation privacy issue Privacy preservation	Data hashing Record participants' data access behaviour	Blockchain
P6	(Huang et al., 2017)	IoT device measurements	The centralised approach is unable to provide transparency, auditability, and immutability of the data exchange process	Permission management through access contracts Record exchange history	Blockchain Smart contract
P7	(Rodriguez-Garcia et al., 2021)	Electronic Medical Records (EMR)	Auditable mechanism to address individual privacy concerns	Records and execute agreements during data request, offer, and payment process with smart contracts	Blockchain Smart contract

No	Reference	TT	MO	SP	AT
P8	(Reniers et al., 2019)	Multi-disciplinary optimisation (MDO) Airplane component design	Safety-critical and confidential data exchange	Logging data access or alter Data validation with homomorphic hashing	Blockchain
P9	(Reniers et al., 2020)	Multi-disciplinary optimisation (MDO)		Logging mechanism A request-response scheme in exchanging symmetric file key	Blockchain Smart contract
P10	(A. Khan & Anjum, 2022)	Electronic Health Records (EHR)	Transparency, traceability, auditability, trust, privacy, and security concerns Centralised approach prone to a single point of failure	Data access logging	Blockchain Smart contract
P11	(Anderson & Edwards, 2010)	Electronic Health Records (EHR)	Exploring strategies to enable trust between stakeholders	Inter-investigator pattern Inter-institutional pattern Mandated federal data-sharing pattern	Feedback loops
P12	(Pasquier et al., 2017)	Confidential data	Auditable Information Flow Control (IFC)	Defining tag and label for each data item Data-centric logging	IFC model with taint tracking Linux Security Module
P13	(Kohli & Suarez, 2016)	Payment data ERP data	Manual processing of securing payment data between ERP	Processing history Audis trails of payment handling	Encryption SSH Transfer Protocol

No	Reference	TT	MO	SP	AT
				Synchronisation with the banking portal	
P14	(Xia et al., 2022)	Beneficial data Confidential data	Difficulties in controlling beneficial data	Storing encrypted plain text audit log	"Gatekeeper" broker
P15	(Esteves et al., 2022)	Confidential data	Developing and building trust to let service providers share the private data	Measure data uses of the users	Paradigm Open Digital Rights Language (ODRL)
P16	(Serrano et al., 2014)	Confidential data Organisational Data	Organisations hinder from sharing their data due to cyber security risks	Logging	Controlled Multilateral Sharing Information Exchange Policy
P17	(Bargh et al., 2014)	Confidential data Citizen data eGovernment	Data misuse violates the privacy of individual data The available technical solutions did not prevent privacy violations Put data dissemination into practice	Reporting all data processing actions by the data processor to the data controller. Real-time and delayed feedback to the data controller Complete and partial feedback to the data controller	Feedback mechanism
P18	(Oh et al., 2014)	Electronic Health Records (EHR)	Patients' concerns to share health data The benefits of a broker system to having an indexed and secured data sharing Opportunity to ensure auditability of the information exchange with the broker	Data logging Create audit decisions based on the log processed by the algorithm	Broker Audit Data Processor Audit Agent Audit Viewer Audit Algorithm

No	Reference	TT	MO	SP	AT
P19	(Wang et al., 2018)	International supply chain International logistic Customs	High integrity and confidential data between businesses shared with government or regulatory authorities Conflict resolution between businesses and regulatory authorities	Define compliance agreement Trace data movement toward the compliance agreement	Compliance monitoring framework
P20	(Choi et al., 2013)	Citizen data eGovernment	Prevent confidential citizen data leakage during information sharing	Enforcing usage control Embedding digital footprint to shared information	Security Framework
P21	(Chen et al., 2021)	Confidential data	Privacy concerns in sharing confidential and valuable data Opportunity to gain more value from sharing data	Verifying data correctness and validity by the *Judge* component. The judge matches the encrypted data with formal rules defined	Blockchain Smart contract
P22	(Bajoudah et al., 2019)	IoT device measurements	Participants' reputation in the data marketplace Possibility of loses due to transactions with unreliable participant	Periodic checkpoints during data exchange (data offer, trade, and receipt)	Blockchain Smart contract
P23	(Luo et al., 2022)	Valuable data	Data security issue hinders participant to share their data	Notary Mechanism, specific node to verify and audit the occurring events	Blockchain

2.3 SYSTEMATIC LITERATUR REVIEW RESULTS

2.3.1 Importance and Functionalities of a Clearing House

In this section, the importance of a clearing house and the functionalities required by a clearing house will be discussed. Understanding the importance will be necessary to determine whether a trust component like a clearing house is needed in data exchange. Table 7 below describes the importance factor along with supporting functionalities offered in the previous studies. There are several importance factors to enable trust in data exchange, such as Auditability, Authorisation, Privacy Preservation, and Traceability were derived from observing the literature and the concepts elaborated in Section 2 above.

Table 2-7 Clearing House Importance and Functionalities

No	Importance	Functionalities
IM1	Auditability [P1, P3, P6, P7, P8, P10, P12, P13, P14, P18, P19, P23]	- Log data transaction/participants activity [P1, P3, P6, P7, P8, P10, P12, P13, P14, P18, P19, P23]
IM2	Authorisation [P2, P3, P6, P9, P14, P20]	- Encrypt data [P2, P3, P9, P14] - Clearing and Settlement [P2, P6, P20] - Post-transaction Rating System [P2, P22]
IM3	Privacy Preservation [P1, P2, P3, P4, P5, P7, P8, P10, P12, P15, P17, P18, P20, P21, P23]	- Collect and validate metadata [P1, P12, P20] - Collect and validate data [P4, P8, P17, P18] - Encrypt data [P2, P3, P5, P8, P20, P21]
IM4	Traceability [P1, P3, P5, P6, P9, P10, P13, P16, P17, P22]	- Log data transaction/participants activity [P1, P3, P6, P8, P9, P10, P13, P15, P16, P17, P18, P22, P23]

Based on the table above, most of the studies required the existence of a trust component to preserve privacy in data exchange due to the possibility of data leakage due to participants violating the agreement or policy agreed upon. The effect of such data leakage is crucial for individual or organisational reputation, revenue, or another impact that may generate losses or damage(Cheng et al., 2017). Thus, IM3 addresses that preserving privacy through technical or semantic techniques is necessary (Bargh et al., 2014). Technical preservation involves data encryption, which is the most used technique found in the literature. Data encryption ensures only authorised participants have access to the data, by decrypting the data with the same key used to encrypt the data. As a result, the clearing house will not be able to perform an analysis and process the data unless the participants are involved in the decryption process in the clearing house. However, this will result in a spike in data transmission between the connector and the clearing house (Bastiaansen et al., 2020). On the other hand, preserving privacy also can be done semantically by collecting and validating the data or the metadata only. However, collecting and validating the whole data is not recommended due to privacy or sensitive information that data may contain (Bastiaansen et al., 2020).

Recording the data exchange through logging, for example, is the most recommended functionality such a clearing house should offer. Logging enables clearing houses to provide traceability (IM4) and auditability (IM1) of the data being exchanged. Traceability ensures transparency between participants by recording all or specified participants' behaviour during exchanging the data in order to minimise the risk of participants violating the data exchange agreement (Huang et al., 2020). The violations such as data producer sending bad quality data until the data consumer disregard the usage policy of the data will

derive conflict between parties (Reniers et al., 2020). Therefore, the trace or history of events recorded by clearing house further can be utilised as a tool to perform an audit or resolve a conflict over data exchange performed (Wang et al., 2018).

Finally, Importance IM2 highlights the clearing house could offer an authorisation service to ensure only authorised participants are allowed to participate in the data exchange. While data encryption is one of the favoured methods, there are also several interesting ways to let such a clearing house authorise the transaction. For example, the clearing and settlement process found also supports the authorisation process. The clearing and settlement process ensures only verified and authorised participants (Sober et al., 2022), such as through completing payment obligations or each identity successfully verified are allowed to access and exchange the data (Bastiaansen et al., 2020; Huang et al., 2017). Clearing house may also provide participants with other participants' ratings based on each previous transaction (Sober et al., 2022), therefore participants may opt whether to participate or not in the data exchange if other participants are having a good or bad rating (Bajoudah et al., 2019).

2.3.2 The Type of Transaction Requires a Clearing House

This section discusses the type and category of transaction that requires a trust-intermediary such clearing house. There are a number of transaction types found in the literature, however, there are also some types that are similar to each other. The table below describes the category of transaction also the type of transaction that belongs to the category. There are four categories based on the observation from types found in the literature, namely, Personal Data which mainly concerns individual data, financial data, organisational data, and others.

Table 2-8 Transaction Types Require Clearing House

No	Category	Type of Transaction
TT1	Personal Data [P1, P3, P5, P6, P10, P11]	- Health records [P3, P5, P7, P10, P11, P18] - Citizen data [P17, P20] - Personal assets [P1, P22]
TT2	Financial Data [P4, P13, P14]	- Payment data [P13] - Valuable data [P4, P13, P14, P23]
TT3	Organisation Data [P2, P3, P4, P5, P8, P9, P11, P12, P13, P14, P15, P16, P19, P21]	- Manufacturing data [P4, P8, P9] - Logistic data [P19] - Confidential data [P3, P4, P5, P8, P9, P11, P12, P14, P15, P16, P19, P21] - Organisational assets [P2, P4, P5, P8, P9, P11, P13, P15, P16, P19]
TT4	Others [P1, P2, P6]	- IoT device measurement [P1, P2, P6, P22]

The table above indicates most of the studies are utilising clearing houses for exchanging organisational data. Organisational data (TT3) is usually containing sensitive and valuable information such as manufacturing, logistics, confidential, and organisation's assets data based on each purpose. Moreover, financial data such as payment or trade and other valuable data is also sensitive. For example, aircraft manufacturing data is highly sensitive as it is containing designs or procedures to ensure safety (Reniers et al., 2019). Moreover, organisational data may contain private information such as employee identity, customer data, new product design, transaction records, or also assets in the company (Belhi et al., 2022). Therefore, a clearing house is needed to ensure only authorised participants may participate in the data exchange to minimise the risk of this sensitive information being leaked. On the other hand, the data

consumer also needed to be ensured the data being shared are of high quality in line with the cost on which they have to pay (Kohli & Suarez, 2016).

Similarly, Transaction type TT2 argues that personal data are also sensitive and private. In this digital era, personal data are no longer stored in paper but as digital data. For example, the rise of e-government enable citizen to get governmental service easily and quickly as their data is already captured in the government system (Choi et al., 2013). There is also health records of an individual that can be used by hospitals or health institution to increase their service by utilising patients' data to understand which treatment is effective for a specific health issue (Anderson & Edwards, 2010). On the other hand, personal data that are not handled by the government, hospital, or other institutions is also valuable and sensitive. Personal assets, for example, can be digital media or also can be a record of a personal device like a mobile phone. These assets may be exchanged to generate a new value by selling or letting a corporation or institution use or lease the data (Pincheira et al., 2020). However, it is then required to have such components to ensure the privacy of these personal data is preserved during data exchange.

Then, the transaction type TT3 financial data are grouped into two kinds, such as payment data and valuable data. Both types of data typically contain sensitive and private information (Kohli & Suarez, 2016). For example, personal payment may contain the account number, personal lifestyle, or private description of an item or anything being paid. For organisations, payment data may contain the organisation's partner account numbers, information about the organisation's activities, or other sensitive information (Belhi et al., 2022; Xia et al., 2022). Finally, there are also IoT device measurements. Currently, IoT devices are widely used both personally and organisationally to support their activity. These data may not be consumed or spread to an unwanted individual or organisation due to the value of the measurement captured by the device and several may contain private and sensitive information (Bajoudah et al., 2019; Huang et al., 2017; Sober et al., 2022).

2.3.3 Effect of Clearing House on Data Exchange

Introducing or implementing a specific component may derive several effects on functionality, performance, cost, or also security in a specific ecosystem. In the table below, the effect of implementing a trust component such clearing house to ensure trust in data exchange is elaborated.

Table 2-9 Clearing House Effect in Data Exchange

No	Effects
EF1	Ensure participants trust [P1, P2, P3, P4, P6, P9, P10, P11, P14, P15, P17, P18, P20, P22]
EF2	Increase data exchange effectivity [P1, P3, P5, P7, P12, P16, P19]
EF3	Increase development and operational cost [P2, P9, P10, P23]
EF4	Increase privacy and security of data exchange [P3, P4, P6, P7, P8, P10, P11, P12, P13, P14, P15, P16, P17, P18, P20, P21, P22, P23]
EF5	Maintain development and operational cost [P3, P12]
EF6	Increase data value [P3, P5]
EF7	Ensure conflict resolution [P1, P3, P4, P5, P7, P8, P9, P11, P16, P17, P18, P19, P20, P22, P23]

As discussed in the previous sections, many individuals, organisations or institutions concerns the privacy and security which may discourage them to participate in exchanging their data. A clearing House or intermediary similar to a clearing house has shown an impact in increasing the privacy and security of the data being exchanged (EF4). Each initiative presents various solutions to ensure the privacy and security

of the data exchange. Also, to the question on ensuring participants share their data privately and securely in the data exchange ecosystem. For example, using blockchain technology or encrypting the data which is shown to solve participants' concerns regarding their privacy and security in exchanging their data (Luo et al., 2022).

Security of the data exchange is not only limited to the process of sharing the data but also to the process after the data is exchanged. For example, the possibility of the data provider does not provide data with good quality and complete or the data consumer does not follow the data usage regulations agreed upon in the contract are exist in data exchange (Bastiaansen et al., 2020). As a result, this reason also might hinder both participants from fully participating in the exchange ecosystem (Oh et al., 2014). Therefore, most studies agreed that a trust corporation such clearing house is needed to provide the effect EF7 which clearing house enables conflict resolution preceding the data exchange. Clearing House is provided with some functionalities to help the audit process in the future, by recording participants' behaviour for instance (Bargh et al., 2014).

Overall, ensuring these measures may also help participants' trust in participating to exchange their data (EF1). As a result, if privacy and security concerns are considered by the clearing house and participants' trust is ensured then the participants' participation will increase and may bring new value to each of them (Pincheira et al., 2020). While a clearing house may increase these measures and the effectiveness of data exchange (EF2), it is important to understand and select the correct building blocks and technology to develop and deploy a clearing house in order to maintain the development and operational cost of a clearing house (EF5). For example, choosing an unsuitable technology may introduce the possibility of clearing houses increasing the development and operational cost (EF3). For example, the use of Ethereum while it is ensuring secure transactions which leads to ensuring trust between the participant, is very expensive to implement and develop (Luo et al., 2022; Sober et al., 2022).

2.3.4 Architecture Components of a Clearing House

In this section, the architectural components of a clearing house will be discussed. There are four major components of a clearing house, including data storage, data access management, data governance, and security then finally data reporting. The table below presented each initiative component used to develop a clearing house.

Table 2-10 Clearing House Architecture Component

No	Architecture Component	Description
AC1	Data Storage	- Metadata [P1, P2, P14, P17, P19] - Decentralised Storage [P1, P2, P7, P8, P9, P10, P11, P12, P13, P15, P16, P17, P22] - Hash [P3, P5, P6, P7, P9, P12, P18, P20, P21, P23]
AC2	Data Access and Security	- Smart contract [P1, P2, P3, P4, P6, P9, P10, P21, P22, P23] - Usage control policy [P5, P7, P8, P11, P12, P13, P14, P15, P16, P17, P18, P19, P20] - Data Marketplace [P2]
AC3	Data Provenance	- Smart contract [P1, P2, P3, P4, P6, P9, P10, P21, P22, P23]

		- Usage control policy [P5, P7, P8, P11, P12, P13, P14, P15, P16, P17, P18, P19, P20]
AC4	Data Reporting	- Logger [P1, P3, P4, P5, P7, P8, P10, P11, P12, P13, P14, P15, P16, P17, P18, P19, P20, P22] - Data Marketplace [P2] - Smart contract [P6, P9, P21, P22, P23]

Introducing an intermediary between participants in a data exchange brings a question of how the intermediary should store and access the data (AC1). Most studies emphasise the data should only be kept and stored in its origin or the data provider due to privacy and security reasons. Thus, only the transaction record (Esteves et al., 2022) or access history is recorded (A. Khan & Anjum, 2022). As a result, traffic between the clearing house and the data storing place will be increased in case of the clearing house needed to find or report the data for audit purposes (Bastiaansen et al., 2020). While storing whole data is not recommended as it is introducing a new security gap, several studies also suggest storing the hash or the encrypted data in the clearing house. Otherwise, In line with the specification of IDS Clearing House, it is also possible to let the clearing house only store the metadata.

Next, the data access and security component is mainly responsible for handling the authorisation and authentication in the clearing house (AC2). It is necessary to ensure only authorised participant has access to the data. For example, by utilising a usage control policy that contains information over what are the allowed or permitted actions of specific participants in data exchange (Xia et al., 2022). On the other hand, several building blocks found in the literature are based on blockchain technology. Therefore, it does not surprising that smart contract is favoured to provide data access and management service toward clearing house. A smart contract may be beneficial to ensure that only authorised participants are allowed to access the data (Reniers et al., 2020). Moreover, the smart contract also offers the possibility of a settlement process (Huang et al., 2017). Therefore, the smart contract is also supporting the data provenance of a clearing house (AC3).

Finally, the clearing house is required to provide a report concerning conflict resolution, claim handling, or other auditing purposes (AC4). As described in the table, the recording of participants' behaviour during data exchange through logging is favoured by most studies. The terms of logging here are used to differentiate with logging in smart contracts (Bargh et al., 2014). While both options are discovered in most studies. However, smart contract logging may increase development and operational cost due to the technology used (Luo et al., 2022; Sober et al., 2022). In contrast, non-smart contract logging is highly customisable to suit the need of the clearing house. Moreover, the possibility of modern logging provides more insight and measures to increase the functionality of the logging mechanism.

3. DESIGN

In this chapter, the findings from the systematic literature review are used to determine the requirements and specifications on a clearing house. Based on these requirements, a set of architectures is also defined. The defined architectures include the standard viewpoints of the clearing house and the viewpoint from the logistic data space.

3.1 SOFTWARE REQUIREMENTS AND SPECIFICATIONS ON A CLEARING HOUSE

In this section, software requirements derived and gathered from the key elements of the systematic literature review result discussed above are elaborated. As discussed earlier, there is currently limited information on clearing house development in data exchange. Therefore, the requirements presented below will offer a clear description and view regarding the system functionality and constraints on designing and developing a clearing house (Sommerville, 2011). The requirements below are priorities using the MoSCoW method. There are four levels of prioritisation: Must, Should, Could, and Won't. Requirement with Must prioritisation is necessary to be included in the software, and the prioritisation is gradually decreased until the Won't requirement (Hudaib et al., 2018).

3.1.1 Functional Requirements

Must:

FR 1. The clearing house must be able to provide clearing and settlement services.
FR 2. The clearing house must be able to discharge a transaction if the defined contract and rules are no longer valid.
FR 3. The clearing house must only clear transactions that are verified financially, legally, and technically according to the agreed rules and contract by each participant.
FR 4. The clearing house must support the enforcement of the agreed contract and rules.
FR 5. The clearing house must store only the metadata of the data being exchanged.
FR 6. The clearing house must record every behaviour and activity of each participant.
FR 7. Trace and recorded participants' behaviours must be stored in a secure way.

Should:

FR 8. The clearing house should be able to provide a report on a specific time basis.
FR 9. The clearing house should be able to provide metadata and recorded participants' behaviour upon request in case of conflict resolution purposes.

3.1.2 Non-Functional Requirements

Should:

NFR 1. The clearing house should be highly available and reliable.
NFR 2. The clearing house should be able to handle high traffic volume in real-time.
NFR 3. The report should be generated within a specific time window.

3.2 ARCHITECTURE OF A CLEARING HOUSE

In the previous section, the functional and non-functional requirements on a Clearing House are discussed. In order to reflect and illustrate how the clearing should be designed and developed according to the requirements defined, in this subsection, the requirements are translated into an Enterprise Architecture viewpoints in ArchiMate language. There are four viewpoints that will be discussed in this subsection, such as application behaviour viewpoints, application collaboration viewpoints, service realisation viewpoints, and technology usage viewpoints. Each defined viewpoint is specifically directed to anyone in software engineering or software development domains.

3.2.1 Motivation Viewpoint

The ArchiMate motivation viewpoint highlights the context and insight of the motivations, drivers, assessments, and goals in an enterprise architecture. Figure 3-1 below depicts the motivation of this study in designing and implementing a trusted intermediary called IDS Clearing House. Based on the problem statement discussed earlier, this study aims to address the concerns of Logistics Enterprise in participating in data sharing, as well as the concerns of the International Data Space Association in data space research and development. Firstly, the logistics enterprise is concerned with the lack of trust which later may introduce losses to them through experiencing fraud or violation in sharing their data. Therefore, the proposed architecture and its instance should be able to ensure that only validated and authorised participants are allowed to participate in the data sharing. In addition, features such as traceability and auditability must be enabled as well in the outcome of this study.

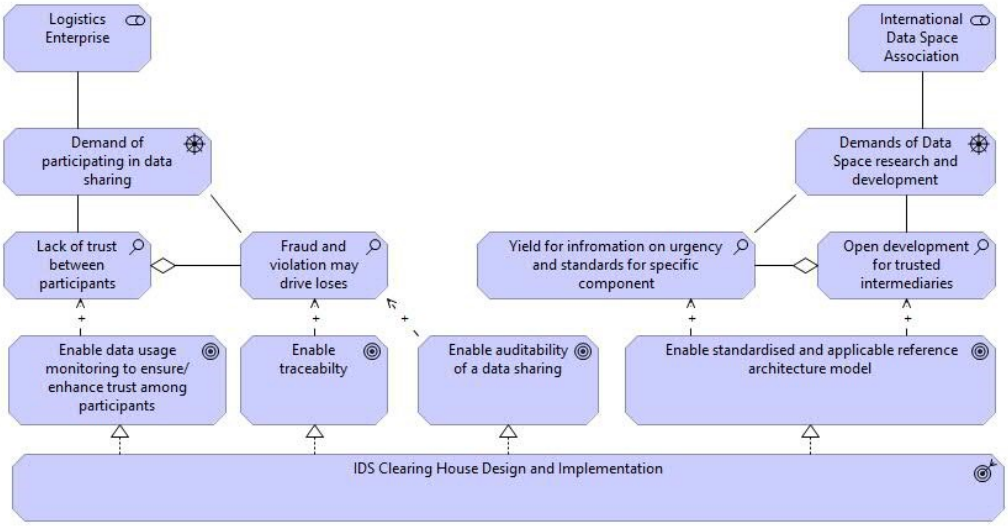


Figure 3-1 Motivation Viewpoint for Logistic Enterprises to Adopt Clearing Houses

Secondly, for the International Data Space Association in which also includes system architects, software engineers and developers, users, also other parties which highly involved in data space research and development, there is ongoing research and development for trust intermediary components such as a clearing house. However, several crucial information such as standards and urgency of such components have not yet been thoroughly researched. As a result, this study should be able to ensure that the

outcomes are able to act as a reference architecture for these parties to develop a clearing house, especially in other business cases and requirements.

3.2.2 Application Behaviour Viewpoint

According to the functional requirement FR1 in Section 3.1, a clearing house application in a data transaction must offer at least four services, such as Clearing, Settlement, Logging, and Claim Handling services. These services are aligned with the clearing house's architecture in IDS Reference Architecture Model 4.0. As shown in Figure 3-2 below, an IDS Clearing House has four services, including Clearing and Settlement, Usage Control Claim Validation, Logging, and Billing services(Otto, B. et al., 2019).

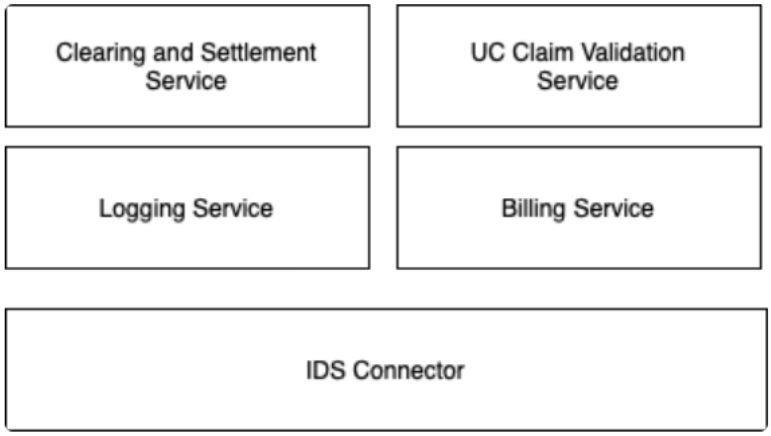


Figure 3-2 IDS Clearing House Architecture in IDS RAM

Based on the requirements defined, architecture in Figure 3-2 above, and description in the IDS's documents, an ArchiMate viewpoint is derived to illustrate the internal behaviour and main functionality of each service in an application behaviour viewpoint in Figure 3-3 below. In addition, this viewpoint also illustrates how each functionality will realise the service expected from a clearing house. Not only limited to exploring the proposed clearing house functionalities and behaviour but this viewpoint may also be used to further design the software in another modelling language[8].

Four main services that a clearing house must include are depicted as an application component. Each application has its own service which realised by the functionality and behaviour of each application. These services, namely, Clearing, Settlement, Logging, and Claim Handling, are exposed to each API endpoint. Starting with the Clearing service, which is responsible for providing clearing functionality that includes validating obligations, such as contract and usage policies agreements, payment arrangements, and consumer identity authentication. If a consumer is able to fulfill these obligations, then transactions may be cleared for settlement and other processes within the transaction. These processes reflect to the functional requirement FR3 described earlier.

[8] https://pubs.opengroup.org/architecture/archimate2-doc/chap08.html

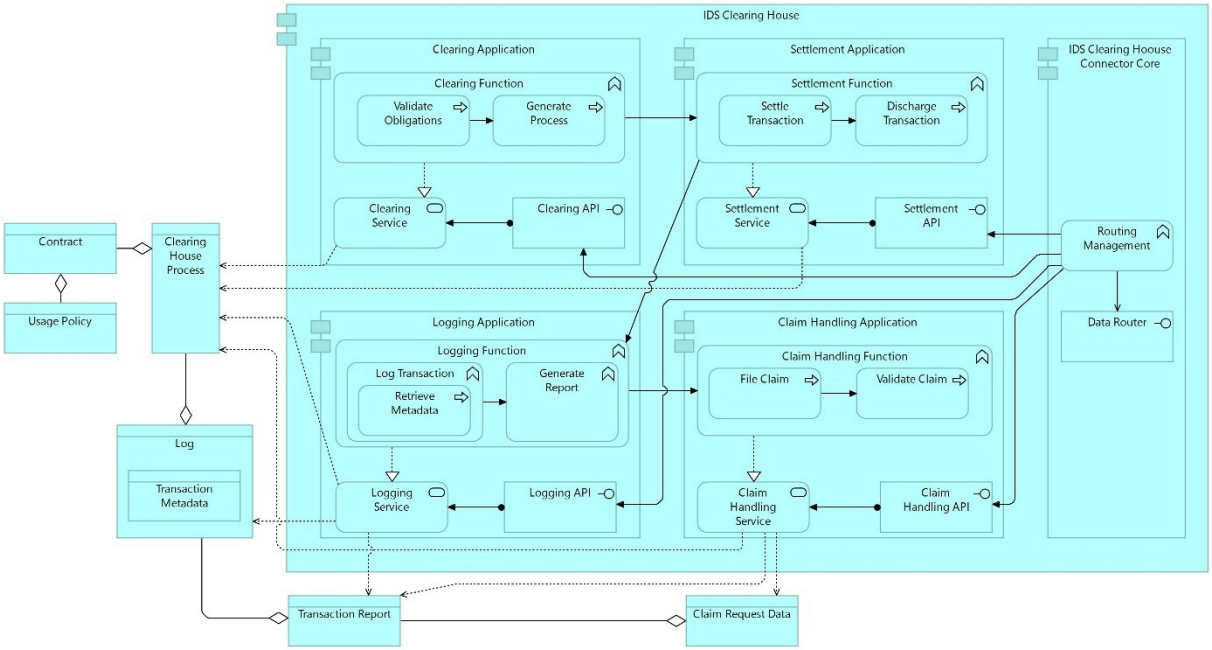


Figure 3-3 Application Behaviour Viewpoint

After a transaction is cleared, a Clearing House Process is generated which aggregates the contract and usage policies defined and agreed upon in the transaction. Later, this Clearing House Process is necessary to be used and referred to during the utilisation of the Clearing House's services during the transaction. For example, the Settlement service is triggered whenever a transaction is cleared and a process is generated. The settlement service may refer to the Clearing House Process to obtain information from the contract and policies to perform the settlement process and function. In a case where a transaction is finished, no longer valid, or blocked, the transaction will be discharged.

Similarly, the logging service is also able to be invoked whenever a Clearing House Process is generated. Each log and trace record of the transaction will be recorded and referred to the described Clearing House Process. These log and trace records are based on the metadata of the transaction. Later, a report may be generated by using the Clearing House Process ID. In case of violation or fraud discovered after the transaction happens, a claim may be requested. Each claim request will be validated before determining the justification required to compensate for the violation and fraud that happened.

While the IDS RAM 4.0 shows the Clearing House is composed of Clearing and Settlement, Logging, Claim Validation, and Billing services, the Billing service is not included in this study as the focus of this study is on the data transaction clearing and settlement, not the commercial nor the financial clearing and settlement. However, if required and implemented in the future, a Billing application should somewhat similar architecture as other main services. The diagram in Figure 3-4 below, illustrates the architecture of the Billing service.

Based on IDS RAM 4.0, there are multiple ways of Billing allowed in IDS based on the contract and usage policy agreed upon. For example, consumers may pay the transaction fee beforehand or after each transaction is made. Therefore, the Billing service must utilise the Clearing House Process as well in order to be able to gather information regarding payment and billing. If the payment is made after each transaction is performed, then the transaction report will be used to calculate the transaction fee. Later, the billing service will send the bill in the form of an electronic invoice on which data consumers may choose the payment platform available. For example, in the Netherlands, payment options like iDEAL will be suitable for small businesses participating in the data exchange.

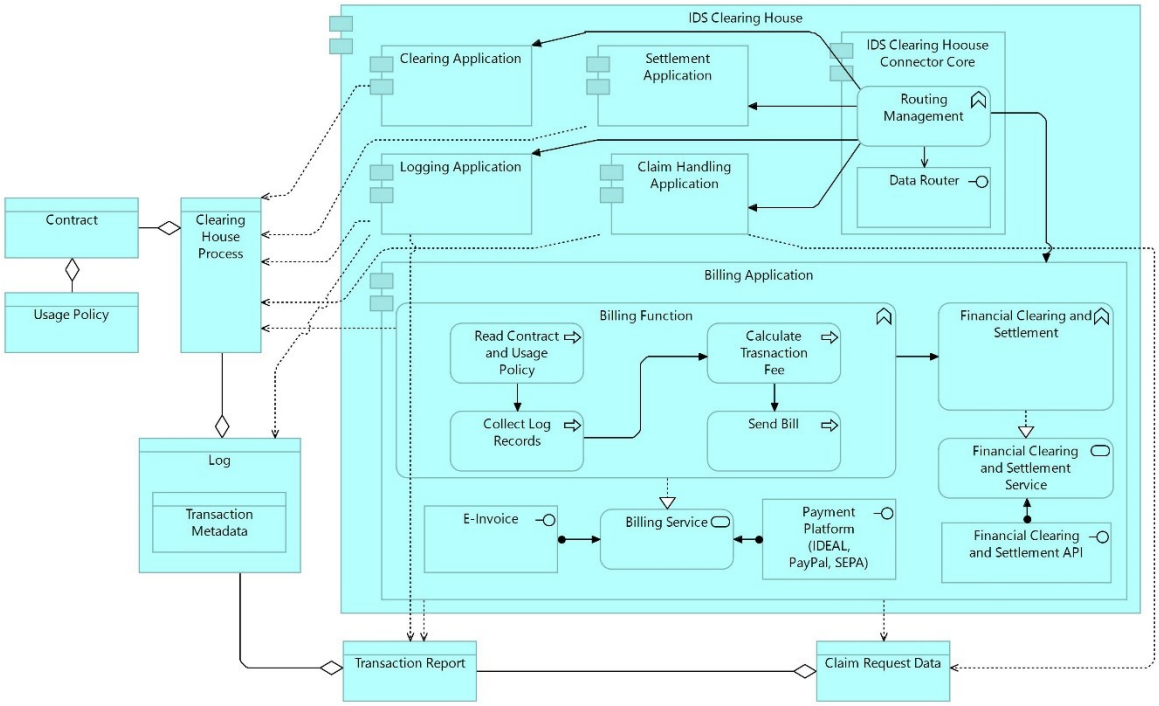


Figure 3-4 Application Behaviour Viewpoint Extended with Billing Application

Each payment made must be cleared and settled by a financial clearing and settlement service. These functionalities should be made available by the bank or another appointed trust financial party. The benefits of having such services also enable the justification platform of the validated claim over fraud or violation that occurred. For example, if a data consumer is found guilty of violating the contract, the Claim Handling may transfer the claim justification process and validation to the financial clearing and settlement service. Finally, All of these main services above are available to data providers and data consumers through an IDS Connector Core. In the upcoming sub-section, the collaboration between each of Clearing House's services and the data providers' and consumers' connectors are explored and discussed.

3.2.3 Application Cooperation Viewpoint

In the previous section, it is shown that Clearing House is composed of at least four different services which realised by each own application module. In order to have data providers and data consumers be able to access each service and functionality offered by the Clearing House, the Clearing House must have a communication module to expose each service to each participant. In order to enable a secure and sovereign transaction, IDS strongly recommends enabling each communication between participants through an IDS Connector (Pettenpohl et al., 2022b). Therefore, an IDS Clearing House must be deployed over an IDS Connector in order to enable the transaction with the participants' connector. Figure 3-5 below shows the architecture of an IDS Clearing House deployed on an IDS Connector environment.

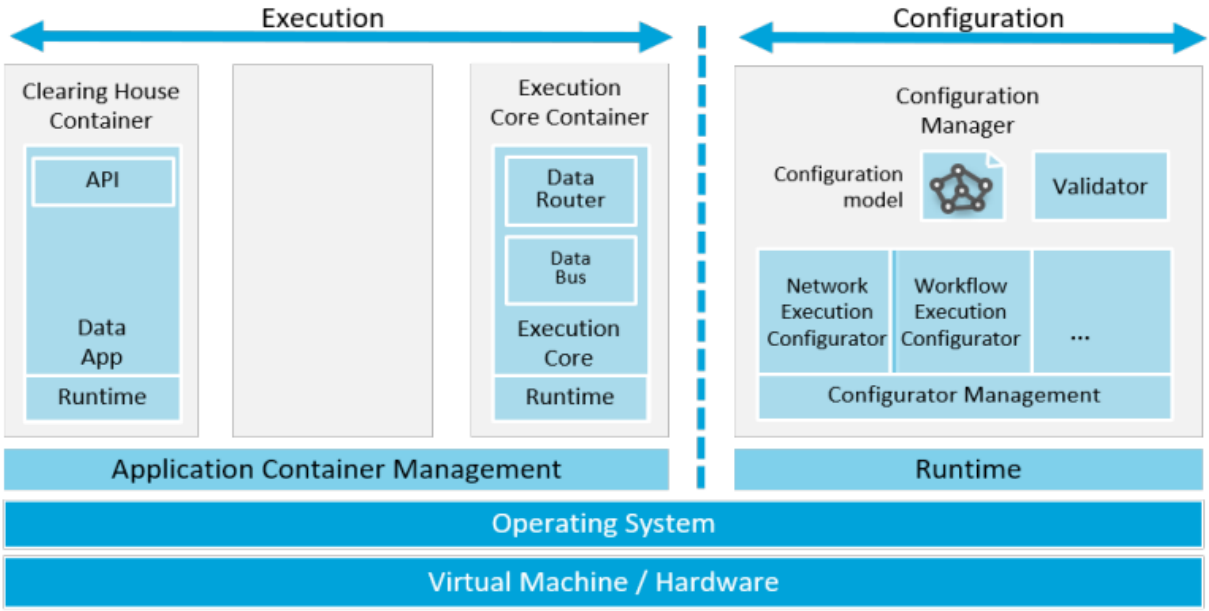


Figure 3-5 IDS Clearing House on IDS Connector Architecture

There are two main containers used in Clearing House deployment on an IDS Connector. Firstly, the execution core container which consists of an Enterprise Integration Framework (EIF). The EIF enables the communication between connectors and orchestration within the services of the Clearing House through a message bus and message router (Otto, B. et al., 2019). There are multiple EIFs available such as Apache Camel, Mule Enterprise Service Bus (ESB), Spring Integration, and WSO2 ESB. In this study, the Apache Camel will be utilised. In sub-section 3.2.4, a justification of the technology stack used to implement the integration framework is provided.

Secondly, there is the Clearing House container which contains of Data App APIs of each service such as Clearing, Settlement, Logging, and Claim Handling services. The use of the integration framework besides enabling communication with other connectors, it is also useful to minimise API calls since the Clearing House consists of many distinguishing and collaborating services through creating or defining certain routes[9]. These routes are managed by the Routing Management function of the integration framework. Each route then is exposed and accessible through the Data Router endpoint which enables communication with other connectors. In the same way, each application or service of Clearing House may also communicate with each other through the Routing Management function. Figure 3-6 below illustrates the Application Cooperation viewpoint between the IDS Clearing House and the IDS Connectors.

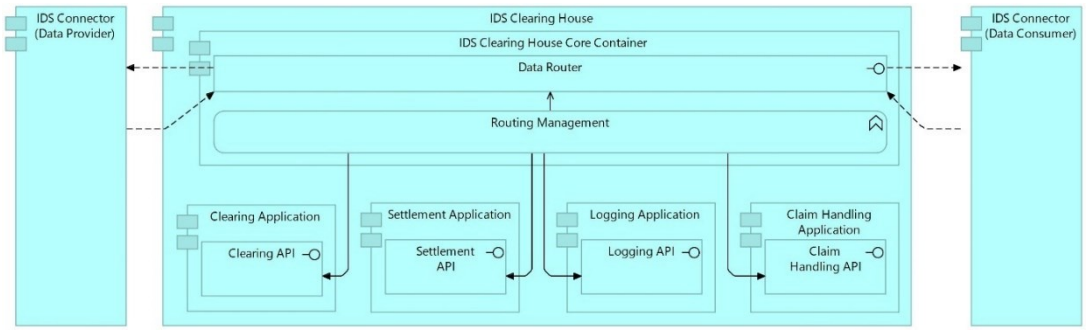


Figure 3-6 Application Cooperation Viewpoint

[9] https://camel.apache.org/

Based on the diagram above, there are message exchanges between the IDS Clearing House with the corresponding IDS Connectors of the participants. Each message sent by the connector will receive a result or response message from the Clearing House. For example, an IDS Connector of a Data Provider is requesting Transaction Clearing to the Clearing House. If the transaction is cleared, the Data Provider will receive a message that the transaction successfully cleared along with the generated Clearing House Process ID. The communication between the IDS Clearing House and the IDS Connectors is shown in the sequence diagram below.

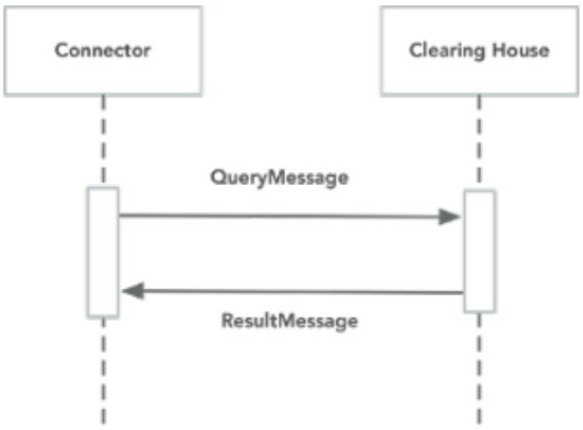


Figure 3-7 IDS Clearing House and IDS Connector Communication Sequence Diagram

Centralised and Decentralised Clearing House Comparison

Figure 3-6 above highly suggests the clearing house proposed in this study is deployed centrally instead of decentralised. A centralised clearing house can be referred to as a trusted intermediary, which enables centralised monitoring, transaction history storage, and also centralised coordination. This approach, however, requires a third party which can be a trusted auditor or auditing institution (Dalmolen et al., 2018). While introducing a third party in a data-sharing transaction may open an opportunity to generate new value, specifically for auditing companies to gain benefit from their services, this may lead to other security concerns where these participating third parties have low integrity.

On the other hand, clearing houses can also be deployed on each node individually meaning that each node participating in the data sharing should have a clearing house installed or deployed on their node. The main benefit of having this approach is to eliminate the need for central or trusted authorities to participate in the transaction. As a result, decentralised approach is more resilient against attacks or security issues such as data misuse by the third party participating in the data sharing (Huang et al., 2017). The introduction of Blockchain technology also increased the popularity of this approach as performed by several previous studies explored in section 2 above.

While decentralised approach is argued to have better security compared to a centralised one, the scalability issue is one of the complex challenges that decentralised clearing house may have. For instance, implementing technology such as public blockchain (i.e. Ethereum, Bitcoin) for clearing houses requires each node to validate and process each message. If there are two nodes, then there will be two times validation and process needed. As a result, communication and network latency might be increased. Furthermore, this also leads to operational costs as discussed in Section 2.3.3.

Regardless, figure 3-8 below illustrates the architecture of the proposed clearing house in a decentralised approach. The clearing house should have the same functionalities and features in order to maintain the main objective of ensuring auditable data sharing. The most noticeable change made in this approach is,

that the data router endpoint is no longer serving two distinct connectors but only a single connector. Each connector should be connected to its own clearing house implementation. As a result, storing transaction history may require more steps than in centralised one. For example, by adding a chain every time messages are logged if using the Blockchain approach (Dalmolen et al., 2018).

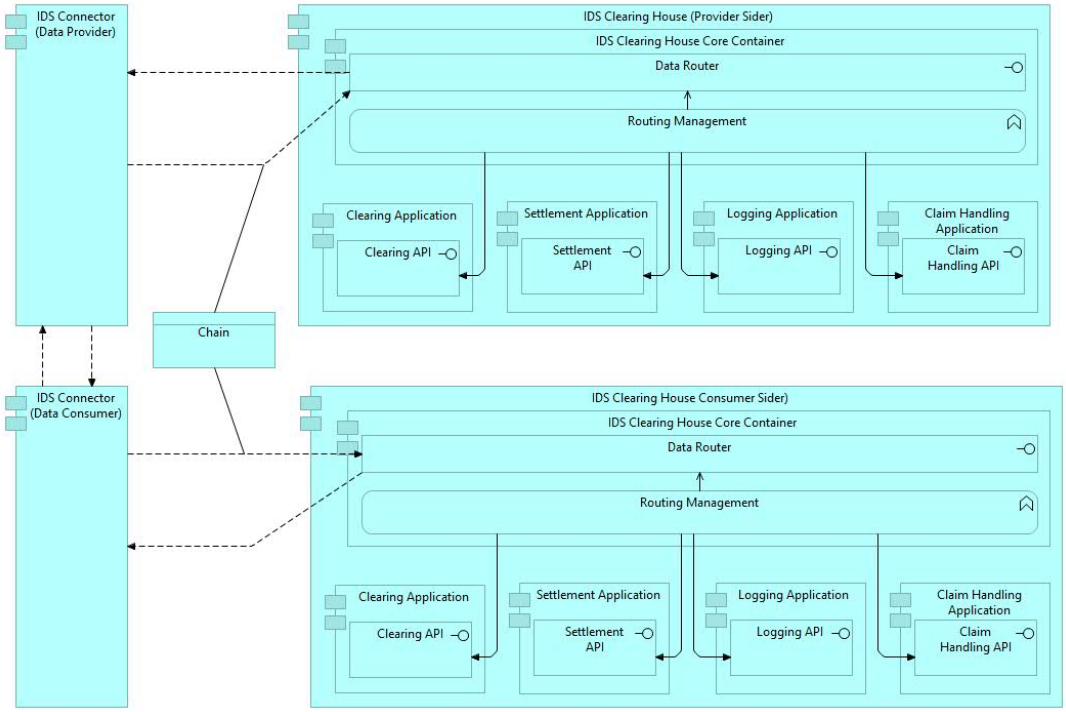


Figure 3-8 Application Cooperation Viewpoint - Decentralised Approach

3.2.4 Service Realisation Viewpoint

In this section, an interaction between human or business actors with the Clearing House is described. As the Clearing House contains four main different services, the Service Realisation viewpoint illustrates how each service could serve business actors' functionalities or activities in performing data transactions which is auditable and traceable. As illustrated in Figure 3-9 below, each use of a Clearing House begins with the agreement between both Data Provider and Data Consumer whether or not a Clearing House involvement is necessary or required to participate in the transaction.

After both agree to use and specify which Clearing House they would like to use or involve, the Data Provider may include the Clearing House in its routing. Through this routing, a connector may start with a request for data transactions clearing to the Clearing House. Later, the process is continued with the following process, such as the Settlement process with settling and discharging the transaction. Then, the Logging process which monitors and records every transaction and message queued to the Clearing House.

Then, the data provider may also request a report of the transaction to the clearing house for auditing purposes. This request is made available by the Logging application through the logging service. Data providers also may file a claim request if they discover a violation of the contract and usage policy by the data consumer. If the claim is accepted, the clearing house will prompt a justification to the data consumer to pay a compensation fee to the data provider. All of these processes are available and accessible by the data provider through the identification of the generated process of the transaction.

On the other hand, the data consumer will have distinct processes with the clearing house. As the data consumer consents to utilise a clearing house, the metadata of each activity and behaviour performed while consuming the data will be recorded. However, data consumers also may file a claim in case of fraud discovered in the consumed data. If the claim is valid, then the clearing house will prompt the data provider to reshare the correct data or compensate the loss of the data consumer to justify this issue. These processes offered by its corresponding service are described in subsection 3.2.1 earlier.

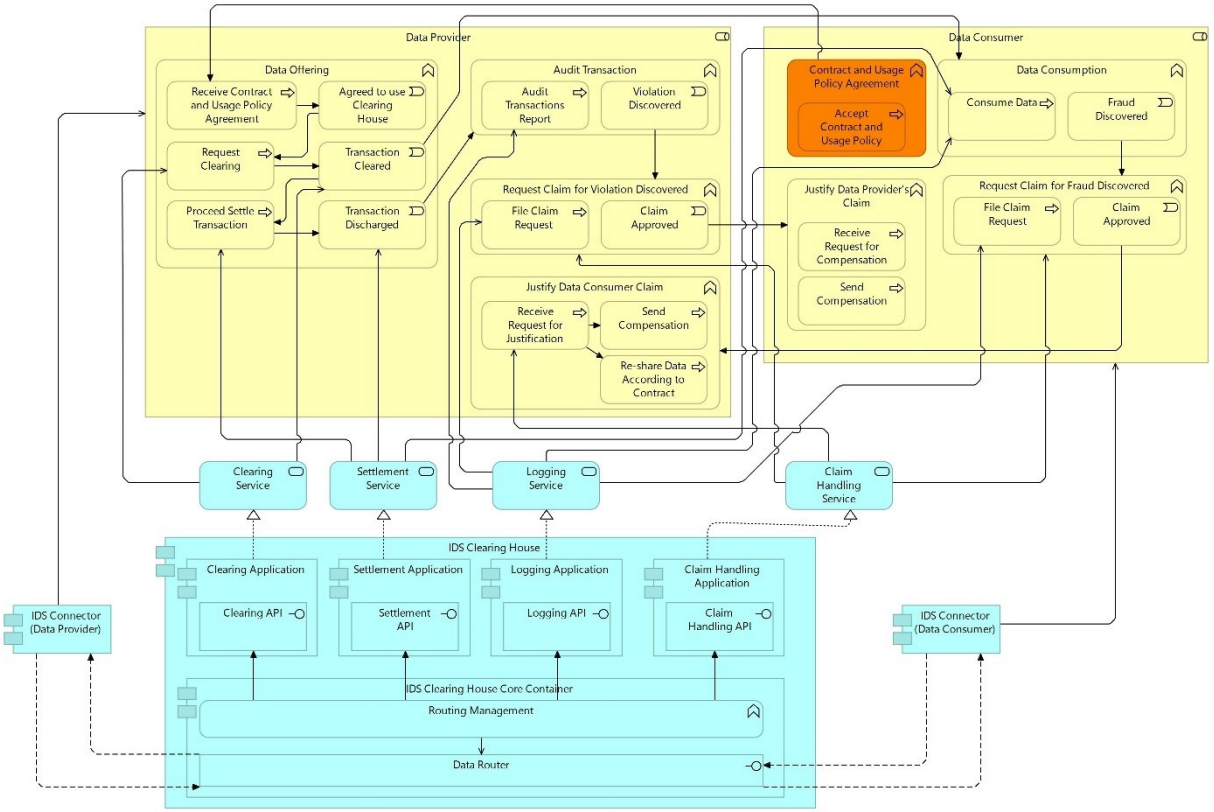


Figure 3-9 Service Realisation Viewpoint

3.2.5 Technology Usage Viewpoint

The IDS RAM 4.0 and specification have been addressed in the recommended deployment environment of the Clearing House. Based on Figure 3-5 earlier, IDS Clearing House is deployed as a containerised application following the architecture of an IDS Connector. Therefore, in Figure 3-10 below, each application module supporting the IDS Clearing House is deployed as a container, including the core connector which is responsible for the communication between other connectors and within the application in the Clearing House.

As a result, Apache Camel is preferred in this study to be used as a technology for implementing the integration framework of the connector. As described in Figure 3-6 in Section 3.2.3, Apache Camel has a Data Router endpoint and Routing Management. Routes are defined in two ways, either by using Java-based DSL configuration or by using XML-based DSL configuration. Furthermore, choosing one out of two DSL options in defining the route should not be an issue as both work the same[10].

[10] https://camel.apache.org/manual/dsl.html

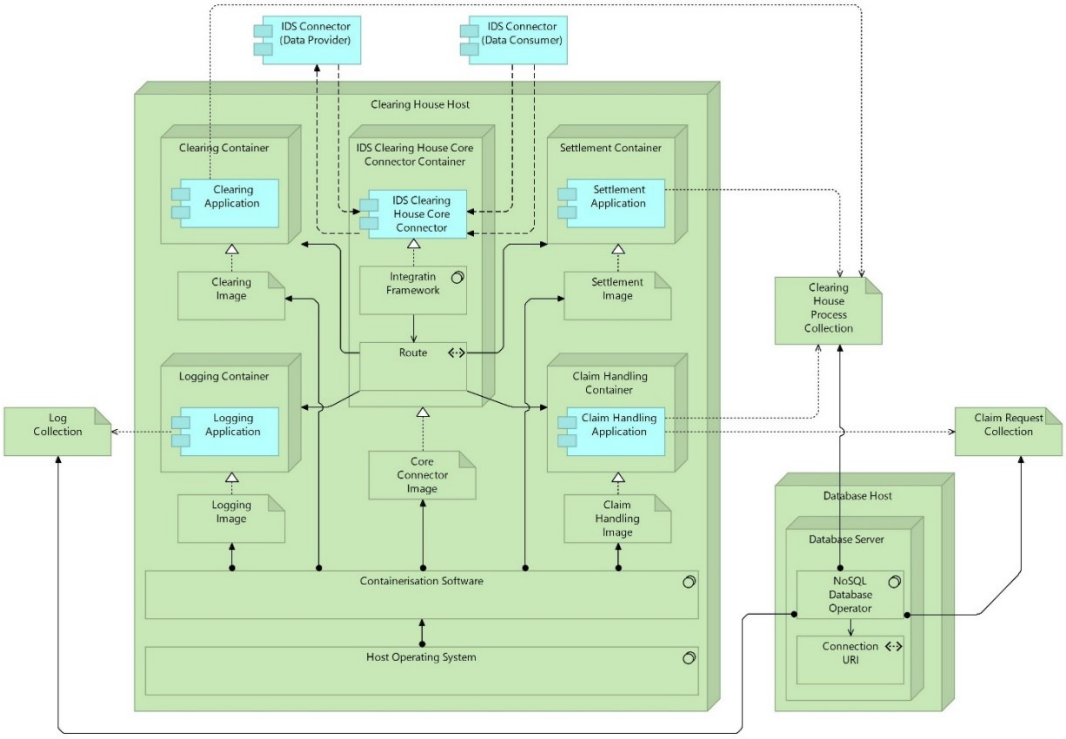


Figure 3-10 Technology Usage Viewpoint

Enterprise Integration Framework Routing

There are several routes that the clearing houses' EIF must have, such as clearing, settlement, report generation, and claim filing routes. These routes are supposed to be accessed by the users of the clearing house. In order to let a user use the services of the clearing house, these routes must be exposed to an API endpoint which can also be considered as an API Gateway. This API endpoint needs to be specified in each participant's connector. While there are many different routes, Apache Camel is able to perform Message Routing based on a specific condition. For example, each API request should have a header. In this header, it can be specified for example a "dst" parameter to tell the router which service is requested by the users.

Besides the API endpoint, the routes must be defined. Basically, the routes should contain the following basic components such as:

1. **From**: Indicating the start point of the route. There are several ways to start a route such as REST endpoint, direct, timer, and so forth. These components also describe where the data is coming from.
2. **To**: Indicating the destination or endpoint of the data being sent. Routes are allowed to have multiple of this component in case of a result or data from a previous destination to be forwarded to the next destination.
3. **Choice**: As discussed earlier, Apache Camel is able to route specific requests based on a specific condition. This Choice Definition component works similarly to the "if" statement in a programming language. So, in a Choice Definition, "Choice" marks the beginning of the conditional statement of the route. Then, there is the "When" component which contains the specific conditions. Lastly, the "Otherwise" component which similar to the "else" statement.

4. **Log:** Report the information of data being processed. It can also be used to show which step is the data currently in.

Table 3-1 below contains code snippets on how a route should be defined for both DSLs according to the description above:

Table 3-1 Apache Camel Route DSLs

Java-based DSL

```
from("sourceEndpoint")
   .choice()
   .when(<<fields>>).<<conditions>>
         .log(“condition X fulfilled”)
         .to("destinationEndpoint")
   .otherwise()
         .log(“no condition fulfilled”);
```

XML-based DSL

```
<routes xmlns="http://camel.apache.org/schema/spring">
   <route>
      <from uri="sourceEndpoint"/>
      <choice>
         <when>
            <simple>_conditions_</simple>
            <log message="Condition X fulfilled"/>
            <to uri="destinationEndpoint"/>
         </when>
         <otherwise>
            <log message="No conditions fulfilled"/>
         </otherwise>
      </choice>
   </route>
</routes>
```

Data Storage

In regard to the functional requirements FR6, FR7, FR8, and FR9, there are multiple ways of storing data in applications. For instance, based on the type of the database there are two types, relational and non-relational database. The relational database, often associated with SQL Database, relies on tables and rows to store the data. Also, a relational database uses “keys” to identify and enable the relationship between each record in the database. There are several Relational Database Management Systems (RDBMS) such as MySQL, Microsoft SQL Server, and PostgreSQL.

The main advantage of using RDBMS is having a concrete and reliable structure. However, this also leads to disadvantages where a complex system with a complex data structure requires more traffic in order to maintain the consistency of the data structure. Therefore, in this case, a non-relational Database or NoSQL database will come in handy and deliver more efficient and better performance in handling data

with complex structures in high traffic (M. Z. Khan et al., 2023). Several alternatives to NoSQL databases are MongoDB, Firebase, Cassandra, and Redis.

In this study, MongoDB is chosen to support the storing of the logging data or any data related to the clearing house. Several considerations for choosing MongoDB over other databases, both Relational and Non-relational, MongoDB is proven to have a high performance in handling high traffic and data (M. Z. Khan et al., 2023). Moreover, the flexibility of the Non-relational database also brings advantages to ensure the adaptability of the Clearing House with distinct data type and structure in other cases. Importantly, due to the requirement of storing the data in a secure way, storing an encrypted value is also proven to be better in MongoDB (Costa et al., 2022).

In regard to data storage, MongoDB stores data as a collection in a document. Each collection is stored as a JSON object. In table 3-2 below, illustrate the schema of a Clearing House Process. As discussed earlier, this process will be referred to during the whole communication with the Clearing House. The schema defines a Clearing House Process must contain a unique ID, provider ID which is the owner of the process, and consumer ID which is the user of the process. The ID of the process is the one that will be used in each service offered by the Clearing House. Meanwhile, both provider and consumer IDs are used to validate the integrity of each participant. For example, by looking at the history of the claim accused toward them or also further be used with an identity provider or other authenticator in a dataspace.

Table 3-2 Clearing House Process Schema

```
{
  "type": "object",
  "properties": {
    "id": {
      "type": "string"
    },
    "providerId": {
      "type": "string"
    },
    "consumerId": {
      "type": "string"
    }
  },
  "required": ["id", "providerId", "consumerId"]
}
```

Then, the following Figure 3-3 shows the schema of a transaction log. A log must have a unique ID, in addition, the processId also needs to be stored for reporting and claim handling purposes later. Furthermore, each time a log is made, the exact local time of the Clearing House host will be recorded as well. There are multiple information to be logged such as IP addresses, MAC addresses, or other metrics. Finally, the transaction metadata is also required to be logged for auditing purposes later in case a violation or fraud is discovered.

Table 3-3 Log Schema

```
{
  "type": "object",
```

```
  "properties": {
    "id": {
      "type": "string"
    },
    "processId": {
      "type": "string"
    },
    "dateTimeCreated": {
      "type": "string",
      "format": "date-time"
    },
    "remarks": {
      "type": "string"
    },
    "transaction-metadata": {
      "type": “object"
    }
  },
  "required": [
    "id",
    "processId",
    "dateTimeCreated",
    "remarks",
    "transaction-metadata"
  ]
}
```

Finally, the claim request schema is defined in Table 3-4 below. Claim request needs a claimRequestId and ProcessId. Then, claimantId contains the ID of the participant who is the victim. On the other hand, the accusedId contains the ID of the participant who performs the violation and fraud. ClaimDescription should describe what is being claimed for. Also, the transactionLogId is the proof of the violation and fraud that occurred, later the auditor may use this information to determine the status of the claim itself in the status field. There is also ClaimDate to indicate when the claim is filed.

Table 3-4 Claim Request Schema

```
{
  "type": "object",
  "properties": {
    "claimRequestID": {
      "type": "string"
    },
    "claimantId": {
      "type": "string"
    },
    "accusedId": {
      "type": "string"
    },
    "claimDescription": {
      "type": "string"
```

```
    },
    "claimDate": {
      "type": "string",
      "format": "date-time"
    },
    "status": {
      "type": "string"
    },
    "transactionLogId": {
      "type": "string"
    }
  },
  "required": [
    "claimRequestID",
    "claimantId",
    "accusedId",
    "claimDescription",
    "claimDate",
    "status",
    "transactionLogId"
  ]
}
```

4. DEMONSTRATION

In this chapter, an instantiation based on the designs and architectures discussed above is developed. The instantiation takes the form of a prototype that addresses the main functionalities and features of a clearing house. Later, this prototype will be used as a tool to demonstrate the importance and uses of a clearing house in specific use case scenarios regarding auditable data sharing.

4.1 PROTOTYPE DEVELOPMENT

The processes and decisions made in regard to the prototype development are discussed in this section. However, the development will be focused on enabling the functionalities rather than the technicalities (i.e., technology stack or components used) of the proposed clearing house. Furthermore, this sub-section will also discuss the deployment of the prototype which is later to be used in demonstrating the clearing house in several use case scenarios.

4.1.1 Development Environment

The prototype is developed with a Spring Boot 3.1[11] framework in Java 17[12] programming language. Mainly, Spring Boot and Java are chosen due to familiarity reason and their popularity in microservices development. Moreover, both also have active maintenance, support, and a large community. In addition, based on the architecture in the previous section, it is understood that a clearing house should be deployed on an IDS connector. Then, there is an example of IDS Connector implementation[13] which is based also on Java and Spring Boot framework. Therefore, the language and framework chosen are argued to be suitable for IDS component development.

The project development started by generating the Spring Boot project using Spring Initializr[14] to easily generate the project with the needed dependencies and libraries. Based on the project's composition depicted in Figure 4-1 below, the project is using Gradle instead of Maven due to familiarity, and it is proven to perform better than Maven and other build automation tools (Prakash, 2022). Then, the following dependencies are needed for all the services except for the core connector module.

1. **Spring Web:** Specifically providing features and tools to develop web applications and RESTful web services. For instance, Spring MVC (Model-View-Controller) and mapping to HTTP requests (GET, POST, PUT, DELETE).
2. **Spring Data MongoDB:** Integrates Spring application to MongoDB database with Plain Old Java Objects (POJOs) mapping, queries, and repository support.
3. **Lombok:** Reduces Java boilerplate by providing annotations for common Java methods at compile time such as Setters, Getters, and Constructors.

Meanwhile, for the core connector applications it needs a different dependency. The Enterprise Integration Framework (EIF) chosen to develop the connector is Apache Camel, in line with the architecture defined earlier. Spring Boot is also popular in terms of EIF development with Apache Camel. In this prototype, the

[11] https://spring.io/projects/spring-boot
[12] https://www.oracle.com/java/technologies/javase/jdk17-archive-downloads.html
[13] https://github.com/International-Data-Spaces-Association/DataspaceConnector/tree/main
[14] https://start.spring.io/

Apache Camel used is 4.0.0-RC1[15] [16], which was the latest version when this report was written. Then the dependencies are composed as follows:

1. **Camel Spring Boot Starter:** provides the necessary configurations and dependencies to easily integrate Apache Camel into the Spring Boot application.
2. **Camel Spring Boot Jackson**: responsible for JSON processing, allowing Camel routes to work with JSON data using Jackson's serialization and deserialization capabilities.
3. **Camel Spring Boot HTTP:** handles HTTP-based communication within Camel routes in Spring Boot applications, facilitating the integration of RESTful services and HTTP endpoints.
4. **Camel Spring Boot Servlet:** handles incoming requests and responses via the servlet API.

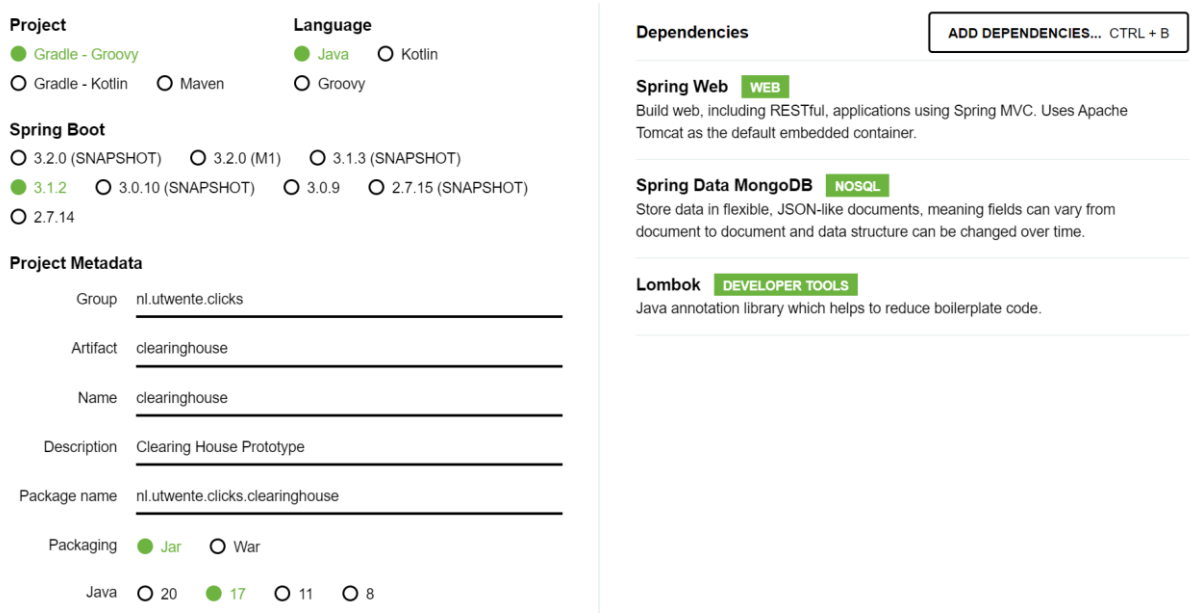


Figure 4-1 Spring Initializr Set Up

Referring to the designs explored earlier, there are five distinct applications within the clearing house. The applications include Clearing, Settlement, Logging, Claim-handling, and Connector Core applications of the clearing house. Each application function is respectful to the designs in Section 3 above. All applications except the core connector are developed in a Model-Service-Controller (MSC) pattern which is illustrated in figure 4-2 below. The pattern is chosen due to providing a structured and organised project. The application modules developed in this pattern should have the following package structure:

1. **Main:** contains the main class of a Spring Boot application, which serves as the entry point of the application to be executed in the Spring context.
2. **Model:** contains the representation of data or business objects. Such as domain classes, stubs, also Mongo repositories.
3. **Service:** contains the business logic of the application. It also bridges the controller and model.
4. **Controller:** Handle incoming HTTP requests and define the RESTful endpoints of the application.

[15] https://camel.apache.org/releases/release-4.0.0-RC1/
[16] https://mvnrepository.com/artifact/org.apache.camel.Spring Boot

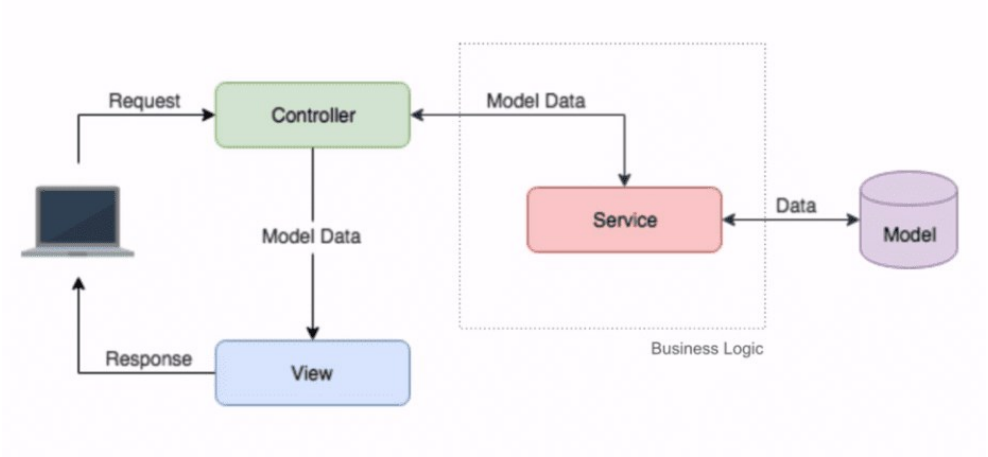


Figure 4-2 Model-Service-Controller Pattern

In contrast, the core connector does not have the MSC pattern as well but instead, it only contains the Main and Route Management class. The main application works the exact same as the other applications. Then, the Route Management class is mainly responsible for constructing and managing the route of the applications. In this project, routes are defined in Java-based DSL instead of XML. The main reason is due to the readability offered by this type of routing. The Appendix 7.1 describes the basic route defined in this prototype. In addition, the XML version of the route is also provided.

4.1.2 Deployment Environment

After the development of the prototype is finished, the prototype is then deployed in a containerised environment. Following the architecture in section 3.2.5, each application should be containerised with respect to the specification of IDS Clearing House and IDS RAM. In this study, Docker is chosen. Each application contains a Dockerfile to instruct how the image of each application should be deployed. The Dockerfile basically contains the following instructions:

Table 4-1 Dockerfile Configuration

```
FROM gradle:8.0.2-jdk17-alpine AS build-[application_name]
COPY --chown=gradle:gradle . /home/gradle/src
WORKDIR /home/gradle/src
RUN gradle clean build --no-daemon

FROM eclipse-temurin:17-jdk-alpine

EXPOSE [application_port]

RUN mkdir /app

COPY --from=build-[application_name] /home/gradle/src/build/libs/*.jar /app/[application_name].jar

ENTRYPOINT ["java", "-jar", "app/[application_name].jar"]
```

Based on the Dockerfile above, each application should have the preferred TCP port defined. It is also important to know that once a port is assigned to a specific application, other applications may not be able to run on the same port. Therefore, each application's port is then defined as follows:

1. Core Connector: 8080

2. Clearing Application: 8081
3. Settlement Application: 8082
4. Logging Application: 8083
5. Claim-handling Application: 8084

After each application's Dockerfile is defined, the docker-compose.yml file should be defined. This file helps to define and manage multiple containers. For instance, it helps to build, remove, start, and stop all containers in one go instead of one by one. Then, the docker-compose file should have the following information:

Table 4-2 Docker Compose Configuration

```
version: '3'

services:
  clearing-house-core-connector:
    build:
      context: ./path-to-connector-module
      dockerfile: Dockerfile
    container_name: coreconnector-container
    ports:
      - "8080:8080"

  clearing-app:
    build:
      context: ./path-to-clearing-module
      dockerfile: Dockerfile
    container_name: clearing-app-container
    ports:
      - "8081:8081"

  settlement-app:
    build:
      context: ./path-to-settlement-module
      dockerfile: Dockerfile
    container_name: settlement-app-container
    ports:
      - "8082:8082"

  logging-app:
    build:
      context: ./path-to-logging-module
      dockerfile: Dockerfile
    container_name: logging-app-container
    ports:
      - "8083:8083"

  claim-handling-app:
    build:
      context: ./path-to-claim-handling-module
      dockerfile: Dockerfile
```

```
    container_name: claim-handling-app-container
    ports:
      - "8084:8084"

  mongodb:
    image: mongo:latest
    container_name: my-mongodb-container
    ports:
      - "27017:27017"
    volumes:
      - mongodb-data:/data/db

volumes:
  mongodb-data:
```

Based on the docker-compose above, it also defines the Mongo database. As a result, whenever the docker-compose is built or started, it runs the MongoDB container which is pulled from the docker hub. However, if there is a MongoDB existing already and the storage should use that database, then the "MongoDB" service and volumes can be removed from the docker-compose file. In addition, both scenarios should affect the application.properties in order to let the application understand which database should be used. The following table gives examples of MongoDB settings in the application.properties file.

Table 4-3 Application Properties Configuration for MongoDB

Docker MongoDB
spring.data.mongodb.host=mongodb spring.data.mongodb.port=27017
Local / Specific host MongoDB
spring.data.mongodb.host=[hostname] spring.data.mongodb.port=27017

After everything is configured, then the following command: "**docker compose up**" can be executed on the directory where the docker-compose.yml file is stored. Then, after all images are built successfully, the applications will be executed right away. By running the "**docker container ls**", all running containers will appear.

```
CONTAINER ID   IMAGE                                          COMMAND                  CREATED         STATUS         PORTS                     NAMES
60ca5b6f65b0   mongo:latest                                   "docker-entrypoint.s…"   6 minutes ago   Up 6 minutes   0.0.0.0:27017->27017/tcp  my-mongodb-c
ontainer
5d3e917c0ed4   clearinghouse-clearing-house-core-connector    "java -jar app/core-…"   6 minutes ago   Up 6 minutes   0.0.0.0:8080->8080/tcp    coreconnecto
r-container
9b47110a45bb   clearinghouse-clearing-app                     "java -jar app/clear…"   6 minutes ago   Up 6 minutes   0.0.0.0:8081->8081/tcp    my-clearing-
app-container
a33e3e41e919   clearinghouse-logging-app                      "java -jar app/loggi…"   6 minutes ago   Up 6 minutes   0.0.0.0:8083->8083/tcp    logging-app-
container
7de95f283b7f   clearinghouse-settlement-app                   "java -jar app/settl…"   6 minutes ago   Up 6 minutes   0.0.0.0:8082->8082/tcp    settlement-a
pp-container
3990a6b14200   clearinghouse-claim-handling-app               "java -jar app/claim…"   6 minutes ago   Up 6 minutes   0.0.0.0:8084->8084/tcp    claim-handli
ng-app-container
PS C:\Users\Gebruiker1>
```

Figure 4-3 Docker Containers of Clearing House Services

4.1.3 Integration with CLiCKS Project

In order to fully simulate how the proposed clearing house is able to interact and serve the participants through their connectors, in this study, a focus on the integration with connectors from the CLiCKS project is aimed. The CLiCKS' connectors are occupied with "Usage Logging" rules which are intentionally prepared for the clearing house integration. Along with the developers of the CLiCKS connector, an integration strategy was established. The strategy consists of several steps, starting with understanding the business logic of the clearing house toward the connector's existing business logic. This includes updating the sequence diagram of the CLiCKS's connector into the following diagram:

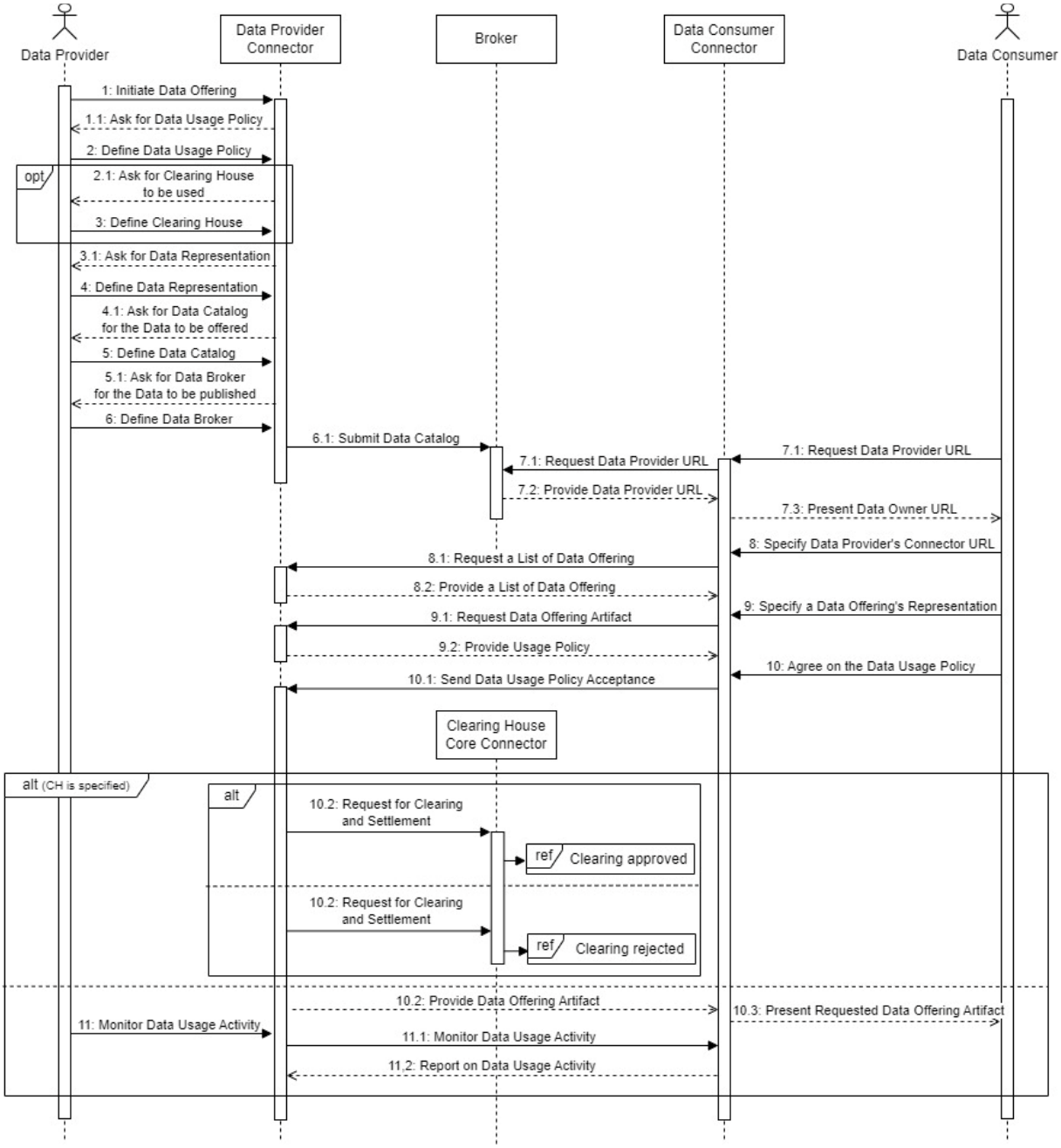


Figure 4-4 Clearing House Connector and CLiCKS Connector Sequence Diagram

Based on the business logic of the existing participant's connector, the clearing house invocation can be defined as soon as the Data Provider specifies the "Usage Logging" rule in their offered data. It is expected that the interface of the connector is able to specify or select the target clearing house. The target clearing

house can be defined earlier similarly to the message broker. In order to specify the clearing house, users may access the Clearing House menu in the connector user interface. Then, the user will be able to add, edit, or delete the clearing houses.

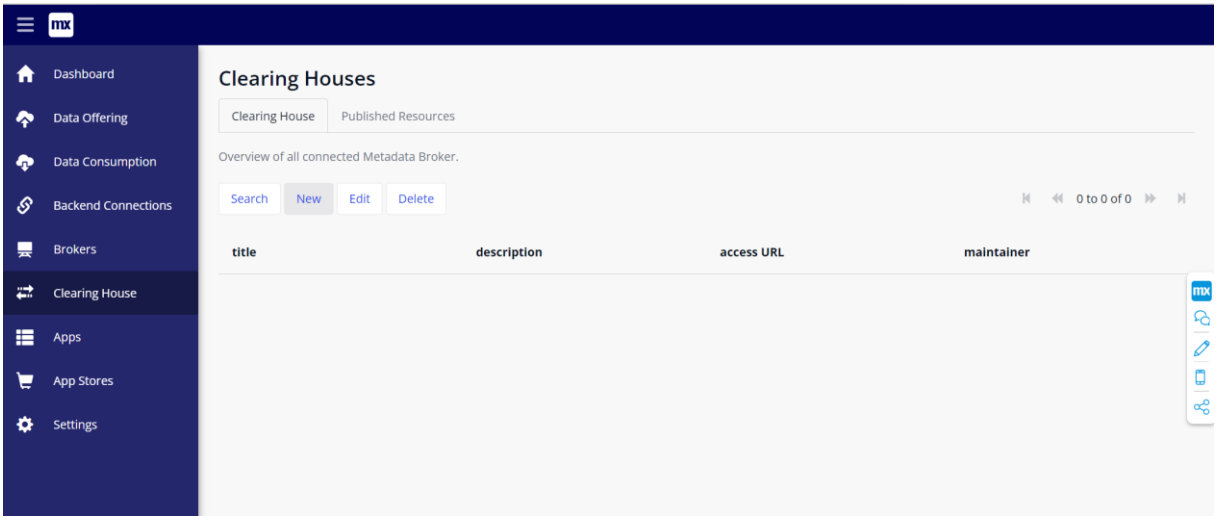


Figure 4-5 Clearing House Menu in CLiCKS Connector

To add a new one, the user is required to specify the name, description, access URL, or the endpoint of the connector of the clearing house, and maintainer information of the specified clearing house.

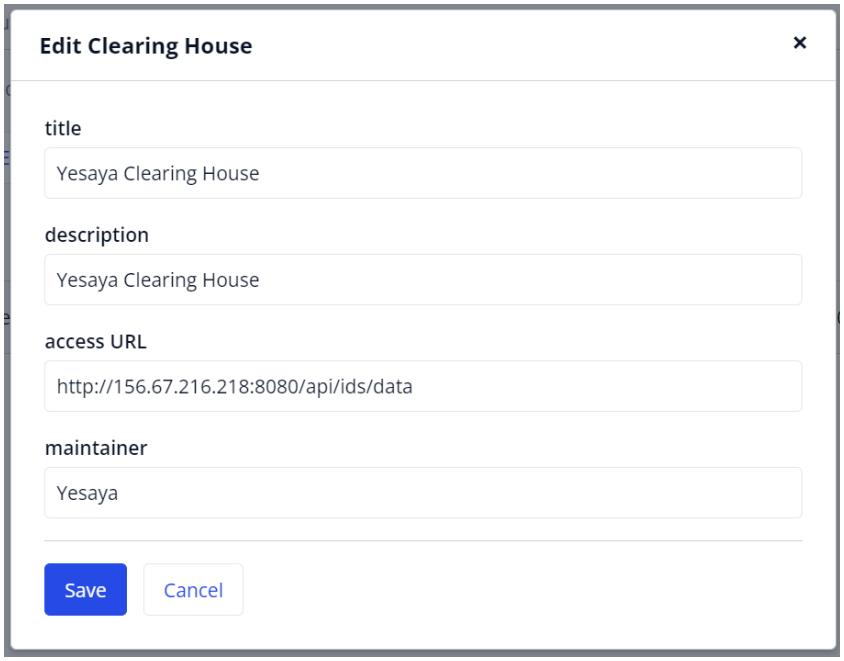


Figure 4-6 Add or Edit Clearing House Information Window

Then, to realise sequence numbers 2.1 and 3, the data provider will be able to specify which clearing house will be used in the transaction during initiating data offering and creating the resource to be shared. clearing house can be specified when choosing the "Usage Logging" policy.

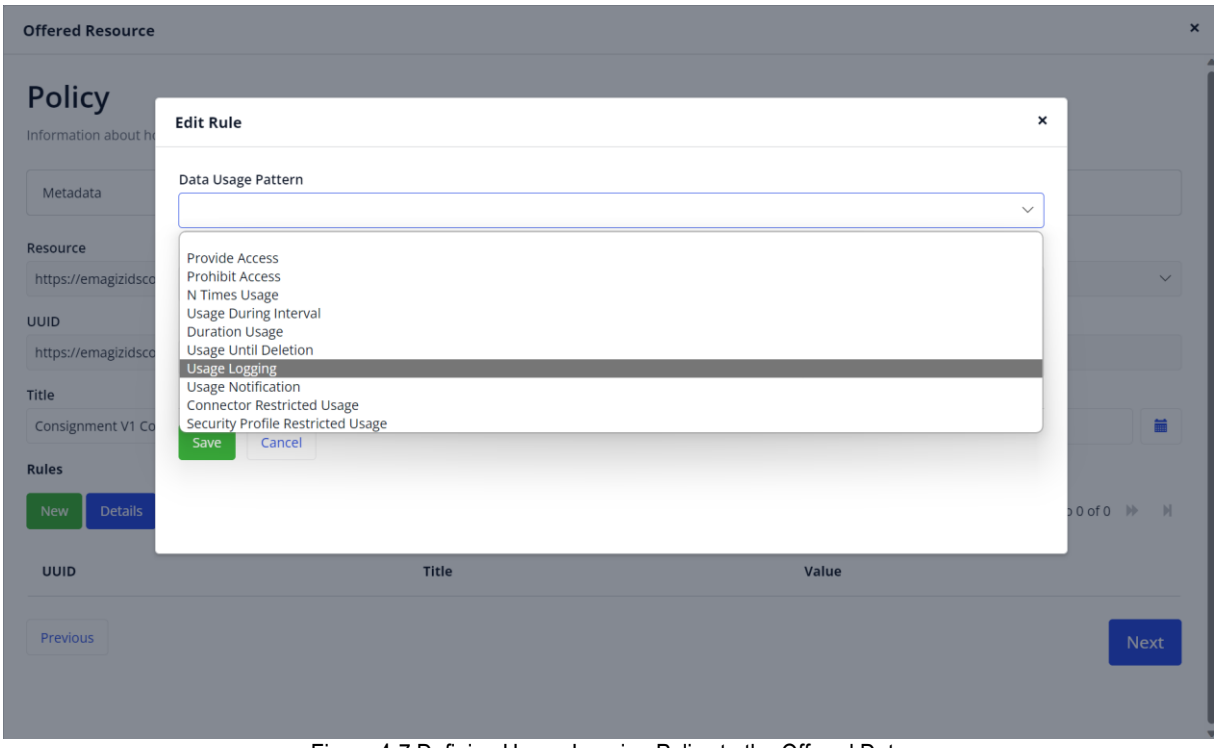


Figure 4-7 Defining Usage Logging Policy to the Offered Data

In this usage policy, the data provider may select the clearing house defined earlier to be used in the data sharing.

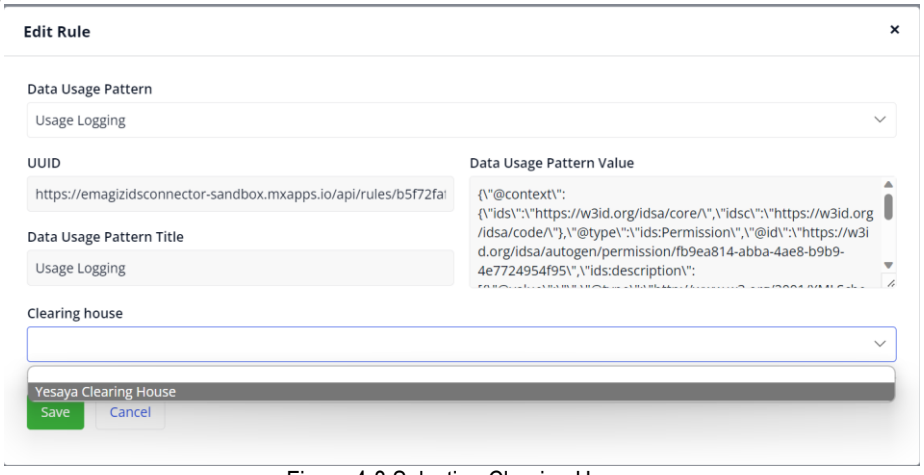


Figure 4-8 Selecting Clearing House

After the clearing house is selected, when the Data Consumer would like to consume the data, it will invoke the clearing house services such as clearing, settlement, and logging immediately.

4.2 USE CASE SCENARIO

In order to prove that the designs and their instantiation are able to address the goal of enabling auditability in data sharing, there are three use case scenarios derived from the IDS Clearing House Specifications. These scenarios include commercial or financial clearing and settlement, and violations that can be performed either by the Data Provider or Data Consumer. Based on those documents, only the last two scenarios will be discussed including the violation performed by the Data Provider and also violation

performed by the Data Consumer. In addition, this scenario currently is demonstrated through the Postman application.

4.2.1 Successful Clearing and Settlement of a Transaction

The first use case scenario will be exploring on how is the expected successful scenario of a data-sharing transaction. In this case, the demonstration will include successful clearing, settlement, and logging. Assuming, both participants have adhered to the contract and usage policies agreed upon. The following Table 4-4 describes this use case scenario.

Table 4-4 Use Case Scenario: Happy Flow - Successful Clearing and Settlement

Use Case Identification	
Use Case Name	Transaction Cleared and Settled
Objective	Transaction performed successfully
Users/Actors	Data Provider, Data Consumer
Trigger	Data Provider and Data Consumer adhered to the contract and usage policy
Preconditions	
1. Contract and usage policies exist	
Basic Flow	
Step	*User Action*	*System Action*
1	Data Consumer accepts the contract	

2	Data Provider receives the accepted contract and requests for Clearing Service	2.1. Clearing House Connector receives clearing request
		The request is sent to the endpoint of the clearing house by specifying “clearing” as a destination in the header
		http://localhost:8080/camel/clearinghouse?dst=clearing
		
		The request should also contain the request body required by the clearing service, which is the contract request body as follows:
		{
		"id":"contract1234",
		"Resourceid": "resource1234",
		"title": "Traffic History Data",
		"startDate": "21-07-2023",
		"endDate" : "25-07-2023",
		"consumer" : "logistiekbedrijft-1",
		"provider" : "logistiekbedrijft-2"
		}
		
		2.2. Clearing House Connector route request to Clearing Service

		coreconnector-container | 2023-07-24T07:51:56.246Z INFO 1 --- [nio-8080-exec-1] s
		tart-clearing-house : Listening to destination: clearing
		coreconnector-container | 2023-07-24T07:51:56.248Z INFO 1 --- [nio-8080-exec-1] r
		oute3 : Start using Clearing House service
		coreconnector-container | 2023-07-24T07:51:56.248Z INFO 1 --- [nio-8080-exec-1] r
		oute3 : Validating contract and generating PID
		
		2.3. Clearing Service clears the transaction
		The clearing service receives the request and request body. Then, it immediately validates the contract also the integrity of the participating connector. The clearing service will validate the integrity by looking at the history of the claim accused toward the participant. Instead of directly invoking the logging service, the clearing service will invoke the route available in the connector.
		coreconnector-container | 2023-07-24T07:51:56.596Z INFO 1 --- [nio-8080-exec-3] r
		oute9 : Getting claim over fraud/violation history acc
		used on: logistiekbedrijft-1
		
		2.4. Clearing Service generates the Clearing House Process
		clearing-app-container | 2023-07-24T07:51:57.013Z INFO 1 --- [nio-8081-exec-1] s
		ervice.ClearingService : Request Clearing House Process' generation
		clearing-app-container | 2023-07-24T07:51:57.140Z INFO 1 --- [nio-8081-exec-1] s
		ervice.ClearingService : Generating new Clearing House Process
		
		The database now has the generated clearing house process
		{
		"_id": "49356e8c-f36d-42ac-aeb9-f14443f9f6e4",
		"contractID": "contract1234",
		"providerID": "logistiekbedrijft-2",
		"consumerID": "logistiekbedrijft-1",
		"_class": "domain.model.ClearingHouseProcess"
		}
		
		2.5. Clearing House Connector forwards clearance result
		When a transaction is cleared successfully, it returns “true” to the connector.

2	Data Consumer Syncs the artifact	3.1. Clearing House Connector receives settlement request
		The settlement service can be invoked through the connector by accessing the following URI with a specifying header processId referring to the Clearing House process created earlier.
		http://localhost:8080/camel/clearinghouse?dst=settlement
		
		In addition, the settlement service also requires a requestBody which consists of the metadata of the artifact being settled and also the usage policies defined. The following request body is an example:
		
		{
		"artifact": {
		"uuid": "art123",
		"remoteid": "rmt123",

```
        "title": "Traffic monitoring june 2023",
        "accessURL":
    "https://www.viamichelin.nl/web/Verkeer/Verkeersinfo-
    Enschede-7511-Overijssel-Nederland",
        "automatedDownload": true,
        "username": "gebruikers1",
        "password": "admin",
        "numAccessed": 0,
        "byteSize": 255,
        "checkSum": "chk123",
        "artifactJSON": "json",
        "apiKey": "api123",
        "routingType": "1"
      },
      "usagePolicy": {
        "uuid": "up123",
        "title": "used until date and 2 times",
        "value": "-",
        "pattern": "-",
        "timesUsage": 2,
        "intervalStart": "21-07-2023",
        "intervalEnd": "25-07-2023",
        "usageDuration": 6
      }
    }
```


3.2. Clearing House Connector route request to Settlement Service


```
coreconnector-container       | 2023-07-24T08:25:15.125Z  INFO 1 --- [nio-8080-exec-1] start-clear
ing-house                     : Listening to destination: settlement
coreconnector-container       | 2023-07-24T08:25:15.128Z  INFO 1 --- [nio-8080-exec-1] route4
                              : Start settlement service
coreconnector-container       | 2023-07-24T08:25:15.128Z  INFO 1 --- [nio-8080-exec-1] route4
                              : Settling transaction for PID: ${header.processId}
```


3.3. Settlement Service settles the transaction

The settlement service then receives the request parameter and request body from the connector and proceeds to settle the transaction. During this settlement process, the settlement service may validate the usage policy and the metadata. For example, transaction validity through the number of access or date of access.


```
settlement-app-container      | 2023-07-24T08:25:15.743Z  INFO 1 --- [nio-8082-exec-1] service.Set
tlementService                : Transaction settled successfully
```


3.4. Settlement Service invokes Logging Service

If the transaction is settled successfully, then the settlement service immediately prompts a logging request to the EIF router.


```
coreconnector-container       | 2023-07-24T08:25:15.793Z  INFO 1 --- [nio-8080-exec-1] route4
                              : Settlement history: {"logId":"1
dbdd405-b2c5-4aa7-9f4f-8c3e15ec5595","clearingHouseProcessId":null,"logOwnerId":"xxx","accessDateT
ime":"2023-07-24T08:25:15.741125008","dateAccessValid":true,"nu
mAccessValid":true,"providerReachable":true,"consumerReachable":true,"remarks":"Transaction settle
coreconnector-container       | 2023-07-24T08:25:15.794Z  INFO 1 --- [nio-8080-exec-1] route5
                              : Logging transaction with PID:
```

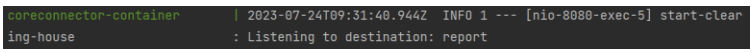
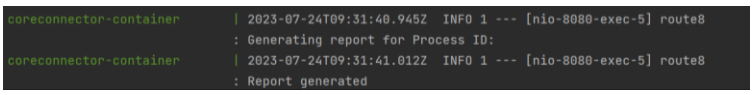
3.5. Logging Service record the transaction

```
coreconnector-container       | 2023-07-24T08:25:16.185Z  INFO 1 --- [nio-8080-exec-1] route5
                              : Transaction logged
```

After successfully logging the transaction, the log is stored in the database.


```
{
  "_id": "1dbdd405-b2c5-4aa7-9f4f-8c3e15ec5595",
  "logOwnerId": "xxx",
  "accessDateTime": "2023-07-24T08:25:15.741125008",
  "dateAccessValid": true,
  "numAccessValid": true,
  "providerReachable": true,
  "consumerReachable": true,
  "remarks": "Transaction settlement succeed on: 2023-07-
24T08:25:15.743000174",
  "artifactBefore": {
    "uuid": "art123",
    "remoteid": "rmt123",
    "title": "Traffic monitoring june 2023",
    "accessURL":
"https://www.viamichelin.nl/web/Verkeer/Verkeersinfo-
Enschede-7511-Overijssel-Nederland",
    "automatedDownload": true,
    "username": "gebruikers1",
    "password": "admin",
    "numAccessed": 0,
    "byteSize": 255,
    "checkSum": "chk123",
    "artifactJSON": "json",
    "apiKey": "api123",
    "routingType": "1"
  },
  "artifactAfter": {
    "uuid": "art123",
    "remoteid": "rmt123",
    "title": "Traffic monitoring june 2023",
    "accessURL":
"https://www.viamichelin.nl/web/Verkeer/Verkeersinfo-
Enschede-7511-Overijssel-Nederland",
    "automatedDownload": true,
    "username": "gebruikers1",
    "password": "admin",
    "numAccessed": 0,
    "byteSize": 255,
    "checkSum": "chk123",
    "artifactJSON": "json",
    "apiKey": "api123",
    "routingType": "1"
  },
  "_class": "domain.model.Log"
```

		}

3	Data Provider monitors the activity	4.1. Clearing House Connector receives report generation request
		 
		
		4.2. Clearing House Connector route request to Logging Service
		 
		
		4.3. Logging Service generates the report
		[
		{
		"logId": "1dbdd405-b2c5-4aa7-9f4f-8c3e15ec5595",
		"processId": "49356e8c-f36d-42ac-aeb9-f14443f9f6e4",
		"logOwnerId": "xxx",
		"accessDateTime": "2023-07-24T08:25:15.741125008",
		"dateAccessValid": true,
		"numAccessValid": true,
		"providerReachable": true,
		"consumerReachable": true,
		"remarks": "Transaction settlement succeed on: 2023-07-24T08:25:15.743000174",
		"artifactBefore": {
		"uuid": "art123",
		"remoteid": "rmt123",
		"title": "Traffic monitoring june 2023",
		"accessURL": "https://www.viamichelin.nl/web/Verkeer/Verkeersinfo-Enschede-7511-Overijssel-Nederland",
		"automatedDownload": true,
		"username": "gebruikers1",
		"password": "admin",
		"remoteAddress": null,
		"numAccessed": 0,
		"byteSize": 255,
		"checkSum": "chk123",
		"artifactJSON": "json",
		"apiKey": "api123",
		"routingType": "1"
		},
		"artifactAfter": {
		"uuid": "art123",
		"remoteid": "rmt123",
		"title": "Traffic monitoring june 2023",
		"accessURL": "https://www.viamichelin.nl/web/Verkeer/Verkeersinfo-Enschede-7511-Overijssel-Nederland",
		"automatedDownload": true,
		"username": "gebruikers1",

		"password": "admin", "remoteAddress": null, "numAccessed": 0, "byteSize": 255, "checkSum": "chk123", "artifactJSON": "json", "apiKey": "api123", "routingType": "1" } }]

4.2.2 Data Consumer Files a Claim

The next use case scenario is regarding a claim filed by the Data Consumer due to contract and usage policy violation by the Data Provider. In this case, the violation might include: sharing low-quality data, sharing incomplete data, or sharing data that does not conform to the contract and usage policy agreed upon. While there are multiple ways these violations might happen, in this scenario, the sharing of incomplete data is demonstrated. Table 4-5 below demonstrates this use case scenario.

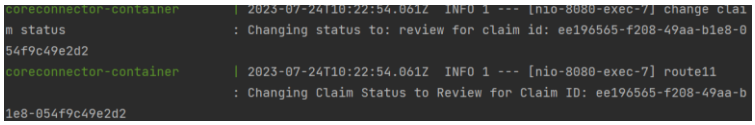
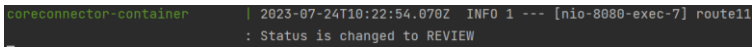
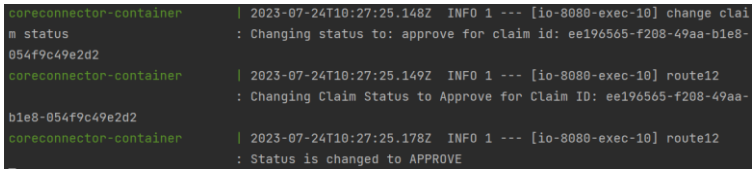
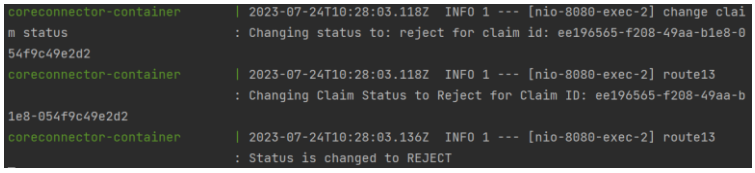
Table 4-5 Use Case Scenario: Sad Flow - Data Consumer Files a Claim

Use Case Identification		
Use Case Name	Data Consumer Files a Claim	
Objective	Justify Data Consumer claim upon the violation performed by the Data Provider	
Users/Actors	Data Consumer, Data Provider, Third Party Auditor	
Trigger	Data Provider shared incomplete data	
Preconditions		
1. Contract and usage policies exist		
2. Claim handling service is available		
3. Data Provider creates incomplete data resource and artifacts		
Basic Flow		
Step	*User Action*	*System Action*
1	Data Provider requests for transaction clearing and settlement	2.1. Clearing Service clears the transaction 2.2. Clearing Service generates the Clearing House Process 2.3. Clearing Service invokes Settlement Service to settle the transaction 2.4. Settlement Service settles the transaction 2.5. Settlement Service invokes Logging Service 2.6. Logging Service record the transaction
2	Data Consumer receives access to the shared data	2.1. Settlement Service discharges the transaction
3	Data Consumer access the shared data	3.1. Settlement Service settles the transaction 3.2. Settlement Service invokes Logging Service 3.3. Logging Service record the transaction

4	Data Consumer discovers the retrieved data is incomplete	-

5	Data Consumer files claim to Clearing House	5.1. Clearing house connector receives claim filing request
		The claim can be requested through the following URI: http://localhost:8080/camel/clearinghouse?dst=claim
		
		with also providing the following request parameters:
		1. claimantId: contains consumerId in this case
		2. accusedId: contains providerId in this case
		3. description: contains explanation of violation performed by the accused participant.
		4. logTransactionId: act as a prove to complement the claim
		
		5.2. Clearing house connector route the request to the Claim-handling service
		 
		
		5.3. Claim Handling Service stores the claim
		 
		
		The stored claim in the database:
		{
		"_id": "ee196565-f208-49aa-b1e8-054f9c49e2d2",
		"claimantId": "logistiekbedrijft-1",
		"accusedId": "logistiekbedrijft-2",
		"claimDescription": "Data Incomplete",
		"claimDate": "2023-07-24T09:43:19.936198283",
		"status": "rejected",
		"transactionLogId": "1dbdd405-b2c5-4aa7-9f4f-8c3e15ec5595",
		"_class": "domain.model.ClaimRequest"
		}

6	Third-Party Auditor reviews the claim	6.1. Clearing house connector receives a request to change the claim status from OPEN to REVIEW To change the status of a claim, the following URI is used: http://localhost:8080/camel/admin/claim-status?status=[targetstatus]&claimRequestID=[claimRequestID] There are four statuses available: 1. Open 2. Review 3. Approve 4. Reject

		6.2. Clearing house connector route the request to the Claim-handling service  6.3. Claim status is changed to review 
7	Third-Party Auditor decides the claim approval	7.1. If the claim is valid, the status can be changed into "Approved" by using the same URI to review but the status is set to approve http://localhost:8080/camel/admin/claim-status?status=approve&claimRequestID=[claimRequestID]  7.2. Otherwise, reject by specifying "reject" in the status parameter: http://localhost:8080/camel/admin/claim-status?status=reject&claimRequestID=[claimRequestID] 
8	Data Provider receives claim justification request	-
9	Data Provider justifies the claim	-

4.2.3 Data Provider Files a Claim

The next use case scenario is regarding a claim filed by the Data Provider due to contract and usage policy violation by the Data Consumer. Data Provider may figure out a violation performed by the Data Consumer in two ways: by auditing the log report and findings by the Policy Enforcement Points. In this study, the focus will be on discovering violations through auditing the log reports generated by the Clearing House. Table 4-6 below demonstrates this use case scenario.

Table 4-6 Use Case Scenario: Sad Flow - Data Provider Files a Claim

Use Case Identification	
Use Case Name	Data Provider Files a Claim

Objective	Justify the Data Provider's claim upon the violation performed by the Data Consumer based on auditing the log report
Users/Actors	Data Provider, Data Consumer
Trigger	Data Consumer violates the contract and usage policy

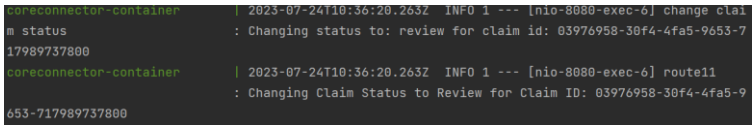
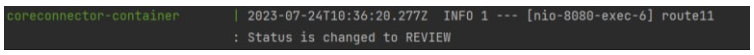
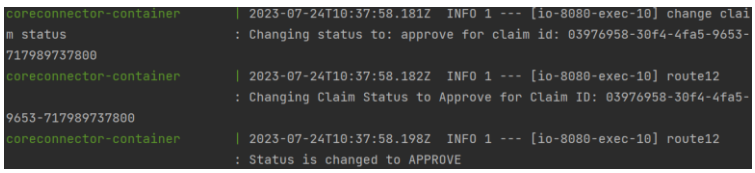
Preconditions

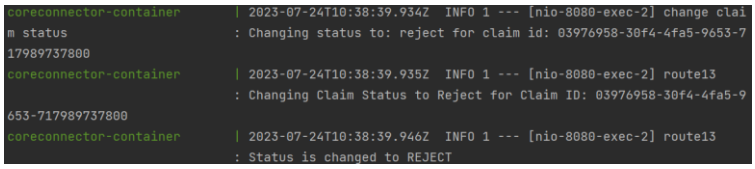
1. Contract and usage policies exist
2. Claim handling service is available
3. Clearing and Settlement have been performed

Basic Flow

Step	*User Actions*	*System Actions*
1	Data Consumers access the shared data	1.1. Settlement Service settles the transaction 1.2. Settlement Service invokes Logging Service 1.3. Logging Service record the transaction
2	Data Consumer proceeds to consume the data while violating the contract and usage policy	2.1. Settlement Service settles the transaction 2.2. Settlement Service invokes Logging Service 2.3. Logging Service record the transaction
3	Data Provider requests the transaction report to the Clearing House	3.1. Logging Service receives report generation request 3.2. Logging Service generates the report
4	Data Provider discovers violation in the transaction record	-

5	Data Provider files claim to the Clearing House	5.1. Clearing house connector receives claim filing request
		The claim can be requested through the following URI: http://localhost:8080/camel/clearinghouse?dst=claim
		
		with also providing the following request parameters:
		1. claimantId: contains providerId in this case
		2. accusedId: contains consumerId in this case
		3. description: contains an explanation of the violation performed by the accused participant.
		4. logTransactionId: act as a prove to complement the claim
		
		5.2. Clearing house connector route the request to the Claim-handling service
		`coreconnector-container | 2023-07-24T09:43:19.776Z INFO 1 --- [nio-8080-exec-8] start-clearing-house : Listening to destination: claim`
		`coreconnector-container | 2023-07-24T09:43:19.776Z INFO 1 --- [nio-8080-exec-8] route7 : Filing Claim`
		5.3. Claim Handling Service stores the claim
		`coreconnector-container | 2023-07-24T09:43:20.126Z INFO 1 --- [nio-8080-exec-8] route7 : Claim filed with status OPEN`
		
		The stored claim in the database:
		`{ "_id": "03976958-30f4-4fa5-9653-717989737800", "claimantId": "logistiekbedrijft-2", "accusedId": "logistiekbedrijft-1",`

		"claimDescription": "Access more than agreed usage policy", "claimDate": "2023-07-24T10:35:10.317000374", "status": "Open", "transactionLogId": "1dbdd405-b2c5-4aa7-9f4f-8c3e15ec5595", "_class": "domain.model.ClaimRequest" }
6	Third-Party Auditor reviews the claim	6.1. The clearing house connector receives a request to change the claim status from OPEN to REVIEW To change the status of a claim, the following URI is used: http://localhost:8080/camel/admin/claim-status?status=[targetstatus]&claimRequestID=[claimRequestID] There are four statuses available: 1. Open 2. Review 3. Approve 4. Reject 6.2. Clearing house connector route the request to the Claim-handling service  6.3. Claim status is changed to review 
7	Third-Party Auditor decides the claim approval	7.3. If the claim is valid, the status can be changed to “Approved” by using the same URI to review but the status is set to approve http://localhost:8080/camel/admin/claim-status?status=approve&claimRequestID=[claimRequestID]  7.4. Otherwise, reject by specifying “reject” in the status parameter: http://localhost:8080/camel/admin/claim-status?status=reject&claimRequestID=[claimRequestID]

		sss 
8	Data Provider receives claim justification request	-
9	Data Provider justifies the claim	-

5. VALIDATION

In order to ensure that the contribution in this research including the minimum software requirements and specification, architectures, and designs aligned to the goal of enabling auditability in data sharing, a validation round is planned. The validation round inquires for expert opinion on the proposed solution. In this section, the validation scenario will be discussed. Later, an analysis and report based on the expert's opinion will also be discussed.

5.1 VALIDATION DESIGN

In this sub-section, the validation scenario will be explored. According to the validation plan described in Figure 5-1 below, the validation consists of two steps. Firstly, the validation round will begin with a presentation to engage the participating panels to understand the main objectives and motivations of this study. Later, a demonstration of the validation model will be conducted. The validation model is discussed in the sub-section 5.1.1. In regards to the participating panel, sub-section 5.1.2 elaborates on the composition of the experts in the participating panel. Secondly, each participant will be handed out a questionnaire that explores their opinion regarding the model being validated. The structure of the questionnaire is discussed in sub-section 5.1.3.

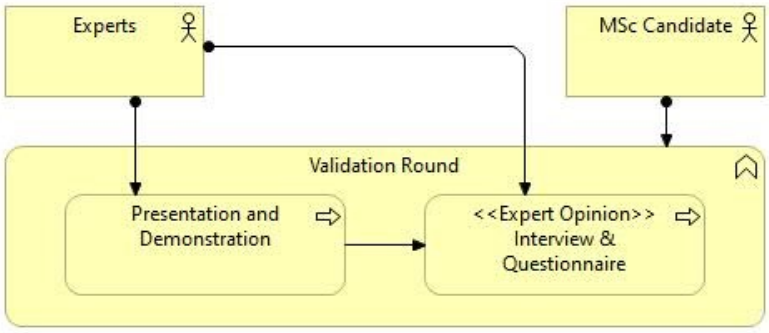


Figure 5-1 Validation Round Plan

5.1.1 Validation Model and Demonstrator

The validation Model in a Design Science Research methodology plays a crucial role in ensuring the artifact fulfills the intended objectives and makes a valuable contribution within the relevant problem domain (Wieringa, 2014). The validation model used in this validation round is the prototype which was discussed previously in section 4.1. The model will act as a demonstrator as well to be used during the presentation and demonstration phase. The demonstrator should consist of all scenarios defined in section 4.2 such as, a happy flow where all participants adhere to the contract and usage policies agreed upon, and also two sad flows which address the situation whenever a participant violates the contract and usage policies.

5.1.2 Participating Panel

In order to maintain the objectives of ensuring that the validated model achieves the intended goal and contribution to auditable data sharing, at least three experts and active contributors in the data space

development are invited to participate in validating this study's result. Table 5-1 below explores the composition of the participating panel. Mainly, the experts in the participating panel are working directly with specific data space. For example, experts from TNO are known to SCSN (Smart Connected Supplier Network) and Industrial Data Space. Then, the expert from academia is highly related to several data space research. Finally, there is also an expert in application monitoring which will help validate the logging functionalities of the proposed clearing house.

Table 5-1 Participating Panel Composition

ID	Role	Organisation	Experience
E1	PhD Researcher	TU Delft	- 3 years as a PhD Researcher on the topic of Business Models for Data Platforms - 2 years as a Business and Technology Integration
E2	Scientist Innovator	TNO	- 5 years as a Scientist Innovator in the architecture and design of data ecosystem and data sharing infrastructure - 4 years as a website developer and network administrator
E3	Software Engineer	CAPE Groep	- 4 years as a Software Engineer specialising in application monitoring - 1 year as a Software Developer

Each expert has been contacted and invited through university email and personal LinkedIn messages. The invitation requests the panel to participate in the validation round between the first and second week of August (1 to 11 August). The given and proposed timespan allows each expert to match the validation round meeting with each schedule. Table 5-2 below describes the availability and confirmed date of each expert in the participating panel.

Table 5-2 Validation Round Schedule Confirmation

ID	Proposed Date	Location	Status
E1	7 August 2023	Online, MS. Teams	Done
E2	8 August 2023	Online, MS. Teams	Done
E3	9 August 2023	Online, Google Meet	Done

5.1.3 Validation Pointers

After finalising the presentation and demonstration phase, each expert will receive a form containing several validation pointers. These validation pointers are defined in order to ensure that the proposed outcome of this study conforms to the main objective and motivation of the study. In addition, each also enables this validation round to capture opinions and feedback from the participating panel. Table 5-3 below describes the validation pointers or questions to be answered by the panel in the validation round. Each pointer is based on the goals defined in the motivation viewpoint in section 3.2.1. The goal reflects the goal in the motivation viewpoint discussed earlier. Then, each goal has at least one question to validate the referred goal. Each question will be assessed with a five scale using the Likert Scale, with the lowest to highest: Strongly Disagree (1), Disagree (2), Neutral (3), Agree (4), Strongly Agree (5). In order to justify each assessment, each expert is allowed to write any opinions and justification regarding the score given towards the question asked.

Table 5-3 Validation Pointers

No	Goals	Questions	Score	Remarks/Feedback
1	Validate participant's integrity to participate in sharing data	To what extent is the proposed architecture enabling trust between participants?		
2	Enable traceability	To what extent is the proposed architecture enabling traceability in data sharing?		
3	Enable auditability	To what extent is the proposed architecture enabling auditability in data sharing?		
4	Enable standardised and applicable reference architecture model	To what extent is the proposed architecture applicable in other data space use cases?		

5.2 QUESTIONNAIRE AND FEEDBACK ANALYSIS

After conducting the validation round with all experts in the participating panels, in this section, all feedback and the opinions of each expert regarding the assessment and validation were given to the proposed architecture. Table 5-4 below elaborates the assessment given by each expert towards the architecture on achieving the main objective of enabling auditable data sharing in a logistic data space. The table contains scores from each expert, then an average score per question, and the total average score is also measured.

Table 5-4 Questionnaire Results

No	Questions	E1	E2	E3	AVG	S_DEV
1	To what extent is the proposed architecture enabling trust between participants?	4	4	3	3.67	0.47
2	To what extent is the proposed architecture enabling traceability in data sharing?	4	5	3	4	0.82
3	To what extent is the proposed architecture enabling auditability in data sharing?	4	4	3	3.67	0.47
4	To what extent is the proposed architecture applicable in other data space use cases?	5	4	4	4.33	0.47
				AVG	3.92	
				S_DEV		0.56

5.2.1 Feedback from E1

E1 discovered the idea of enabling traceability and auditability in data space is interesting and promising, supported by his statement claiming traceability is an important aspect of sovereign data exchange. In regards to the goal of enabling trust and traceability between participants, the demonstrator is capable of fulfilling these goals and has potential. Through the traceability provided, E1 found it very useful to prevent data misuse by verifying data usage. Expectedly, this should also increase participant's trust. However, E1 suggested that scoping the trace is important. For example, by reflecting on data product taxonomy he found will be helpful.

In addition, through the trace provided by the demonstrator, E1 argued that as long as there is a trace then auditability will be enabled. While the demonstrator yield for third parties to audit the transaction, E1 wonder how extensive will the auditing process be performed by these auditing companies. Moreover, aligned with his research in TU Delft about digital platforms and data marketplace, he looked forward to the business and economic value that an IDS Clearing House could deliver in the future. Specifically, for these auditing companies.

Finally, E1 testified that the demonstrator and the proposed architecture provide a solid starting point for further exploration. For example, by finding use case studies of real-world implementation which is important to confirm the feasibility and business model aspects. Overall, E1 expressed that the findings, motivation, and the proposed outcome of this research are promising and good.

5.2.2 Feedback from E2

E2 is a Scientist Innovator specialising in technical architecture and design of data ecosystems and data sharing infrastructure. In addition, E2 was once co-authoring the research regarding decentralised approaches for IDS Clearing House. With this experience and specialisation, E2 gave a lot of insight into the proposed architecture and the demonstrator. Firstly, regarding enabling trust between participants, he argued that as long as both participants consent to involving a third party then it is indeed increasing trust between both participants. Moreover, third-party involvement also limits the chance of data sharing tampering by one or both participants.

Secondly, E2 argued that metadata collected by the clearing house is sufficient enough to provide traceability. Considering the centralised approach proposed, he suggested the clearing house may store the hashes of the data being shared which he argued may be able to detect anomalies in the data sharing easily. Moreover, while his previous research recommended decentralised approach to increase the security and privacy of the data being shared, he testified the proposed architecture is suitable to be developed in decentralised manner (Dalmolen et al., 2018). He also suggested exploring other ledger technology instead of a chain.

Thirdly, E2 stated that auditability depends strongly on the data governance and the interaction between participants and the clearing house. E2 highly suggested developing a governance framework, such as ensuring the original data is always available upon request. However, this can be reserved for future research or investigation. Fourthly, regarding the generality of the architecture, E2 testified the proposed architecture is solid and well-defined. In addition, the architecture can be adapted and applied to other data spaces and use cases with ease. However, in another case such as whenever two clearing houses are required to communicate with each other, the architecture may not be able to accommodate this scenario.

Finally, E2 found the future direction of the findings of this research to be interesting and valid. For example, E2 would like to see how the demonstrator or proposed architecture be integrated with other existing data space connectors such as trusted connectors[17] or eclipse data space connectors[18]. Then, E2 also emphasised the applicability and impact of these findings in real-world use cases or scenarios. Especially, in measuring the trust that the proposed architecture and demonstrator should provide.

5.2.3 Feedback from E3

E3, as a software engineer in Cape Groep specialising in application monitoring, testified that the proposed architecture and demonstrator are indeed enabling trust between data consumer and provider in exchanging the data. However, E3 raised an expectation on how each can trust the clearing house itself. As the architecture suggested the participation of a third-party auditor to operate or deploy the clearing house, E3 implied the importance of having a domain-specific clearing house in order to nurture trust between data providers and consumers toward the clearing houses.

Through his expertise in data logging, E3 had several remarks regarding the traceability features provided. First, E3 concerned the roles and authorisation within the clearing house. In addition, E3 also suggested determining how long will the data be stored in the clearing house considering the amount of data that might be passing through and stored in the clearing house. Secondly, E3 implied the importance of standardisation of the information logged and processed by the clearing house. These courses of action will be useful and helpful for the data providers and consumers in choosing the preferred clearing house.

Additionally, E3 also suggested reflecting on ISO 19011:2018[19] which provides guidelines for auditing management systems. In this guideline, there are seven principles of auditing to which an auditing management system should comply. These principles are integrity, fair presentation, due professional care, confidentiality, independence, evidence-based approach, and risk-based approach. E3 argued most of these principles have been fulfilled, except the “integrity” principle which focuses on fairness, honesty, and, a responsibility[20] that is rather not provided by the demonstrator.

In regards to auditability provided by the clearing house, E3 reflected back to the previous point regarding readability and trust for the data being stored in the clearing house. Moreover, E3 also pointed out that the standardisation of log information is also beneficial during auditing processes. Finally, E3 highly suggested the clearing house be domain specific instead of generalised due to different problems or use cases existing in different data spaces. However, E3 testifies to the potentiality of IDS Clearing House to have a great demand like a broker in e-commerce as long as IDS is able to demonstrate and solve data-sharing problems successfully.

[17] https://industrial-data-space.github.io/trusted-connector-documentation/
[18] https://github.com/eclipse-edc/Connector
[19] https://www.iso.org/standard/70017.html
[20] https://safetyculture.com/topics/iso-19011/#:~:text=ISO%2019011%20establishes%20benchmarks%20for,systems%20and%20establishing%20audit%20programs

6. FINAL REMARKS

In this section, a discussion of the study and the outcomes answer the research question defined in section 1 earlier. Then, challenges and limitations that have influenced the design decision and the final outcome of the research will be elaborated. In addition, several ideas for research continuation in the future will be discussed. Finally, to conclude this study, a conclusion is written summarising the study has been performed.

6.1 CONCLUSION AND DISCUSSION

At the beginning of this study, One main research question which was then divided into four sub-research questions was defined based on the motivation and problem statement of the study. These questions are then addressed thoroughly across specific sections in this report. In this sub-section, each question will be revisited and discussed on how the performed study has addressed and answered the problem and objectives being asked.

6.1.1 What is the state of the art of Clearing Houses in data exchange?

In section 2, there are two parts of the study utilised to gather information regarding the state of the art of a clearing house or trust intermediaries in data exchange. Firstly, an exploratory study was conducted. This part of the study defined the definitions and terminologies that were referred to throughout this study. In addition, an expectation of a clearing house concept in a data exchange was also explored. Clearing houses in data exchange was inspired by the clearing houses in the financial domain which has existed for quite a long time. However, both have different purposes, technologies, and functionalities could offer. Therefore, a comparison of both was discussed. Including, a comparison with a Two-phase commit concept which is also somewhat similar to the expected clearing house in data exchange. Then, these findings became the foundation to conduct more thorough research in academic literature through a systematic literature review.

A systematic literature review (SLR) was conducted to explore more about the importance of implementing such an intermediary in a data exchange transaction. Based on the findings, a trust intermediary such clearing house was mainly implemented to ensure privacy, traceability, and auditability on a data-sharing transaction. Those reasons were enabled by logging or feedback mechanisms over the performed data sharing. In addition, the importance of a clearing house was also in line with the type of transaction which is highly sensitive and valuable data such as personal data, organisational data, or also device measurement data. These kinds of data are prone to privacy leaking or fraud, therefore it is necessary to have a clearing house as a trusted intermediary to ensure trust, traceability, and auditability in a data-sharing transaction.

Then, the SLR went deeper into gathering knowledge of the effect of implementing such intermediaries in the data-sharing environment. While clearing houses or similar intermediaries may increase the efficiency and transparency of data sharing, it can also be costly in terms of development and operation. For example, some technologies may have a greater development and deployment cost compared to one another. Technology such as blockchain, increases security in exchanging data, however, upscaling the clearing house is much more complex and costs higher than implementing the clearing house in other technologies.

Finally, each literature in the SLR was extracted to gather deeper information regarding the architectural components used to develop the clearing house. Most of the studies choose decentralised technologies such as Blockchain and smart contracts, while there is also some alternative such as building the clearing house centrally. Several ideas on how the data is stored and governed were also explored.

6.1.2 How can a clearing house be designed to operate in a Logistic Data Space?

In section 3, the findings from the exploratory study and the systematic literature review on the state of the art of clearing houses were brought into the artifacts proposed in this study. In accordance with the IDS RAM and IDS Clearing House specifications, a set of minimum software requirements and specifications were discussed. Basically, a clearing house in data sharing must have four basic services, including Clearing, Settlement, Logging, and Claim-handling. Then, based on this set of requirements, a set of architectural designs was established.

The architecture designs were established in the ArchiMate language. There are several viewpoints created in order to separate the concerns and domains related to this study. The architecture design begins with the motivation viewpoint which addresses the driver of this study to be conducted in the first place. These drivers then should be addressed by each goal realised by the outcome of this study. Then, an application behaviour viewpoint was established to illustrate how each service application should function and operate. In addition, due to the focus of this study being on clearing houses in data-sharing transactions, the financial functions of the clearing house will not be demonstrated later. However, an architecture including the Billing Service was also discussed.

In the extension of the application behaviour viewpoint, another three viewpoints were established as well including the application cooperation viewpoint which will be discussed in the following sub-section. Then, the service realisation viewpoint thoroughly explores how each service in the clearing house enables the business processes such as requesting from clearing, and settlement until when an unwanted situation happens such as fraud or violation, the business processes on looking for justification through auditing and claim filing were also explored. Finally, the technology usage viewpoint was also discussed which contains design decisions regarding technology choices. The motivation for each decision was derived from the SLR result conducted. For example, the decision to use a centralised approach instead of a decentralised one in regard to the efficiency and scalability offered.

6.1.3 How can the designed clearing house be used with other components in the Logistics Data Space to achieve auditable data exchange?

An IDS Clearing House is typically an IDS Connector with a clearing house's application such as clearing, settlement, logging, and claim-handling. Therefore, a clearing house architecture must follow the architecture of an IDS Connector. For example, the clearing house application must be occupied with a core connector application that functions to route within services and expose the services offered to the participating connector through an endpoint.

Furthermore, a set of use case scenarios was elaborated in demonstrating the functionality of the proposed clearing house in achieving auditable data sharing in logistic data space. To achieve this, an instantiation of the clearing house architecture took the form of a prototype. Furthermore, three main use case scenarios explored the happy and sad flows of a data transaction. In the happy flow, a scenario where a successful clearing and settlement were established due to the conformance of each participant

in data sharing toward the contract and usage policies agreed upon. Then, the sad flow covered two scenarios whenever a data provider or a data consumer violated the contract and the usage policies.

In relation to the logistic data space, an integration between the proposed clearing house and the connector developed in the CLiCKS project was established. These integration strategies were discussed in section 4. In order to have the participants' connectors (Data Provider and Data Consumer) able to access and use the clearing house service, each must define the API endpoint of the clearing house core connector. As a result, several adjustments were made in the CLiCKS connector to enable these integrations.

Furthermore, based on the demonstration using the use case discussed earlier, the clearing house was able to show how each functionality worked. Starting from the clearing service which validates the integrity of the participants through the auditing and claim reporting history. Later, the clearing house was also able to demonstrate how a settlement process was done by validating and ensuring that the transaction was valid. The settlement service was also able to invoke the logging service through the route exposed by the routing management of the connector.

Then, the logs stored by the logging service were also able to be generated as a report which contained some information that were useful for auditing purposes such as the validity of the transaction, until the transaction metadata which can be used to be compared to see if any changes were made. Finally, whenever an unwanted situation happened such as a violation or fraud, both data provider and data consumer were able to file a claim request to the clearing house. Later, trusted third parties operating the clearing house are also able to validate the claim and describe the justification needed.

6.1.4 How usable, useful, and general is the proposed clearing house?

Besides the demonstration discussed in Section 4, a set of validation rounds were conducted. This stage involved three experts in data space development, including three experts. All three experts complement each other, which consists of academia, engineer or contributor in one of the data spaces, and one engineer specialising in application monitoring. During this validation round, each expert was presented with the background and motivation of this study. In addition, each is also able to look at how the prototype demonstrates the functionalities in ensuring and enabling auditable data space. Then, the validation round was concluded by asking for the opinion, feedback, and assessment of the experts towards the proposed outcome in achieving the main goal and objectives discussed earlier. In order to guide the assessment process, a set of questions was prepared and assessed on a five-scale Likert scale.

Based on the validation round, most experts have testified the proposed clearing house is promising, both in terms of economically and technically. The functionalities and features provided were able to demonstrate several goals such as trust, traceability, and auditability. However, most experts highly suggest ensuring that the proposed clearing house has complied with existing standards. Moreover, scoping the authority and responsibility of the third party running the clearing house also needed to be defined carefully and clearly, with respect to those existing standards.

Generally, all of the experts agreed on the generality provided by the architecture to be applied or adapted in other use case scenarios and other data spaces with ease. In order to increase the usability and efficiency of the proposed clearing house, several discussions on making a set of frameworks to be used by the third party. Moreover, the proposed clearing house also suggests that it can be developed in another manner or with other technologies. For example, to be developed decentralised with ledger technology.

6.2 LIMITATIONS AND FUTURE RESEARCH

6.2.1 Limitation

There were several limitations encountered during this study, and each will be discussed in this sub-section. Starting from the instantiation or prototype of the proposed architecture. Due to limited time constraints on completing this study, several alternatives and practices can be explored and implemented. For example, the chosen project environment can be written in a more secure language and frameworks such as Rust. Also, an interactive user interface will be useful in helping the clearing house operator in navigating and using the services. In addition, an encryption algorithm could have been explored to ensure security in storing the data received from the participants' connectors. Moreover, validation of the integrity of the participants could have also been incorporated with components such as identity provider, dynamic attribute provisioning service (DAPS), and also policy enforcement point.

In regards to testing the modularity and adaptability of the clearing house to different systems also was not included in this study. For example, by using the connectors available in the GitHub repository of IDSA such as TrustedConnector and Dataspace Connector, to deploy the clearing house and also to test the communication of each connector to the clearing house. In addition, an option to extend the existing clearing house solutions developed by AISEC Fraunhofer was considered. However, it was unsuccessful to deploy and run the clearing house. Besides the time limit, there were several issues which hindered the findings to be integrated with the existing solutions. The issues such as dependency issues, or also routing issues. Similar issues are also open in their GitHub repository[21].

6.2.2 Future Research

Based on the findings, discussion, and limitations of this study, there are several ideas and recommendations for the continuation of this study. First, is exploring and validating thoroughly the effects of implementing the proposed clearing house on the performance of data exchange. There are Several metrics can be selected such as reliability, resilience, and importantly the security of the data processed by the clearing house itself. In addition, the extension of this project by implementing the Billing service will be an interesting study. The billing service may expose the possibility of commercialising the clearing house in data sharing.

Secondly, performing a thorough comparison of the proposed clearing house implemented in the data space in a centralised and decentralised manner. While several adjustments should be made such as changing the way of storing logs by using Blockchain technology for example. The comparison may also include the performance test as discussed in the previous idea. In addition, it is also interesting to see how the proposed clearing house performed on an existing connector such as Dataspace Connector and TrustConnector. Finally, applying the proposed clearing house on different dataspace would be an interesting study as well.

[21] https://github.com/International-Data-Spaces-Association/DataspaceConnector/issues

REFERENCES

3.5.5 Clearing House. (n.d.). Retrieved 4 April 2023, from https://docs.internationaldataspaces.org/ids-ram-4/layers-of-the-reference-architecture-model/3-layers-of-the-reference-architecture-model/3_5_0_system_layer/3_5_5_clearing_house

Anderson, N., & Edwards, K. (2010). Building a chain of trust: Using policy and practice to enhance trustworthy clinical data discovery and sharing. *Proceedings of the 2010 Workshop on Governance of Technology, Information and Policies*, 15–20. https://doi.org/10.1145/1920320.1920323

Bajoudah, S., Dong, C., & Missier, P. (2019). Toward a Decentralized, Trust-Less Marketplace for Brokered IoT Data Trading Using Blockchain. *2019 IEEE International Conference on Blockchain (Blockchain)*, 339–346. https://doi.org/10.1109/Blockchain.2019.00053

Barenji, A. V., & Montreuil, B. (2022). Open Logistics: Blockchain-Enabled Trusted Hyperconnected Logistics Platform. *Sensors*, *22*(13), Article 13. https://doi.org/10.3390/s22134699

Bargh, M. S., Meijer, R., Choenni, S., & Conradie, P. (2014). Privacy protection in data sharing: Towards feedback based solutions. *Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance*, 28–36. https://doi.org/10.1145/2691195.2691279

Bastiaansen, Bramm, Georg, Ceballos, Juan, Gall, Mark, & Kolenstart, Maarten. (2020). *Specification: IDS Clearing House* (1.0). Zenodo. https://doi.org/10.5281/ZENODO.5675765

Belhi, A., Gasmi, H., Hammi, A., Bouras, A., Aouni, B., & Khalil, I. (2022). *A Broker-Based Manufacturing Supply Chain Integration with Blockchain: Managing Odoo Workflows Using Hyperledger Fabric Smart Contracts*. *640 IFIP*, 371–385. Scopus. https://doi.org/10.1007/978-3-030-94399-8_27

Caytas, J. (2016). *Developing Blockchain Real-Time Clearing and Settlement in the EU, U.S., and Globally* (SSRN Scholarly Paper 2807675). https://papers.ssrn.com/abstract=2807675

Chen, P., Shi, P., Xu, J., Fu, X., Li, L., Zhong, T., Xiang, L., & Kong, J. (2021). TeeSwap: Private Data Exchange using Smart Contract and Trusted Execution Environment. *2021 IEEE 23rd Int Conf on High Performance Computing & Communications; 7th Int Conf on Data Science & Systems; 19th Int Conf on Smart City; 7th Int Conf on Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys)*, 237–244. https://doi.org/10.1109/HPCC-DSS-SmartCity-DependSys53884.2021.00057

Cheng, L., Liu, F., & Yao, D. (2017). Enterprise data breach: Causes, challenges, prevention, and future directions: Enterprise data breach. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *7*, e1211. https://doi.org/10.1002/widm.1211

Choi, J. (Jong U., Ae Chun, S., Kim, D. H., & Keromytis, A. (2013). SecureGov: Secure data sharing for government services. *Proceedings of the 14th Annual International Conference on Digital Government Research*, 127–135. https://doi.org/10.1145/2479724.2479745

Costa, M., Rodrigues, M., Baptista, P., Wanzeller, C., Martins, P., & Abbasi, M. (2022). Database Encryption Performance Impact on PostgreSQL and MongoDB. In J. L. Reis, E. P. López, L. Moutinho, & J. P. M. dos Santos (Eds.), *Marketing and Smart Technologies* (pp. 121–127). Springer Nature. https://doi.org/10.1007/978-981-16-9268-0_10

Dalmolen, S., Bastiaansen, H., Moonen, H., Hofman, W., Punter, M., & Cornelisse, E. (2018). *Trust in a multi-tenant, logistics, data sharing infrastructure: Opportunities for blockchain technology*.

Drees, H., Kubitza, D., Theissen-Lipp, J., Pretzsch, S., & Langdon, C. (2021). *Mobility Data Space—First Implementation and Business Opportunities*.

Esteves, B., Asgarinia, H., Penedo, A. C., Mutiro, B., & Lewis, D. (2022). Fostering trust with transparency in the data economy era: An integrated ethical, legal, and knowledge engineering approach. *Proceedings of the 1st International Workshop on Data Economy*, 57–63. https://doi.org/10.1145/3565011.3569061

Firdausy, D. R., De Alencar Silva, P., Van Sinderen, M., & Iacob, M.-E. (2022). A Data Connector Store for International Data Spaces. In M. Sellami, P. Ceravolo, H. A. Reijers, W. Gaaloul, & H. Panetto (Eds.), *Cooperative Information Systems* (Vol. 13591, pp. 242–258). Springer International Publishing. https://doi.org/10.1007/978-3-031-17834-4_14

Herbst-Murphy, S. (2013). *Clearing and Settlement of Interbank Card Transactions: A MasterCard Tutorial for Federal Reserve Payments Analysts* (SSRN Scholarly Paper 2348276). https://doi.org/10.2139/ssrn.2348276

Huang, Z., Su, X., Zhang, Y., Shi, C., Zhang, H., & Xie, L. (2017). A decentralized solution for IoT data trusted exchange based-on blockchain. *2017 3rd IEEE International Conference on Computer and Communications (ICCC)*, 1180–1184. https://doi.org/10.1109/CompComm.2017.8322729

Huang, Zhu, P., Xiao, F., Sun, X., & Huang, Q. (2020). A blockchain-based scheme for privacy-preserving and secure sharing of medical data. *Computers and Security*, *99*. Scopus. https://doi.org/10.1016/j.cose.2020.102010

Hudaib, A., Masadeh, R., Qasem, M. H., & Alzaqebah, A. (2018). Requirements Prioritization Techniques Comparison. *Modern Applied Science*, *12*(2), 62. https://doi.org/10.5539/mas.v12n2p62

Iacob, M.-E., Charismadiptya, G., Van Sinderen, M., & Piest, J. P. S. (2019). An Architecture for Situation-Aware Smart Logistics. *2019 IEEE 23rd International Enterprise Distributed Object Computing Workshop (EDOCW)*, 108–117. https://doi.org/10.1109/EDOCW.2019.00030

Imeri, A., & Khadraoui, D. (2018). The Security and Traceability of Shared Information in the Process of Transportation of Dangerous Goods. *2018 9th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*, 1–5. https://doi.org/10.1109/NTMS.2018.8328751

Jafari, N., Azarian, M., & Yu, H. (2022). Moving from Industry 4.0 to Industry 5.0: What Are the Implications for Smart Logistics? *Logistics*, *6*(2), Article 2. https://doi.org/10.3390/logistics6020026

Khan, A., & Anjum, A. (2022). Blockchain-based distributed platform for accountable medical data sharing. *Proceedings of the 14th IEEE/ACM International Conference on Utility and Cloud Computing Companion*, 1–8. https://doi.org/10.1145/3492323.3503506

Khan, M. Z., Uz Zaman, F., Adnan, M., Imroz, A., & Rauf, M. (2023). *Comparative Case Study: An Evaluation of Performance Computation Between SQL And NoSQL Database*. *Volume 01*, 10.

Kitchenham, B., & Charters, S. (2007). *Guidelines for performing Systematic Literature Reviews in Software Engineering*. *2*.

Kohli, M., & Suarez, E. (2016). Centralized Solution to Securely Transfer Payment Information Electronically to Banks from Multiple Enterprise Resource Planning (ERP) Systems. *2016 International Conference on Information Technology (ICIT)*, 275–282. https://doi.org/10.1109/ICIT.2016.062

Lechtenbörger, J. (2009). Two-Phase Commit Protocol. In L. LIU & M. T. ÖZSU (Eds.), *Encyclopedia of Database Systems* (pp. 3209–3213). Springer US. https://doi.org/10.1007/978-0-387-39940-9_2

Luo, Y., Liu, M., Wang, J., Yuan, C., & Liu, T. (2022). When Secure Data Sharing Meets Blockchain: Overview, Challenges and Future Prospects. *The 2022 4th International Conference on Blockchain Technology*, 1–8. https://doi.org/10.1145/3532640.3532641

Oh, S. E., Chun, J. Y., Jia, L., Garg, D., Gunter, C. A., & Datta, A. (2014). Privacy-preserving audit for broker-based health information exchange. *Proceedings of the 4th ACM Conference on Data and Application Security and Privacy*, 313–320. https://doi.org/10.1145/2557547.2557576

Otto, B., Steinbuss, S., Teuscher, A., & Lohmann, S. (2019). *IDS Reference Architecture Model* (Version 3.0). Zenodo. https://doi.org/10.5281/ZENODO.5105529

Pasquier, T. F. J.-M., Singh, J., Eyers, D., & Bacon, J. (2017). Camflow: Managed Data-Sharing for Cloud Services. *IEEE Transactions on Cloud Computing*, *5*(3), 472–484. https://doi.org/10.1109/TCC.2015.2489211

Pettenpohl, H., Spiekermann, M., & Both, J. R. (2022a). International Data Spaces in a Nutshell. In B. Otto, M. ten Hompel, & S. Wrobel (Eds.), *Designing Data Spaces: The Ecosystem Approach to Competitive Advantage* (pp. 29–40). Springer International Publishing. https://doi.org/10.1007/978-3-030-93975-5_3

Pettenpohl, H., Spiekermann, M., & Both, J. R. (2022b). International Data Spaces in a Nutshell. In B. Otto, M. ten Hompel, & S. Wrobel (Eds.), *Designing Data Spaces: The Ecosystem Approach to Competitive Advantage* (pp. 29–40). Springer International Publishing. https://doi.org/10.1007/978-3-030-93975-5_3

Piest, J. P. S., Iacob, M. E., & Sinderen, M. J. van. (2021, July 7). A federated interoperability approach for data driven logistic support in SMEs. *10th International Conference on Interoperability for Enterprise Systems and Applications, I-ESA 2020*. 10th International Conference on Interoperability for Enterprise Systems and Applications, I-ESA 2020: Interoperability in the era of artificial intelligence. https://research.utwente.nl/en/publications/a-federated-interoperability-approach-for-data-driven-logistic-su

Pincheira, M., Donini, E., Giaffreda, R., & Vecchio, M. (2020). *A Blockchain-Based Approach to Enable Remote Sensing Trusted Data*. 652–657. Scopus. https://doi.org/10.1109/LAGIRS48042.2020.9165589

Prakash, M. (2022). Software Build Automation Tools a Comparative Study between Maven, Gradle, Bazel and Ant. *International Journal of Software Engineering & Applications*, *13*(3), 1–20. https://doi.org/10.5121/ijsea.2022.13301

Pretzsch, S., Drees, H., & Rittershaus, L. (2022). Mobility Data Space. In B. Otto, M. ten Hompel, & S. Wrobel (Eds.), *Designing Data Spaces: The Ecosystem Approach to Competitive Advantage* (pp. 343–361). Springer International Publishing. https://doi.org/10.1007/978-3-030-93975-5_21

Reniers, V., Gao, Y., Zhang, R., Viviani, P., Madhusudan, A., Lagaisse, B., Nikova, S., Van Landuyt, D., Lombardi, R., Preneel, B., & Joosen, W. (2020). Authenticated and auditable data sharing via smart

contract. *Proceedings of the 35th Annual ACM Symposium on Applied Computing*, 324–331. https://doi.org/10.1145/3341105.3373957

Reniers, V., Van Landuyt, D., Viviani, P., Lagaisse, B., Lombardi, R., & Joosen, W. (2019). Analysis of architectural variants for auditable blockchain-based private data sharing. *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*, 346–354. https://doi.org/10.1145/3297280.3297316

Rodriguez-Garcia, M., Sicilia, M.-A., & Dodero, J. M. (2021). A Privacy-preserving Design For Sharing demand-driven Patient Datasets Over permissioned Blockchains And P2P Secure transfer. *PeerJ Computer Science*, *7*, 1–25. Scopus. https://doi.org/10.7717/peerj-cs.568

Serrano, O., Dandurand, L., & Brown, S. (2014). On the Design of a Cyber Security Data Sharing System. *Proceedings of the 2014 ACM Workshop on Information Sharing & Collaborative Security*, 61–69. https://doi.org/10.1145/2663876.2663882

Si, X., Liu, B., & Zhao, B. (2022). A Chronic Disease Medication Data Sharing Model Based on Blockchain. *Proceedings of the 5th International Conference on Computer Science and Software Engineering*, 22–26. https://doi.org/10.1145/3569966.3569973

Sober, M., Scaffino, G., Schulte, S., & Kanhere, S. S. (2022). A blockchain-based IoT data marketplace. *Cluster Computing*. Scopus. https://doi.org/10.1007/s10586-022-03745-6

Sommerville, I. (2011). *Software engineering* (9th ed). Pearson.

Two Phase Commit. (n.d.). Martinfowler.Com. Retrieved 24 April 2023, from https://martinfowler.com/articles/patterns-of-distributed-systems/two-phase-commit.html

Wang, Y., Hulstijn, J., & Tan, Y. (2018). Regulatory supervision with computational audit in international supply chains. *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*, 1–10. https://doi.org/10.1145/3209281.3209319

Wieringa, R. J. (2014). *Design Science Methodology for Information Systems and Software Engineering*. Springer. https://doi.org/10.1007/978-3-662-43839-8

Xia, S., Zhu, Z., Zhu, C., Zhao, J., Chard, K., Elmore, A. J., Foster, I., Franklin, M., Krishnan, S., & Fernandez, R. C. (2022). Data station: Delegated, trustworthy, and auditable computation to enable data-sharing consortia with a data escrow. *Proceedings of the VLDB Endowment*, *15*(11), 3172–3185. https://doi.org/10.14778/3551793.3551861

APPENDICES

A. ROUTING

Java-based

```
rest("/clearinghouse")
        .post().to("direct:start-clearing-house")

from("direct:start-clearing-house")
        .routeId("start-clearing-house")
        .log("Redirecting to: ${header.dst}")
        .choice()
        .when(header("dst").contains("clearing"))
           .to("direct:clear-transaction")
        .when(header("dst").contains("settlement"))
           .to("direct:settle-transaction")
        .when(header("dst").contains("report"))
           .to("direct:generate-report")
        .when(header("dst").contains("claim"))
           .to("direct:file-claim")
        .otherwise()
           .setBody(constant("Destination is undefined"));

from("direct:clear-transaction")
        .to("http://hostname:port/clearing/clearTransaction?bridgeEndpoint=true")
        .log("Transaction Cleared : ${body}");

from("direct:settle-transaction")
        .to("http:// hostname:port /settlement/settleTransaction?
             processId=${header.processId}&bridgeEndpoint=true")
        .to("direct:log-transaction");

from("direct:log-transaction")
        .to("http:// hostname:port /logging/recordTransaction?bridgeEndpoint=true")
        .log("Transaction logged");

from("direct:generate-report")
        .log("Generating report for Process ID: ${body}")
        .to("http:/ hostname:port /logging/generateReport?bridgeEndpoint=true")
        .log("Report generated");
```

XML-based

```
<routes xmlns="http://camel.apache.org/schema/spring">
   <route id="clearinghouseRoute">
      <from uri="rest:/clearinghouse"/>
      <post/>
```

```
    <to uri="direct:start-clearing-house"/>
  </route>

  <route id="startClearingHouseRoute">
    <from uri="direct:start-clearing-house"/>
    <log message="Redirecting to: ${header.dst}"/>
    <choice>
      <when>
        <simple>${header.dst} contains 'clearing'</simple>
        <to uri="direct:clear-transaction"/>
      </when>
      <when>
        <simple>${header.dst} contains 'settlement'</simple>
        <to uri="direct:settle-transaction"/>
      </when>
      <when>
        <simple>${header.dst} contains 'report'</simple>
        <to uri="direct:generate-report"/>
      </when>
      <when>
        <simple>${header.dst} contains 'claim'</simple>
        <to uri="direct:file-claim"/>
      </when>
      <otherwise>
        <setBody>
          <constant>Destination is undefined</constant>
        </setBody>
      </otherwise>
    </choice>
  </route>

  <route id="clearTransactionRoute">
    <from uri="direct:clear-transaction"/>
    <to uri="http://hostname:port/clearing/clearTransaction?bridgeEndpoint=true"/>
    <log message="Transaction Cleared : ${body}"/>
  </route>

  <route id="settleTransactionRoute">
    <from uri="direct:settle-transaction"/>
    <to
uri="http://hostname:port/settlement/settleTransaction?processId=${header.processId}&amp;bridgeE
ndpoint=true"/>
    <to uri="direct:log-transaction"/>
  </route>

  <route id="logTransactionRoute">
    <from uri="direct:log-transaction"/>
    <to uri="http://hostname:port/logging/recordTransaction?bridgeEndpoint=true"/>
    <log message="Transaction logged"/>
  </route>
```

```
    <route id="generateReportRoute">
        <from uri="direct:generate-report"/>
        <log message="Generating report for Process ID: ${body}"/>
        <to uri="http://localhost:8082/logging/generateReport?bridgeEndpoint=true"/>
        <log message="Report generated"/>
    </route>
</routes>
```

B. VALIDATION ROUND QUESTIONNAIRE RECAP

ID	Begintijd	Tijd van voltooien	E-mail	Naam	Tijd van laatste wijziging	Participant ID	Question 1: To what extent is the proposed architecture in enabling trust between participants?	Remarks/Feedback on Question 1 (optional)	Question 2: To what extent is the proposed architecture in enabling traceability in data sharing?	Remarks/Feedback on Question 2 (optional)	Question 3: To what extent is the proposed architecture enabling auditability in data sharing?	Remarks/Feedback on Question 3 (optional)	Question 4: To what extent is the proposed architecture be applicable in other data space use cases?	Remarks/Feedback on Question 4 (optional)	General remarks (optional)
1	8/7/23 13:53:20	8/7/23 14:00:30	anonymous			E-1	4	Traceability is indeed an important aspect of sovereign data exchange: if we can verify data usage, we can prevent data misuse. One important point is scoping, e.g., what type of data? In which scenario? Reflecting on data product	4	Assuming with better UI other than Postman, I can see the potential to see the data logging. But, I was unsure about the summary: e.g., whether I need to see the log data details, or was there any simple option to verify data usage compliance?	4	As long as you have the logs, you can always audit them. The question is, how hard is it for the audit company to do so?	5	This provides a solid starting point for exploration: Real-world implementation (e.g., via case studies) is important to confirm the feasibility and business model aspects	Overall good point, and promising resaerch! Try to articulate the use case better, and clarify the "scope' (what is included

					taxonomy will be helpful.						vs what is beyond the MSc thesis resaerch). Good luck!
2	8/9/23 18:34:06	8/9/23 18:40:22	anonymous	E-3	3	Enabling trust between data consumer / producer yes, But how can we make sure we can trust the clearing house it self?	3	From what I heard, the data / metadata stored as traceability information is not standardized.	3	How can we ensure the realibility and trust for the data stored inside the clearing house it self?	4	I would say, this is a difficult problem and if solved and IDS success will have great demand like eCommerce broker.

3	8/14/23 10:41:46	8/14/23 10:55:58	anonymous	E-2	4	Especially in scenarios where a trusted third party is present that all participants agree with, the proposed architecture will increase the trust between participants. Given the fact that tampering by one of the parties in a data exchange will be much harder.	5	The clearing house architecture allows tracing information shared by parties. Although the actual information is not shared with the clearing house, the hashes of the data provide adequate trust in that this data is shared and anomalies can be easily detected.	4	The auditability depends strongly in the governance of the data space and how each party interacts with the clearing house. Due to fact that only hashes of the actual data are shared, the governance framework must include a set of rules that ensure that the original data is still available upon request. But this is outside of the technical architecture.	4	Only in cases where a common trusted third party is challenging, the proposed architecture will be difficult to implement. As scenarios can be envisioned where even the fact that two parties did exchange data might be sensitive. In these scenarios, a completely different approach must be used. But in a large portion of data spaces, this is not the case and the proposed architecture can be applied without any problems.	The proposed architecture is solid, for future research the impact and real world applicability should be investigated further. Since trust is often a subjective metric, it is difficult to measure from the architecture alone. Especially when this system is used and parties did

encounter disputes the real value of the clearing house can be evaluated.

UNIVERSITY OF TWENTE
Drienerlolaan 5
7522 NB Enschede

P.O.Box 217
7500 AE Enschede

P +31 (0)53 489 9111

info@utwente.nl
www.utwente.nl