The Identity Management Solution that Improves Data Sharing in Logistics

Master Business Information Technology Thesis

03-03-2023

Author:
R. van Zenden Faculty of EEMCS, University of Twente

Committee:

Dr. M.J. van Sinderen Faculty of EEMCS, University of Twente
Dr. E.J.A Folmer Faculty of BMS, University of Twente
E. de Graaf Medior Scientist: Specialist, TNO

Abstract

Within freight transport and logistics documents, such as the bill of lading, are becoming increasingly digital and need to be verified by authorities. To do so, there is a need for an inter-organizational Identity Management (IdM) solution: policies and technologies regarding the process of Identification, Authentication, and Authorization (IAA) to accurately identify, authenticate, and authorize entities and ensure that only authorized entities have access to the resources required to carry out their professional tasks. Together with experts in the field of the Basic Data Infrastructure (BDI), data spaces, freight transport, logistics, and IdM, requirements for an inter-organizational IdM solution were defined that improves the current solutions used during data sharing in logistics. A systematic approach is provided to define terminologies of requirements. The inter-organizational IdM solutions of the iSHARE Trust Framework (iSHARE), International Data Spaces (IDS), Gaia-X, the European Digital Identity Wallet (EUDI Wallet), and the Sovrin Network (Sovrin) were analyzed and compared based on the technical documentation and interviews with experts of the inter-organizational IdM solutions. Guidelines on how these IdM solutions can be improved are given. The comparison shows that the Gaia-X IdM satisfies the requirements the best. To satisfy all requirements a Transparent Data Space Compliance Assessment Service should be developed, and membership credential issuers and compliance credential issuers should either be controlled by the data space or the data space should support multiple issuers of both credential types. The Gaia-X IdM is an open and decentralized solution that is flexible in the use of standards, reduces the number of credentials a user has to manage, removes the single point of failure, and creates trust. Further research can be conducted on the different legal frameworks and the eIDAS Regulation.

Keywords— Identity Management, Identity Management Models, Identity Management Solutions, Freight Transport, Logistics, Industry 4.0, Data Spaces, Design Science, iSHARE, International Data Spaces, Gaia-X, European Digital Identity Wallet, Sovrin Network

Contents

List of Figures

List of Tables

Acronyms

ABAC Attribute-based Access Control
ADM Architecture Development Method
AI Artificial Intelligence
AMQP Advanced Message Queuing Protocol
API Application Programming Interface
APIs Application Programming Interfaces
ARF European Digital Identity Architecture and Reference Framework
B2B Business-to-Business
BDI Basic Data Infrastructure
CA Certificate Authority
CAB Conformity assessment bodies
CIM Centralized Identity Management
DAPS Dynamic Attribute Provisioning Service
DAT Dynamic Attribute Token
DATs Dynamic Attribute Tokens
DID Decentralized Identifier
DIDs Decentralized Identifiers
DIM Decentralized Identity Management
DLT Distributed Ledger Technology
DPKI Decentralized Public Key Infrastructure
DSRM Design Science Research Methodology
DTM Dynamic Trust Monitoring
Non-qualified EAA Non-qualified Electronic Attestation of Attributes
EDI Electronic Data Interchange
EUDI Wallet European Digital Identity Wallet
eFTI electronic Freight Transport Information
ELM Electronic Logistic Marketplace
EU European Union
FIM Federated Identity Management
GDPR General Data Protection Regulation
GXFS Gaia-X Federation Services
IAA Identification, Authentication, and Authorization
IdM Identity Management
IdM4DSL inter-organizational Identity Management for Data Sharing in Logistics
IdP Identity Provider
IDS International Data Spaces
IDSA International Data Spaces Association
IIM Isolated Identity Management
IoT Internet of Things
iSHARE iSHARE Trust Framework
JWE JSON Web Encryption
JWKS JSON Web Key Set
JWS JSON Web Signature

JWT JSON Web Token

LSP Logistic Service Provider

LSPs Logistic Service Providers

MDCT Data Control Technologies

OCM Organizational Credential Manager

ODRL Open Digital Rights Language

ParIS IDS Participant Information Service

PCM Personal Credential Manager

PID Personal Identifiable Data

Qualified EAA Qualified Electronic Attestation of Attributes

QSCD Qualified Signature Creation Device

QTSP Qualified Trust Service Provider

RBAC Role-based Access Control

SCM Supply Chain Management

SCSN Smart Connected Supplier Network

Sovrin Sovrin Network

SP Service Provider

SPs Service Providers

SSI Self-sovereign Identity

SSO Single Sign-on

TSP Trust Service Provider

1 Introduction

The world is shifting towards an industry where through the networking of various technologies over the internet, goods are automatically created, distributed, used, repaired, and recycled without human interaction, also known as the industry 4.0[42]. Industry 4.0 demands digitization and a data-driven approach to optimizing processes.

In 2018 the Dutch Ministry of Economic Affairs and Climate introduced the Dutch digitization strategy: focus on blockchain, Artificial Intelligence (AI), and Internet of Things (IoT)[83][90]. The objectives, full and streamlined digitization of multimodal freight transport[1] and logistics[1], and a future-proof digital infrastructure for smooth, safe, and sustainable freight transport and logistics in the Netherlands and with trading partners. Building upon that, the Ministry of Infrastructure and Water Management identified a 10-year trajectory focused on realizing paperless transportation, a single government platform for freight transportation, and the development of the Basic Data Infrastructure (BDI). Leading to structural and sustainable innovations in the multimodal transport chain[89]. Resulting in financial savings, decreased administrative responsibilities, and increased accessibility.

The digital transition begins with paperless transportation, demanding a single European transportation market without exclusions or limitations from individual nations. To make this happen, legislation has been in effect as of August 2020 (electronic Freight Transport Information (eFTI)) mandating that by 2025 all Member States must accept electronic data delivery of information when legally necessary in the transfer of commodities[91]. Secondly, a strong digital infrastructure is needed to enable the exchange of data. A requirement to exchange data between parties is interoperable systems. However, there is a lack of interoperable systems, technical incapability, legal restrictions, organizational unawareness, insufficient data savviness, and incompetence to have data-based process operations within freight transport and logistics. One of the main challenges is the lack of adoption of an open standard to share legal documents digitally which is being investigated by FEDeRATED[98].

FEDeRATED is an European Union (EU) project for digital cooperation in logistics. FEDeRATED consists of 15 partners based in 6 EU countries to aid the EU and its partners in designing one infrastructure that unites the current information systems of freight transport and logistic enterprises[98]. The partners are simultaneously working on 23 proof of concepts where interrelated components are developed and tested[99]. The infrastructure for sharing data, BDI, is currently in the design phase. BDI is a federated network of platforms and systems in which data can be shared in a decentralized and open manner. One part of that infrastructure is the ability of governmental entities to request documents digitally. However, the infrastructure enabling Identification, Authentication, and Authorization (IAA) for government entities (e.g., customs department, and Ministry of Infrastructure and Water Management) to businesses (G2B) has not been finalized[55][57]. Policies and technologies are required regarding the whole process of IAA to accurately identify, authenticate, and authorize entities i.e., Identity Management (IdM). IdM ensures that only authorized entities have access to the resources required to carry out their professional tasks[126].

TNO, one of the 15 partners working on FEDeRATED, is a Dutch organization that conducts research and provides advisory services in the fields of applied science, technology, and engineering. TNO was founded in 1932 and has over 3500 employees[124][125]. The organization works with governments, companies, and other organizations to develop and apply scientific and technological knowledge to solve practical problems and stimulate innovation. TNO's areas of expertise include energy, health, defense, and the environment, among others[124].

TNO's department Data Ecosystems (DE) is working on one of 23 proof of concepts and focuses on implementing an IdM solution in BDI. TNO aspires to a solution that is:

- Decentralized - allows the participants of the network to collectively control its specifications and policies.
- Open - allows parties to openly join the network and participate.
- Open - has open specifications and/or guidelines that allow to build the IdM implementation.
- Flexible in the use of standards - allows different standards to be used to build the IdM implementation.

This results in the following research objective:

To design an inter-organizational IdM solution that is open, decentralized, and flexible in the use of standards that improves existing IdM solutions used during data sharing in logistics.

[1]For the definition of the term see Appendix A

1.1 Background

1.1.1 Identity Management

IdM includes policies and technologies concerned with the identification, authentication, and authorization of users (a party or machine) to have access to resources (e.g., applications, systems, networks, or data). The policies and technologies allow assigning user access rights and restrictions based on their identity preventing unauthorized access to resources[1][74][126]. Often organizations assign more user access rights than they need to perform their jobs. Attackers can take advantage of compromised user credentials to gain access to organizations' networks and data. Applying user access policies and rules consistently throughout an organization helps protect resources from threats such as hacking, ransomware, and phishing. Hence, IdM is an essential component for security[116].

IdM has become more important over the past decade with the growing number of regulations aimed at protecting sensitive data from any exposure such as the General Data Protection Regulation (GDPR). It has been a rapidly evolving field in recent years with several key developments. Some of the latest trends and innovations in the field include federated identity, biometric authentication, Identity as a Service (IDaaS), and a decentralized identity. A federated identity is an approach to IdM that allows users to use a single set of credentials to access multiple systems and applications. This helps to reduce the burden of having to remember different usernames and passwords for each separate system, while also providing a more secure and streamlined experience for users[10][13][100]. Biometric authentication uses biometric data, such as fingerprints, facial recognition, and iris scans, for secure authentication and identity verification. Biometric authentication offers higher security compared to traditional methods like passwords and is becoming more widely adopted in a range of applications, from financial services to mobile devices[80]. IDaaS is a cloud-based solution that provides IdM services to organizations. IDaaS offers a more flexible and scalable alternative to traditional on-premise IdM solutions and is increasingly being used by organizations of all sizes to secure their digital assets and improve their overall security posture[58]. A decentralized identity offers a new and innovative approach to IdM that gives users greater control over their personal information, enhances privacy and security, and provides a more secure and streamlined experience for accessing online services and applications[94].

Identification
Identification is the act of recognizing and naming an entity. An entity can claim who it says it is based on something it knows, has, or is, for example, a username, identity card (ID or passport), or fingerprint[29].

Identification is an action taken by a party or on its behalf that leads to either the selection of:

1. A single partial identity[1] that the party possesses based on some data, or
2. A single entity from a given group of entities that the party possesses and that is the subject of a specified partial identity

To do so, parties must be able to both identify individual entities and have a conscious representation of the things they know to exist in order to reason about them. One way to do this is to assign a character string (such as a label or name) to an entity, which would then function as an identifier and may also be used to identify parties in communications between them. If one party mentions an identifier to another (which identifies a particular entity for that party), the other party will be able to figure out which entity the first is referring to. These identifiers can thus be used for the identification of an entity (i.e., recognizing an entity)[67].

The identification can be done by a person or actor (e.g. an IT system). In case an IT system is performing the identification, it is required that a subset of the characteristics of the entity present within the IT system is chosen in such a way that the IT system is able to find the required data to perform the identification. An IT system should be able to use the same information that a person uses for the identification of an entity[106][107].

Authentication
Authentication is the process of providing a set of assurances in such a way that the risk of selecting a partial identity or user record, of which the subject is not the same as the identified entity, is acceptable for the owner of the partial identity or user record. Authentication helps maintain the safety of the system and the privacy of users by reducing the chance of attacks from unknown third parties. Examples of authentication methods are single authentication (i.e., password authentication), two-factor authentication, and multi-factor authentication[106]. After an entity is authenticated and its identity is correct, an extra security technique can be added to determine a user's rights or ability to carry out particular actions within a system (i.e., authorization).

Authorization
Authorization is the process of defining and assigning access policies to identified and authenticated entities that make a system aware of the resources that an entity should and should not get access to. This stops entities from accessing resources (e.g., secure areas, services, systems, data) they do not require to access, preventing invalid transactions from occurring[59]. To make sure that the access policies defined during authorization are also enforced when access is requested, access control is required.

Access Control
The process of access control can be broken down into two phases: the authorization of access during the policy definition phase and the approval or denial of access during the policy enforcement phase. Access requests are accepted or rejected based on the defined authorizations during the policy definition phase prior to the policy enforcement phase. In short, authorization involves defining policies for access and access control regulates the enforcement of these defined policies. Whenever a user tries to access a resource, the access control checks the access policies defined during authorization and provides access accordingly. There are several types of access control of which the main models are the following[82]:

- Mandatory access control
  A central authority controls access policies according to different levels of security.
- Discretionary access control
  The owners or administrators of the system, data, or resource set access policies.
- Role-based Access Control (RBAC)
  Use role engineering to restrict access to resources based on individuals or groups with defined business functions rather than the identities of individual users.
- Rule-based access control
  The system administrator defines rules for access to resources that are often based on conditions such as the time of the day or the location.
- Attribute-based Access Control (ABAC)
  Evaluates a collection of rules, policies, and relationships based on the attributes of users, systems, and environmental factors to manage access policies.

1.2 Research Design

This research aims to design an inter-organizational IdM, referred to as inter-organizational Identity Management for Data Sharing in Logistics (IdM4DSL), that is open, decentralized, and flexible in the use of standards that supports governmental entities and Logistic Service Providers (LSPs) to perform IAA and thereby gain trust to exchange data. The Design Science Research Methodology (DSRM) of Wieringa[131] is applied to realize the design of the IdM4DSL. DSRM provides a high-level overview of the phases to conduct during the design of an artifact and a well-described explanation of the actions to perform per phase. Design science research focuses on designing an artifact and validating its performance with the intention to improve the functional performance of the artifact. The primary objective of design science research is to create knowledge that experts in the relevant field can utilize to create solutions to their problems. DSRM is chosen because it provides adequate guidelines to conduct design science in information systems and software engineering research.

The topic of this study is the artifact in a particular context. Designing and studying this artifact in context to address a design problem in the context are the two main tasks. The artifact in this study is an inter-organizational IdM that guides the IAA of governmental entities to LSPs. This chapter discusses the problem context, objectives, and scope of this research. Then the research methods used during this study are briefly covered and last, the research process applied during this study is explained.

1.2.1 Problem Statement

In the early twentieth century, businesses were often vertically integrated; they performed all functions of the supply chain[2] in house (manufacturing, sourcing, warehousing, sales, and logistics). In the late 1900s, businesses shifted and

[2] "A supply chain is the collection of functional activities through which raw materials are converted into finished products for sale to a customer" source: McLaren et al.[86]

included external partners (e.g., suppliers, transportation providers, retailers) in their supply chain. This demanded Supply Chain Management (SCM) (controlling the activities among supply chain partners), supply chain integration (improving the information flow between links in the chain) supply chain optimization, and supply chain coordination (making decisions that minimize information asymmetry and the resulting excess inventory). Organizations started to turn to horizontal integration, synchronizing supply chain information, and processes between organizations. They turn to methods such as Electronic Data Interchange (EDI), an Electronic Logistic Marketplace (ELM), or collaborative SCM systems[86].

EDI is a standardized data format protocol that automates, integrates, and simplifies Business-to-Business (B2B) communication. Each industry has recurring electronic business documents that are sent and received by trading partners. Hence, they are standardized and made independent of the communications or software technology used[78]. The standardized exchange of these electronic business documents is known as EDI transactions. During EDI transactions EDI standards are used to automate and streamline purchase orders, invoices, payments, and more.

ELM is an electronic hub using web-based systems linking shippers and carriers for collaboration or trading. It acts as an intermediary, facilitating the exchange of logistics services[130]. Teleroute[49] is an example of an ELM that connects transport companies with freight providers. Carriers are able to search for loads and for freight forwards to find available truck capacity. The technology provider of the ELM guarantees payment, performs a quality check for all members, and rates members based on their activity level and payment behavior.

A SCM system is software that supply chain companies use to control the flow of materials, information, and resources for any good or service they sell, from the acquisition of raw materials to the point of delivery. Supply chain partners invest in collaborative SCM systems that facilitate the exchange of operational data, the exchange and coordination of tactical information such as supply and demand forecasts, and support joint planning.

For logistics companies to optimize internal processes and their customer's supply chain, they require advanced systems that provide insights into supply chain information and make use of AI technology[121]. This requires data from a variety of systems and hence logistics need to increase their digitization and data-drivenness. Every system is built on data, but data in itself does not provide value. To derive value from data it must be transformed into information through context and enrichment. Hence, managing data is a key success factor for logistics companies[96]. However, the logistics industry's data infrastructure (including ELM systems and (collaborative) SCM systems) is characterized by a large number of local data 'islands' also known as data silos. The different modalities in logistics each use their own standards and system which limits data exchangeability and results in data silos. These data silos make it challenging to integrate the data and derive economic value from it. As a result, the potential of cross-company or collaborative processes cannot be leveraged[39].

A lot of data is exchanged between two companies via an EDI link. If you want to exchange data with a new company, a new EDI link must be set up which is often unique and thus different from previous connections. Each new connection costs the same investment to set up and manage. This model is very expensive and not scalable. An improvement of this model is data exchange over a cloud connection. This often means that a company's data is stored outside the company and becomes visible to the cloud service provider. For these reasons and because many regulations still require data to be available on paper demanding change, the Dutch Ministry of Economic Affairs and Climate introduced the Dutch digitization strategy in 2018 that focuses on blockchain, AI, and IoT[83][90]. Their aim is a full and streamlined digitization of multimodal freight transport and logistics, and a future-proof digital infrastructure for smooth, safe, and sustainable freight transport and logistics in the Netherlands and with trading partners. In response, the Ministry of Infrastructure and Water Management identified a 10-year trajectory focused on realizing paperless transportation, a single government platform for freight transportation, and the development of the BDI. BDI is a federated network of platforms and systems in which data can be shared in a decentralized, data-sovereign, standardized, and open manner. The infrastructure is currently in the design phase.

BDI is developed as an application around Corda, a scalable, authorized peer-to-peer Distributed Ledger Technology (DLT) platform enabling the creation of apps that promote and provide digital trust between parties in controlled marketplaces[17]. Sensitive data needs to be protected; data may only be accessed if the data owner has granted permission or if there is a legal obligation. Therefore, an IAA infrastructure is required to identify a data requestor, authenticate whether the entity or party is who or what it says it is, and verify whether that entity or party has permission to access the data it requested access to. This IAA infrastructure requires policies and technologies which is known as IdM. Currently, the iSHARE Trust Framework (iSHARE) specifications are implemented in the current version of the BDI and work technically. ISHARE is an initiative of the Netherlands' Logistics Top Sector (partially) funded by the Dutch government. ISHARE specifies a set of standards and technical specifications focused on the

governance and trust[1] of participants to enhance data sharing while maintaining data sovereignty[1]. The technical specifications include how the IAA process should be set up in order to create trust between parties that want to exchange data. Instead of creating many bilateral contracts, all participants have to comply with one federated legal framework and are registered to become digitally verifiable. Within the iSHARE Network, multiple Authorization Registry Service Providers are available allowing data space participants to manage data access or usage rights about other data space participants. It also allows data owners to set specific time frames for data availability, register the data that is shared, and with whom the data is shared. Information regarding participants of the network such as their role, website, phone number, and Authorization Registry ID can be received via Application Programming Interfaces (APIs). ISHARE is a stakeholder of the BDI project and therefore would like to implement their specifications[71][72].

Although iSHARE provides well-established specifications to perform IAA in a way that builds trust between parties, the following challenges are faced:

- The specifications are proprietary and controlled by a single foundation, the iSHARE Foundation. The foundation determines the policies and specifications of the Trust Framework. Participants of the iSHARE Network do get an advisory role but the iSHARE Foundation determines whether to follow the advice.
- ISHARE specifies to use OAuth 2.0[3] which implies that communication is over HTTP; OAuth 2.0 is exclusively focused on HTTP-based applications. The Corda nodes communicate via the Advanced Message Queuing Protocol (AMQP) but not via HTTP. As a result, the current prototype version of BDI is not compliant with that specification. Non-compliance to the iSHARE specifications may exclude BDI, currently under review, from becoming certified as iSHARE compliant and consequently exclude it from taking part in the iSHARE Trust Network.

Problem Statement
The current IdM solution implemented in the design of BDI to accurately perform IAA for G2B is not decentralized and flexible in the use of standards as aspired to by TNO.

Research Context
The context in which BDI will be applied is freight transport and logistics. In freight transport and logistics an increasing amount of parties[1] are starting to collaborate by sharing data to reach their common goal more efficiently (e.g., faster time to market). These collaborations are known as data spaces[1]. These data spaces consist of LSPs (e.g. maritime shipping companies, marine terminals, depots, trucking companies) but also government entities such as customs who supervise the import, export, and transport of goods and compliance with safety, health, economy, and environment. To do so customs requires information about, for example, the arrival of a container and the products that are being imported or exported. This information should be made available by the party that is shipping these products. Formerly, this information was exchanged via paper documents and is still being done. However, the majority of parties have moved to digital documents to meet the objectives set by the EU.

Imagine two parties of a data space: Dave (an employee of Douane; the Dutch customs) and Lisa (an employee of a Logistic Service Provider (LSP)[1]. Lisa is the shipper of a container of goods from China and needs to import it into the Netherlands. Dave wants to check whether taxes due on imports are paid and whether goods comply with regulations hence, requires information about such. Just before arrival, Lisa updates the system of her employer that the container is about to arrive at the harbor of Rotterdam. This update is pushed to the system of the Douane, to notify them of the event. The message received by the Douane consists of information such as time of arrival, container ID, and country of departure. The Douane runs a rule-based prediction model that notifies Dave of a suspicious container, the one from Lisa. Dave takes action and requests additional information about the goods inside Lisa's container. This is done via the system of Douane that request additional information from an API provided by Lisa's employer.

Another process is that Lisa's container is randomly selected by Dave for inspection. In that case, Dave directly requests additional information about the goods inside Lisa's container. This is done via the same system of Douane that request the additional information from the same API provided by Lisa's employer. Dave inspects the container and checks whether this matches the information gotten from the API.

Before an update is pushed to the system of Douane or additional information is requested from an API provided by Lisa's employer, IAA takes place. For simplification reasons it is explained as Lisa and Dave are the ones that perform

[3] an authorization mechanism that is primarily intended for use in allowing access to a collection of resources, such as remote APIs or user data. It uses access tokens, which are pieces of information that an application receives, for example, from an Application Programming Interface (API) after authentication and can be used to access resources[56]

IAA while in fact this is performed by machines. Identification involves two elements: the X.509 certificate and the client assertion. The X.509 certificates are issued by a trusted Certificate Authority (CA) that verifies whether Lisa is indeed Lisa or Dave is indeed Dave and digitally signs the X.509 certificate as evidence that the X.509 certificate can be trusted. The client assertion is used by Lisa to prove her identity. The client assertion includes information such as the subject of the client assertion (coming from the X.509 certificate), the audience for which the client assertion is created, the time when the client assertion was issued, and the time when the client assertion expires. The client assertion is added to the request for an access token that Lisa can use to push the information to the system of Douane. Before Lisa sends the request, the header, and payload are encoded and signed with the private key of her X.509 certificate using an encryption algorithm. Authentication is performed by verifying the private key with the associated public key and using the encryption algorithm located in the header. In case the public key is the one linked to the private key with which the client assertion is signed, it results in a hash value that matches the original hash value meaning the signature was indeed set by Lisa. Because Dave trusts the mathematics behind the encryption algorithm and trusts the verification of the CA that Lisa is Lisa, Dave trusts that the request sent is really coming from Lisa. Dave now authorizes Lisa to push the information to the system of Douane and issues an access token to Lisa to do so. Lisa identifies herself with the access token that is included in the request to push information to the system of Douane. Dave verifies the access token and processes the information in the system of Douane.

For simplification reasons it is explained as Lisa and Dave are the ones that perform IAA while in fact this is performed by machines. Additionally, it is explained as a pull or push is done to the system of Douane or Lisa's employer while in fact the pull or push is done via a specific API of the system. The IdM works two ways in the context:

1. A pull (i.e., GET API call) from Lisa to the system of Douane where she identifies herself in order to receive an access token. Then a push (i.e., POST API call) from Lisa to the system of Douane, my container is about to enter the harbor.
2. A pull (i.e., GET API call) from Dave to the system of Lisa's employer where he identifies himself in order to receive an access token. Another pull (i.e., GET API call) to the system of Lisa's employer for additional information.

1.2.2 Research Objectives

The primary motivation for this thesis was to guide the TNO DE department and the scientific community by filling the research gap of an open, decentralized, and flexible in the use of standards inter-organizational IdM solution. This research aims to design an inter-organizational IdM solution that is open, decentralized, and flexible in the use of standards that improves existing IdM solutions used during data sharing in logistics.

An IdM provides the policies and technologies required to accurately perform IAA and ensure that only authorized entities have access to resources. The IdM enables verification that information sent by an entity (e.g., identity information) actually comes from that entity, thereby creating trust such that the parties will exchange data. Hereby, the following main research question is defined:

What inter-organizational IdM solution is open, decentralized, and flexible in the use of standards and improves existing IdM solutions used during data sharing in logistics?

The primary research objective is split up into smaller objectives and research questions:

Research Objective 1 (RO1): Knowledge regarding IdM must be acquired to get an understanding of the state-of-the art, challenges, and causes of those challenges. Having this knowledge enhances communication with stakeholders and experts, and aids in designing an accurate inter-organizational IdM solution to achieve the primary research objective. Hence, the first research question is defined with two sub-research questions:

RQ1: What are the challenges posed by the state-of-the art IdM models?
SRQ1.1: What are the state-of-the-art IdM models?
SRQ1.2: What are the challenges experienced regarding the IdM models and what is causing them?

Research Objective 2 (RO2): The collection of requirements from stakeholders regarding the inter-organizational IdM solution that improves existing IdM solutions used during data sharing in logistics. Hence, the second research question is defined:

RQ2: What are the requirements for an inter-organizational IdM solution that improves the existing solutions used during data sharing in logistics?

Research Objective 3 (RO3): Knowing what state-of-the-art IdM model best fits the requirements of the research to have guidance about the IdM model that should be pursued during the design of the artifact. Hence, the third research question is defined:

RQ3: What state-of-the-art IdM model does best fit the requirements?

Research Objective 4 (RO4): Knowledge regarding the existing inter-organizational IdM solutions is required, how they satisfy the identified requirements, and how they can be improved. Having this knowledge enhances communication with stakeholders and experts, and aids in designing an accurate inter-organizational IdM solution that satisfies all identified requirements and improves existing IdM solutions used during data sharing in logistics. Hence, the fourth research question is defined with two sub-research questions:

RQ4: What inter-organizational IdM solutions are used during data sharing in logistics and to what extent do these solutions satisfy the identified requirements?
SRQ4.1: How do these inter-organizational IdM solutions satisfy the identified requirements?
SRQ4.2: How can the inter-organizational IdM solutions be improved?

Research Objective 5 (RO5): Design an accurate inter-organizational IdM solution in line with the stakeholder's requirements that improves the existing IdM solutions used during data sharing in logistics. To do so we need to know what an infrastructure of an inter-organizational IdM solution consists of that satisfies the identified requirements. RO4 can be used to identify reusable characteristics and components and provides the improvements that require to be solved. Hence, the fifth research question has been defined:

RQ5: What does the infrastructure of an inter-organizational IdM solution consists of that satisfies all requirements?

1.2.3 Research Scope

The scope of this research consists of designing an inter-organizational IdM solution that improves existing IdM solutions used during data sharing in logistics. It will only focus on the IAA of governmental entities to businesses (G2B). In doing so the variety of legal frameworks of, for example, the United States or the United Kingdom, and the eIDAS Regulation are not taken into account; this requires legal expertise which cannot be acquired due to time constraints. The designed artifact provides the characteristics and components of an IdM solution to improve the existing IdM solutions used during data sharing in logistics. The research does not provide guidelines for the actual implementation of the IdM improvements. Hence, there are restrictions on how the results may be operationalized because no metrics to quantify the improvements are defined in this research. During the research it is not possible to conduct interviews with the intended end users of BDI; the Douane and LSPs and with experts outside of TNO.

1.2.4 Research Methods

The following research methods were used during this research:

- Literature review - method to identify, select, and summarize all existing scientific research studies relevant to the research topic[120].
- Interviews - qualitative research technique to collect data by asking questions to experts of the field of research.

1.2.5 Research Process

To answer the research questions mention in section 1.2.2, the DSRM of Wieringa[131] is applied. The design cycle is used which covers the engineering cycle visible in Figure 1 partially; it excludes the Treatment Implementation and Implementation Evaluation phases. It has a clearly defined cycle and allows to apply agile development. With no

experience regarding IdM it is expected to encounter uncertainties and new problems frequently during the process. Hence, a flexible approach that allows for iteratively looping through the design science cycle is wanted. The design cycle consists of three phases for designing the artifact: Problem Investigation, Treatment Design, and Treatment Validation. The first phase Problem Investigation focuses on the problem investigation examining the stakeholders and their goals, and identifies, describes, explains and evaluates the problem that requires improvement. This phase is essential to motivate the artifact to be designed. In the second phase Treatment Design, the requirements for the artifact to be designed are identified and the process of designing the artifact takes place. In the third and final phase Treatment Validation, the artifact is validated to prove that treating the problem context with the artifact produces the effects that meet the requirements.

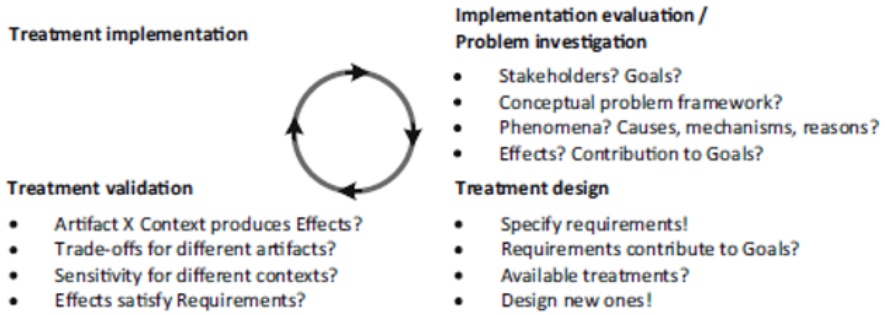

Figure 1: Engineering Cycle Wieringa[131]

The problem investigation is performed based on expert interviews, and a systematic literature review. Semi-structured interviews were conducted with experts #E1 and #E2 (for more information regarding the experts see Table 5) on BDI and data sharing in logistics to understand the problem currently faced with the IdM solutions in order to formulate the problem statement including the context in section 1.2.1. The systematic literature review was conducted to understand the state-of-the-art IdM models and to identify their challenges and causes of those challenges answering RQ1.

In the Treatment Design phase, interviews with experts of TNO (for more information regarding the experts see Table 5) were conducted to specify the requirements as explained in section 3.2 answering RQ2. A comparison between the state-of-the-art IdM models was made based on the requirements to answer RQ3. Based on the comparison a conclusion can be drawn on the state-of-the-art IdM model that should be pursued during the design of the artifact to be designed, IdM4DSL. Again interviews with experts of TNO were conducted to collect existing inter-organizational IdM treatments that adopt the best fitting state-of-the-art IdM model, resulting from answering RQ3. The treatments are investigated in descending order of the total number of suggestions they received and the amount of time available. The existing inter-organizational IdM treatments are analyzed and evaluated based on the identified requirements answering SRQ4.1. A comparison is made between the solutions to identify differences, similarities, and shortcomings. Then possibilities to improve the shortcomings are addressed answering SRQ4.2.

The best fitting model is used as the foundation for designing the IdM4DSL and refined by improving the identified shortcomings. The designed IdM will be validated through expert opinions collected during interviews. During the expert interviews, the IdM4DSL is demonstrated to experts to collect the expected effects of the interaction of the artifact with the problem context and to verify if the IdM4DSL satisfies the requirements.

1.3 Outline

This paper is divided into eight chapters, starting with this chapter providing the introduction to the research and the research methodology applied. Chapter 2 covers the state-of-the-art IdM models by conducting a systematic literature review. Chapter 3 introduces the stakeholders, their goals, the requirements of the artifact, and their contribution to the stakeholders' goals. Chapter 4 covers the comparison of the state-of-the-art IdM models. Chapter 5 provides an analysis of the available inter-organizational IdM treatments. Chapter 6 compares the available inter-organizational IdM treatments and explains how they can be improved. Chapter 7 provides a discussion about the results. Lastly,

the research questions are answered and a conclusion about the research is drawn.

Chapter	Research Question	Information Source
Section 2.2	RQ1, SRQ1.1, SRQ1.2	Literature Review
Section 3.2	RQ2	Expert Interviews
Section 4	RQ3	Literature Review
Sections 5.1, 5.3.2, 5.4.2, 5.5.2, 5.6.2, 5.7.2, and 6	RQ4, SRQ4.1	Expert Interviews
Section 6.1	SRQ4.2	Technical Documentation and Expert Interviews
Section 5.5.2, 6, and 6.1	RQ5	Technical Documentation

Table 1: Paper outline

2 State-of-the-art IdM Models

This chapter covers the state-of-the-art IdM models by performing a qualitative systematic literature review. First, the literature review procedure applied for the systematic literature review is explained after which the findings are presented, and research gaps are identified.

2.1 Literature Review Procedure

The research strategy is inspired by the guidelines of Kitchenham et al. [76]. The literature is obtained from the most relevant accessible University of Twente databases for our scope: Google Scholar, IEEE Xplore, Scopus, and Web of Science. The other databases were focused on other areas such as chemistry, and healthcare contained numerical data or resulted in 0 documents found.

Multiple iterations were performed to determine search strings that resulted in accurate and relevant papers and resulted in three search strings:

1. **("Access Management" OR "Identity Management" OR "Identity and Access Management") AND ("Logistics" OR "Supply Chain" OR "Transport" OR "Industry 4.0" OR "Smart Logistics")**
2. **("Methods" AND "identity" AND "management")**
3. **("Challenges" AND "identity" AND "management")**

Other search strings such as "identification AND authentication, AND authorization AND review", and "identity and access management AND review" resulted in a low accuracy in terms of relevancy.

Applying the first search string among the four databases resulted in 271 papers. To narrow down the number of papers and receive more accurate papers tailored to the needs, the following screening process was conducted, sequentially:

1. A filter on publication years (2000-2022), papers that are openly available (open access), and are written in English
2. A filter on document types: article documents and conference papers
3. Removing duplicates among the four databases
4. Reviewing the abstract and title and assess whether they are useful for answering the research questions

An insufficient number of papers were collected. Hence a second method was applied, backward snowballing based on Wohlin's[132] guidelines. New papers were identified based on the reference list of the papers that resulted from the screening process. For the referenced papers the same screening process was followed. The process of the first search string is shown in Figure 2 and resulted in 4 papers. No useful references were found.

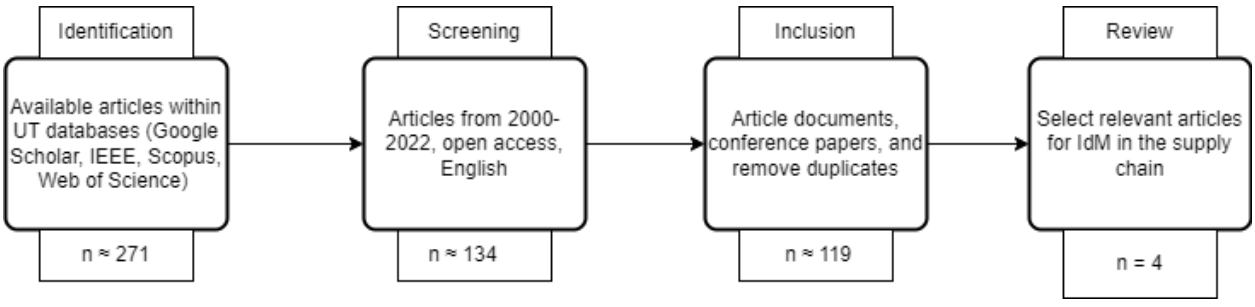

Figure 2: Article selection process search string 1

Applying the second and third search strings among the four databases resulted in 12292 and 8997 papers respectively. Again, the above-mentioned screening process was applied to narrow down the number of papers and receive more accurate papers tailored to the needs. These search strings however resulted in a significant amount of papers with low relevance (456 and 506 respectively). Therefore, an additional step in the screening process was applied: a filter on the subject: computer science (information systems). This process resulted in an insufficient number of papers: 2 papers from the second search string, and 4 from the third search string. Therefore, the backward snowballing method was again used to find additional papers. For the second search string, 2 additional papers were found. For the third search string, 3 additional papers were found.

The process of the second search string is shown in Figure 3 and resulted in 5 papers. The process of the third search string is shown in Figure 4 and resulted in 7 papers. This adds up to a total of 15 papers from the search strings. However, additional online sources are used referenced by TNO and/or used for explanatory purposes.

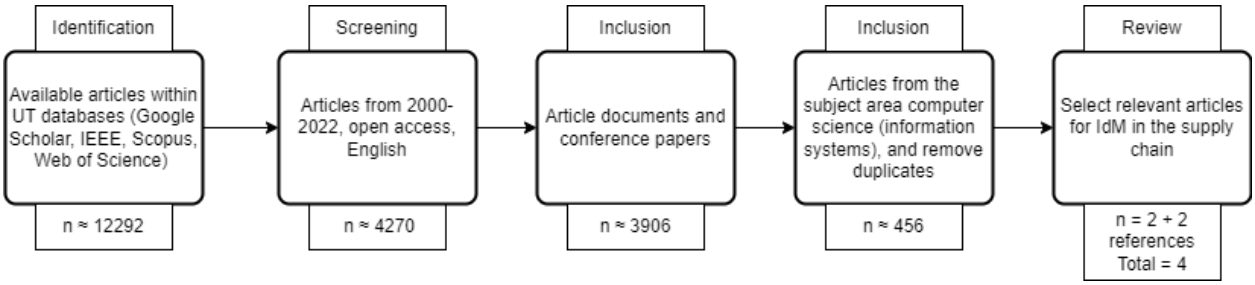

Figure 3: Article selection process search string 2

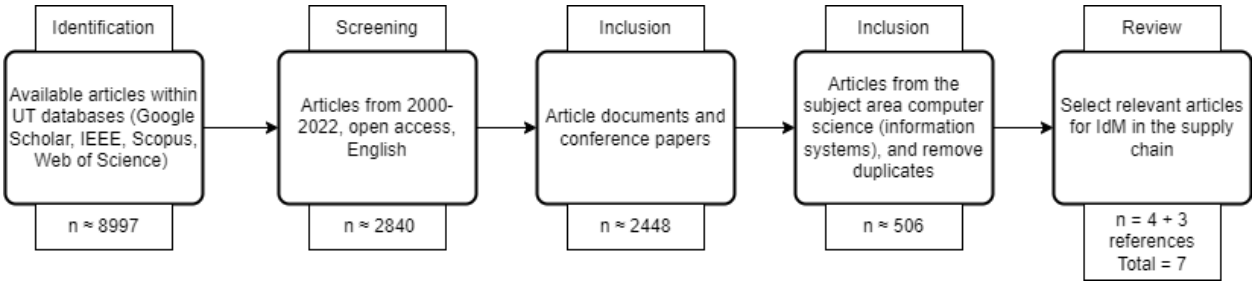

Figure 4: Article selection process search string 3

The papers resulting from the search strategy are shown in Table 2 and discussed in the next section(s).

Reference	Search string	Title
Aldosary and Alqahtani.[1] 2021	3	Survey on Federated Identity Management Systems Limitation and Solution
Smith et al.[119] 2011	3	The identity management challenge
Guo and Wang[53] 2008	1	Application of federated identity management in ERP system
Jensen[74] 2011	3	Benefits of federated identity management - A survey from an integrated operations viewpoint
Cao and Yang[13] 2010	2	A survey of Identity Management technology
Bouras et al.[10] 2020	1	Distributed Ledger Technology for eHealth Identity Privacy: State of The Art and Future Perspective
Smith[118] 2008	3	The challenge of federated identity management
Jensen[75] 2012	3	Federated Identity Management Challenges
Balasubramaniam et al.[6] 2009	2	Identity management and its impact on federation in a system-of-systems context
Maler and Reed[85] 2008	3	The Venn of Identity: Options and Issues in Federated Identity Management
Bertino and Shang[8] 2009	2	Keynote 2: Digital Identity Protection - Concepts and Issues
Frederiksen et al.[41] 2020	3	Identity management: State of the art, challenges and perspectives
Bartolomeu et al.[7] 2019	1	Self-Sovereing Identity: Use-cases, Technologies, and Challenges for Industrial IoT
Cocco et al.[15] 2021	1	Blockchain and Self Sovereign Identity to Support Quality in the Food Supply Chain
Gilani et al.[46] 2022	2	Self-sovereign Identity Management Framework using Smart Contracts

Table 2: Identified papers

2.2 Literature Review Findings

This section addresses **Research Objective 1 (RO1)**: Knowledge regarding IdM must be acquired to get an understanding of the state-of-the-art, challenges, and causes of those challenges. Having this knowledge enhances communication with stakeholders and experts, and aids in designing an accurate inter-organizational IdM solution to achieve the primary research objective. Hence, the first research question is defined with two sub-research questions:

RQ1: What are the challenges posed by the state-of-the art IdM models?
SRQ1.1: What are the state-of-the-art IdM models?
SRQ1.2: What are the challenges experienced regarding the IdM models and what is causing them?

IdM primarily focuses on IAA[1][74]. Five models of IdM were identified[10][13][74]: Isolated Identity Management (IIM), Centralized Identity Management (CIM), Federated Identity Management (FIM), user-centric, and Decentralized Identity Management (DIM). The user-centric model aims to provide users with the ability to store their identity information in one place and control who can access what attributes of their identity information. The DIM model has the same goal and according to Bouras et al.[10], the user-centric model failed because users did not control who could access what attributes of their identity information. The organizations where their identity information was stored had access to their identity information and could decide to delete a user's account. Hence, the user-centric model is not further explored.

The four IdM models IIM, CIM, FIM, and DIM, their challenges, and the causes of these challenges are explained in the next sections.

2.2.1 Isolated Identity Management

IIM is the oldest model where the Service Provider (SP) performs IAA, and the deletion and modification of identities. The storage of identity information and operations is done by the SP[13].

When a user requests access to the service, he is shown a login page requiring the user to provide his username-password combination. The first step in the process is identification which is done via the username. Authentication takes place when the user provides a password that matches the username and clicks on the login button. The SP checks whether the username-password combination is similar to the one that is known in its central database where the username and a hash of the password are stored. If authentication was successful, the SP checks which resources

and services the user may and may not use and grants access accordingly. The authorization policies of a user are determined by a central authority of the SP, for example, the administrator, and is also centrally stored in a database of the SP. Inherited from this model is that each user has a separate username-password combination for each SP.

IIM comes with several challenges:

- Keeping the identity and authorization information up-to-date for each service a company provides is costly and time consuming[53].
- Users have no control over their identity information; it is managed by the SP[46].
- Establishing trust between the user and SP; the user does not know how the SP manages their identity information, whether this is done in a safe and secure way, and in compliance with GDPR[117].
- An account can be deactivated or deleted whenever the centralized party wants[10]. Deleting an account means erasing someone's online identity which he may have worked years on to build and may be of substantial value to him, and irreplaceable[35].
- Users have to manage many username-password combinations, increasing the probability of them getting hacked leading to higher security risks[46].
- Systems are vulnerable to cyberattacks because they are a single point of failure (all identity information is stored centrally) making them an attractive target for hackers[7].

2.2.2 Centralized Identity Management

CIM delegates functions of the SP to the Identity Provider (IdP). The IdP takes care of the IAA, and the deletion and modification of identities. When a user requests access to the service, he is shown a login page requiring the user to provide his username-password combination. The first step in the process is identification which is done via the username. Authentication takes place when the user provides a password that matches the username and clicks on the login button. All user authentication requests at the Service Providers (SPs) are sent to the (for a specific domain) global unique IdP that manages the user identity information[10][13]. The IdP checks whether the username-password combination is similar to the one that is known in its central database where the username and a hash of the password are stored. If authentication was successful, the IdP checks which resources and services the user may and may not use and grants access accordingly. The authorization policies of a user are determined by a central authority of the IdP, for example, the administrator, and are also centrally stored in a database of the IdP. The IdP provides the user with a security token with which he can access the services of the domain he is authorized to access. A domain can only be owned by one organization. Therefore, the distinction between CIM and FIM in this paper will be defined as CIM enabling Single Sign-on (SSO) for accessing services of one domain and FIM enabling SSO for accessing services cross domains.

This explanation is different from Jensen[74] "User authentication is performed by this central entity, which issues identity assertion upon a successful authentication process. The assertion, or security tokens, can then be used to access distributed services across company borders." But similar to Cao and Yang[13] saying "The centralized model requires all users in the same domain, but users always need cross domain or network access. It can't support user privilege delegation and cross domain access well."

CIM comes with several challenges:

- Creates communication overhead and increases the likelihood of a malicious attack[7].
- Systems are vulnerable to cyberattacks because they are a single point of failure (all identity information is stored centrally) making them an attractive target for hackers[7].
- Users have to manage many username-password combinations increasing the probability of them getting hacked leading to higher security risks[46].
- Users have no control over their data; it is managed by the IdP[46].
- Establishing trust between the user and IdP; the user does not know how the IdP manages their identity information, whether this is done in a safe and secure way, and in compliance with GDPR[117].
- An account can be deactivated or deleted whenever the centralized party wants[10] Deleting an account means erasing someone's online identity which he may have worked years on to build and may be of substantial value to him, and irreplaceable[35].

2.2.3 Federated Identity Management

FIM extends CIM by linking different user identifiers within a federation and allowing user authentication across SPs in the federation using the same username-password combination (i.e., federated identity: a single and consistent identity that is usable across platforms, applications, and networks)[10].

Organizations make agreements among themselves to grant access to each other's users. When a user is authenticated by the IdP of the federation, he can access all services of the federation that he is authorized to access[13]. For example, Facebook offers a solution that allows users to log in to other service providers with a social login button. Governments also offer their citizens similar solutions for their services, such as DigiD. In doing so, the organizations trust each other to have their management of identities and accounts in order and therefore give these users access to the services. This usually involves the use of a trust framework that makes data exchange both technically and legally possible.

The IdP is focused on maintaining the federated identity by protecting identity information and making them available to disparate directory services through translation services[81]. An IdP manages the identity information of all users and can be used by multiple partner domains to ensure that their users can use the same username-password combination to access their resources and services. Reducing the amount of username-password combinations a user has to manage; they still have to manage multiple username-password combinations, one for each federation. FIM allows SPs to focus entirely on delivering the service[100].

FIM comes with several challenges:

- Determining who is legally obliged and responsible when something related to the FIM process goes wrong[6][75][118].
- Being compliant with both internal rules and regulations and of partners[6][75][118].
- Trust between partners of the federation[75] caused by:
  - Unclear data ownership agreements resulting in inappropriate use of data and misunderstandings[6].
  - No assurance that the FIM process is controlled and processes are followed sufficiently[75] (the assurance is a challenge).
- Trust between identity providers[6].
- Establishing trust between the user and IdP; the user does not know how the IdP manages their identity information, whether this is done in a safe and secure way, and in compliance with GDPR[1][6][75].
- Compromising one identity provides access to the resources of the whole federation[1][8][75].
- Content-based verification of identity attributes exposing unnecessary user data compromising user's privacy[8].
- Identity providers are single points of failure. In case a IdP gets hacked, the identity information of all federated partners is exposed[41].
- The ability for identity providers to survey users' login activity across multiple sites and use that to learn about citizens' habits and customs[41].
- Users have no control over their data; it is managed by the IdP[46].
- An account can be deactivated or deleted whenever the centralized party wants[10]. Deleting an account means erasing someone's online identity which he may have worked years on to build and may be of substantial value to him, and irreplaceable[35].

2.2.4 Decentralized Identity Management

DIM aims to give users full control over their identity; the user owns his identity, has control over where it is kept, and decides for himself with whom he shares that identity (or parts of it). There are no parties other than the person himself who has exclusive control over the exchange of identity information and related data. The person himself takes the necessary information and decides whether or not to provide it to an SP. Similar to the physical world: there is an agency that issues passports or driver's licenses, but then the document can be used independently[94]. According to Bouras et al.[10] and the paper they referred to from Dunphy and Petitcolas[32], there are two types of DIM models: Self-sovereign Identity (SSI) and decentralized trusted identity.

SSI gives users full control over their identity information and is independent of any SP. A person has a digital wallet. In it, the person can keep digital passes that the person has received from other organizations to prove something such as proof of membership, citizenship, or skill. These passes (or the data on them) are often called credentials: a piece of evidence that declares something about an entity that can be used as evidence. The person can show these credentials when, for example, an organization asks for them when purchasing a service. A person keeping credentials in their digital wallet is in the SSI model known as the holder. A holder of credentials can also be a service, a device, or an organization. An organization that provides credentials to a holder is in the SSI model known as the issuer. An issuer provides credentials, for example, the municipality providing driver licenses. All citizens require to be registered at their municipality and the municipality has a well-defined process to make sure that the identity information provided during the registration is valid. Therefore, driver's licenses are trusted when it is issued by a municipality. A driver's license is of course full of authenticity features, but only a limited portion of them can be checked in real-time. The credentials on the other hand can be verified in real-time and in an undeniable way. It allows verifying whether it was actually issued by a trusted issuer and whether that credential belongs to the holder showing it. Hence, these credentials are known as verifiable credentials and create a basis for trust in a digital world. An organization that asks for a credential when a holder wants to purchase a service, is known as the verifier. A verifier can authenticate the holder based on the credential shown without having to contact the issuer. In the paragraph that follows, it is explained how this is achieved.

SSI knows three concepts: Decentralized Identifiers (DIDs), Decentralized Identifier (DID) documents, and verifiable credentials. A DID acts as a permanent identifier that never changes and represents an entity known as the subject. The DID is a replacement for the username. A DID refers to a DID document that contains information about the subject such as creation time, the public key, method on how to resolve the DID to receive the associated DID document, and more. The DID document is immutable, persistent, and fully owned and controlled by the DID owner[4]. The DID and DID documents form the basis for authenticating a holder based on the verifiable credential shown. The verifiable credential includes, among others, digital signatures of both the issuer and the holder, the DID of both the issuer and holder, and identity information about the holder. A verifier can authenticate a verifiable credential by verifying the digital signatures with which the verifiable credential is signed. To do so, the verifier needs the public keys of both the issuer and the holder which are stored in their DID documents. The verifier can use the verifiable data registry to receive the required information. A verifiable data registry is a decentralized registry that maintains identifiers and schemas required for the verification of a credential written to by the issuer[128]. In order to receive the DID document that a DID refers to, a DID resolver is required: a service that is able to look up and return the associated DID document for a given DID from the verifiable data registry. A DID resolver should be selected that supports the DID method. The verifier looks up the DID in the verifiable data registry and uses the DID resolver to receive the associated DID document. The public key can be used to verify the digital signature with which the issuer signed the verifiable credential shown by the data consumer. In case the public key is the one linked to the private key with which the verifiable credential is signed, it results in a hash value that matches the original hash value meaning the signature has been verified and the credential is authentic. To enable the verifier to prove the integrity of the message and prevent the occurrence of a replay attack, the holder also digitally signs the verifiable credential with its private key. The same process for verifying the digital signature of the issuer is followed but now the data consumer's DID is used. The verifier trusts the verifiable credential shown by the holder because it trusts the issuer's process to validate the holder's identity information used to issue the verifiable credential[10][11][44][45][94][127].

The verifier does not require to store identity information. As a result, less user identity information is exposed when enterprises are hacked significantly reducing costs (administrative, compliance, and operational)[45].

Solutions using verifiable credentials offer a number of features that are distinctive from the physical credentials[94]:

1. They are digital native and thus can be applied in situations where physical credentials are inconvenient or unusable.
2. The model is designed from the privacy-by-design perspective. The organization requesting proof will only receive proof for that specific question asked without additional data that unnecessarily reveals more about someone's identity. For example, it is indicated that a person is over eighteen, but without showing the date of birth. The credential just shows "yes this person is over eighteen years old" or "no this person is not over eighteen years old". In the physical world, this is not possible with the passport: to prove that you are over eighteen, you have to show the date of birth on a document on which also a lot of other data can be seen.
3. Direct digital verifiability. When answering the question "Are you over eighteen?", information about the party that issued the credential (issuer) is included along with the digital signature of that party. This allows the

[4]the subject of a DID does not have to be the DID owner

SP, to verify whether the credential was actually issued by that issuer without consulting the issuer of the certificate. The issuer does not have the possibility to build up a profile of the person showing the credential (what services and resources that person requested access to or accessed).

According to Bouras et al.[10], Gataca[45], Griffith[48], and Leijnse and Scheers[94] most current SSI implementations make use of DLT. A distributed ledger is a database that is widely accessible and cooperatively shared across numerous locations, organizations, or geographies. Before a transaction is added to the ledger it is verified by the nodes of the network based on the common set of policies. If at least 51% of the nodes agree on accepting the transaction, the transaction is written to the ledger. All participants in the network are able to run a node making the solution decentralized[12][66].

Information about the decentralized trusted identity model is limited. In the literature found, it was only mentioned by Bouras et al.[10] and the paper they referred to from Dunphy and Petitcolas[32]. Although Goodell and Aste[47] propose a decentralized digital identity architecture using DLT that looks similar to the decentralized identity model, it cannot be proven that this architecture is the same as the model referred to by Bouras et al.[10], and Dunphy and Petitcolas[32]. The decentralized trusted identity model uses a centralized IdP that issues an identity and performs the authentication based on trusted credentials such as passports or driver's licenses. DLT is used to store the identity attestations (claims about a user's identity) such that third-party services and trusted organizations are able to validate them at a later point in time. It is different from SSI because it does not use the concept of a verifiable data registry. While all of the user credentials are encrypted and stored on his phone, the IdP provides the receiving entity (the entity with which the user shares his credential) with a testimonial of the authenticity of the credential. With SSI, the verification of the authenticity of the verifiable credential is done without the need for communication with the issuer (or IdP as it is called in the decentralized trusted identity model).

DIM comes with several challenges:

- No standard regarding what identity data should be shared in, on, or off the distributed ledger would result in a free-form submission of data on the ledger. Resulting in large transactions which may impact the performance of the system[10].
- Scalability; SSI is in its infancy and because of the lack of widespread adoption there is limited information available on the impact on scalability that billions of working devices will have[7][10].
- IdM consensus protocol; many consensus protocols are focused on cryptocurrencies and preventing double-spending. IdM may not face the issues such as the double-spending problem and therefore the current consensus protocols may not provide the performance and solution to support IdM[10]. The most important element in DLT is the consensus protocol. It controls and maintains the ledger without the need for a central authority.
- Organizational resistance to change has a hindering effect on adoption[7].
- Trust between parties: how can end-users, as well as wallets, be authentically proved and trusted so that verifiers can trust the outcome?[84]. Using a specific system or technology does not inherit trust; a party is autonomous in who or what he trusts.
- Users being in control over their identity information can still be questioned: a verifier can store the identity information shared by the holder and a holder is still reliant on an issuer providing him with credentials.
- Bad user experience has a hindering effect on adoption[84].
- Guaranteeing GDPR requirements: the data on the ledger cannot be deleted or modified at a later date.
- Interoperability between wallets, networks, trust frameworks, and legacy systems; no final standard for decentralized identities, solutions make use of different standards and consensus mechanisms (in case blockchain is used)[22][84].
- Dealing with identity changes; ledgers have the characteristic of being immutable[79].
- Dealing with compromised identity wallets; there are no support services that a user can contact when his wallet is compromised[79].
- Identity recovery; there are no support services where a user can recover his lost password. The user is responsible for keeping his password.

2.2.5 Research Gaps

In the literature found, several definitions of IdM are used. Sometimes, IdM is referred to as Identity and Access Management (IAM) whereas others make a distinction between the two. Our observation is that IdM is not just the technology and policies regarding the IAA of an entity or party and deciding whether that entity or party is allowed to have access to resources such as applications, systems, data, or networks. It is also responsible for access control which is the implementation of authorization decisions through technical and organizational measures. Nevertheless, it indicates the first and second research gap: *the need for a single definition of IdM that can be used across all sectors, so that when IdM is referred to, everyone has the same meaning* and *the need for a clear distinction between IdM and IAM.*

Also, the literature does not provide a paper focused on the paradigm IdM which might be the reason for not having one single IdM definition and papers having different distinctions between the IdM models. For example, Jensen[74] uses a different distinction between CIM and FIM than Cao and Yang[13]. This indicates the third research gap: *the need for a paper that focuses specifically on the evolution of IdM including the technical implementation and clearly defined distinctions between the models.* This paper could also include use cases, benefits, and challenges so that it can be used by others to determine what model is applicable to their use case.

DIM is also unexplored in terms of the general concept, and the two models SSI and decentralized trusted identity. There is no uniform definition of SSI and no papers were found solely focused on the concepts of both models. Especially decentralized trusted identity is unexplored. Only two papers were found briefly mentioning this model. Although Goodell and Aste[47] propose a decentralized digital identity architecture using DLT that looks similar to the decentralized identity model, it cannot be proven that this architecture is the same as the model referred to by Bouras et al.[10], and Dunphy and Petitcolas[32]. If the decentralized digital identity architecture is similar to the decentralized trusted identity model then it is promising and should be explored further. This indicates the fourth research gap: *the need for a paper that focuses specifically on the DIM model explaining the paradigm, explaining the possible model types (e.g., SSI, decentralized trusted identity), and the technology used to realize them.*

According to Bouras et al. [10], Gataca[45], Griffith[48], and Leijnse and Scheers[94] most current SSI implementations make use of DLT. However, DIM does not necessarily have to use DLT; any type of decentralized data storage system works. How DIM can be implemented using decentralized data storage systems apart from DLT is an area that could be further explored. This indicates the fifth research gap: *the need for a paper focused on implementing DIM by using decentralized data storage systems apart from DLT, providing use cases and implementation examples.*

Leijnse and Scheers[94] state that verifiable credentials create a basis for trust in a digital world. A party is autonomous in who or what he trusts making it difficult and complex to prove whether SSI establishes trust or enhances trust between parties. This indicates the sixth research gap: *the need to investigate the effect of SSI on trust between cooperating parties or parties that want to cooperate.*

To conclude, this research provides an overview of the state-of-the-art IdM models with the challenges that come with each of them. It can be used by others to assess which IdM model is relevant for their use case or to use as a basis for in-depth research about the evolution of IdM models or one specific model.

3 Requirements Inter-organizational IdM Solution

The topic of this study is the artifact in a particular context. Designing and studying this artifact in context to address a design problem in that context are the two main tasks. The artifact in this study is an inter-organizational IdM that guides the IAA of governmental entities to LSPs, referred to as IdM4DSL. The problem context is discussed in section 1.2.1. The actors affected by the inter-organizational IdM are known as the stakeholders. Different designs could be made for an artifact that addresses the introduced problem context, but stakeholders' goals evaluate the usability. Hence, this chapter introduces the stakeholders, their goals, the requirements of the artifact, and their contribution to the stakeholders' goals.

3.1 Stakeholders and Goals

The stakeholders of a problem are the actors affected by treating the problem. They are the source of goals, the constraints of this research, and are the source of the requirements for the IdM4DSL.

IdM4DSL is designed to support the main stakeholders, Douane and LSPs (i.e., normal operators), to accurately perform IAA during the data sharing process in logistics. As a result, their processes are optimized contributing to their goals shown in Table 3. During the research, it is not possible to conduct interviews with the normal operators (i.e., end users) to collect their objectives, constraints, and requirements. But the sponsor of this research TNO DE has experts that are working on BDI and are representatives of the intended end users. TNO DE is the sponsor of this research and is working on a proof of concept of BDI where the IdM4DSL is designed for. The author is the researcher and designer of IdM4DSL. Table 3 sums up the possible stakeholders involved in this research, their classification based on Alexander's Taxonomy[131], and their goals.

Alexander's Taxonomy[131]	**Stakeholder**	**Goals**
Normal operators	Douane	Collect import duties and taxes on time and correctly, protecting society from unsafe and unwanted goods, and strengthening the competitive position of the Netherlands and the EU.
Normal operators	LSPs	Transfer the resources and products in the most efficient and effective way possible from the point of origin to the destination and the customer.
Sponsor	TNO DE	Help organizations develop and set up data ecosystems.
Researcher, developer	Author	Research and design an inter-organizational IdM that satisfies the requirements and objectives of the main stakeholder.
Consultants	#E1-#E8	Transfer of domain knowledge and/or practical proof of concepts and/or practical implementations.

Table 3: Possible stakeholders of the be designed inter-organizational IdM, based on Alexander's Taxonomy

IdM4DSL aims to provide a guideline for the application of an inter-organizational IdM treatment during data sharing in logistics to reduce the number of credentials a user has to manage[46], remove the single point of failure[7][41], and create trust between the data sharing parties[1][6][75][84][117].

3.2 Requirements

Semi-structured interviews were conducted with experts of TNO (mentioned in Table 5) to collect all requirements, their meanings, and gain insight into the reasoning behind them. A semi-structured interview is a type of interview in which the interviewer has a general idea of what topics will be covered, but the exact questions asked and the order in which they are asked may vary. This allows the interviewer to follow the flow of the conversation and explore interesting topics or areas of the interviewee's background in more detail. Unlike in a structured interview, where the questions are predetermined and the same for all candidates, a semi-structured interview allows for more flexibility and creativity in the conversation[20][30]. Semi-structured interviews are useful in case the stakeholder does not have a clear understanding of the requirements which is the case in this research.

The semi-structured interviews for defining the terminologies (i.e. semantics) of the requirements have a very specific

approach worthy of explaining. These interviews were set out in three phases. The first phase consists of picking a requirement and defining criteria: a standard by which something may be judged or decided. The criteria will be used to judge whether an IdM model, solution, or IdM4DSL satisfies the requirement or not. In the second phase, the experts come up with a use case to which they can apply the criteria. Before applying the criteria, the experts determine the intended outcome, for example, "yes, we consider Bitcoin as decentralized. Thus when applying the criteria Bitcoin should be judged as satisfying the requirement of decentralized". In the third phase, the criteria are applied to the use case and judged resulting in either the intended outcome or not. When the outcome is as intended, a new use case will be defined and the process is repeated. When the outcome is not as intended, the criteria are redefined and the process is repeated for the same use case. Once all use cases that the experts could come up with resulted in the intended outcome, the criteria are used in order to define the requirement. This process is repeated for all requirements and is visual in Figure 5.

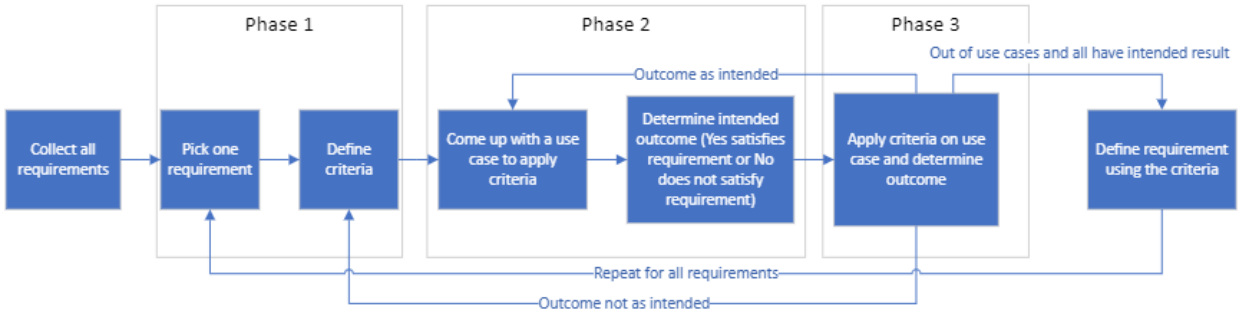

Figure 5: Defining the Requirements

The requirements are classified based on their priority using the MoSCoW method, which is commonly used in software development, product development, and business analysis. MoSCoW stands for[77][97]:

- M - Must have: mandatory requirement to be realized
- S - Should have: high priority that should actually be realized
- C - Could have: can be realized if there is time to spare
- W - Would have: no priority, may be realized in the future
- The lowercase letters 'o' in the abbreviation have no meaning

An analysis was conducted for ranking the requirements based on the priority it has for being part of IdM4DSL. Prioritizing the requirements aids in decision-making about what parts of research should be focused on first and the maintenance of the research. The MoSCoW method is easy to apply, understand, and easy to coordinate with the stakeholders. The prioritization of the requirements is verified by experts #E1 and #E2.

Seven experts (#E1-#E7) were interviewed in order to define the requirements. These experts were selected based on their expertise and vision. Table 5 shows the seven experts that were interviewed, with their identifiers to simplify referencing and their roles explaining their expertise, vision, and reason for interviewing them. All experts are part of the DE department and are experts on data spaces.

In total nine interviews were held with expert #E1 of which eight were also attended by expert #E2. These nine interviews focused on defining the requirements and their terminologies. In between those interviews, feedback was collected during interviews with experts #E3, #E4, #E5, #E6, and #E7 and processed accordingly. The former three experts were interviewed twice taking up to 2 hours each. The latter two experts were interviewed once taking up to 90 minutes each. The full description of the process of defining the requirements can be read in appendix B. The resulting requirements are defined in Table 4. Each requirement has its own identifier to simplify referencing. In the text that follows, bracket notation is used to refer to the requirement (e.g. (#RQ1), (#RQ7)).

Identifier	Priority	Requirement
#RQ1	**Must**	**Issuer Verification**: a mechanism, system, or infrastructure should be available that allows a data provider (i.e., LSP) to verify the issuer of a credential shown by a data consumer (i.e., an agent of Douane).
#RQ2	**Must**	**Open Architecture**: the architecture of the IdM should be openly available to anyone who wants to use the IdM. The architecture should include technical specifications or guidelines of the IdM, allowing developers to build their own implementations.
#RQ3	**Must**	**Proof of Request**: a data provider (i.e., LSP) must be able to prove that a user requested access to a resource at a specific time. In the event that a data consumer (i.e., an agent of Douane) submits an unauthorized request for data, the data provider is able to prove this event to the principal of the data consumer.
#RQ4	**Must**	**Reusable Credentials**: if party A receives a credential from party B, this credential can also be used for identification at other parties such as party C, D, and E. This credential can be reused at other parties because parties C, D, and E trust the issuer of this credential, namely, party B.
#RQ5	**Should**	**Decentralized**: more than one party can control a specific resource and the policies that govern the system's operations. If there are two parties in the data space, they should both supervise the resources and have a vote in the determination of policies for the functioning of the system. More parties mean a higher level of decentralization. The resources of the IdM that should be decentralized: credential issuance, the service providing information regarding members of a data space including the services that they offer, and granting access to a data space. Policies of the IdM include credentials, authentication, access control, data privacy, and security policies.
#RQ6	**Should**	**Open Data Space**: the data space is open for participation by anyone complying with its policies. The compliance assessment must be assessed objectively.
#RQ7	**Could**	**Authorization Policies**: during the process of assigning access rules to authenticated entities, a data provider is able to self-configure authorization policies and assign them to a data consumer regarding data accessibility. Authorization policies could, for example, be provided by RBAC.

Table 4: Requirements

3.3 Contribution Arguments

To justify the requirements defined in Table 4 contribution arguments are given, explaining that an artifact that satisfies the requirements would contribute to a stakeholder goal in the problem context. The template defined by Wieringa[131] is used to explain the contribution arguments: (Artifact Requirement) x (Context Assumption) contribute to (Stakeholder Goal)

If the designed IdM treatment satisfies the requirement of Issuer Verification (#RQ1) assuming the IdM is integrated into the BDI and the ability to verify an issuer of a credential increases a party's level of trust then the designed artifact contributes to the stakeholder's goal to create trust.

If the designed IdM treatment satisfies the requirement of Open Architecture (#RQ2) and assuming the specifications are technology agnostic then the designed artifact contributes to the stakeholder's goal of having an open IdM solution and one that is flexible in the use of standards.

If the designed IdM treatment satisfies the requirement of Proof of Request (#RQ3) assuming that the proof is logged and accessible in case it is needed then the designed artifact contributes to the stakeholder's goal to create trust.

If the designed IdM treatment satisfies the requirement of Reusable Credentials (#RQ4) and assuming the IdM is integrated into the BDI then the designed artifact contributes to the stakeholder's goal of reducing the number of credentials that a user has to manage.

If the designed IdM treatment satisfies the requirement of Decentralized (#RQ5) and assuming that the resources and policies of the data space can also be controlled by more than one party then the designed artifact contributes to the stakeholder's goal of decentralized and to remove the single point of failure.

If the designed IdM treatment satisfies the requirement of Open Data Space (#RQ6) and assuming the assessment was carried out objectively then the designed artifact contributes to the stakeholder's goal of having an open IdM solution.

If the designed IdM treatment satisfies the requirement of Authorization Policies (#RQ7) assuming access requests are logged, can be used as evidence to prove unauthorized requests, and actions are taken by the principal of the agent submitting unauthorized requests then the designed artifact contributes to stakeholder's goal to create trust.

Identifier	Role
#E1	Medior Scientist: Specialist having the vision to apply decentralized solutions. He is the kickstarter and supervisor of this research. He works on the development of the BDI for which this research was initiated, and is an expert on Blockchain, DLT, Corda, interoperability, Zero-knowledge proof, Hyperledger Fabric, Indy, and Aries
#E2	Medior Scientist: Integrator having the vision to apply decentralized solutions. He is working on the development of the BDI for which this research was initiated, and is an expert on Blockchain and smart contracts with knowledge of Corda and Hyperledger Fabric.
#E3	Researcher and expert on SSI, terminology, ontologies, decentralized working, risk management, conceptual modeling, and mental models.
#E4	Medior Scientist: Integrator who develops TNO's International Data Spaces (IDS) prototypes including the IdM and is involved in Gaia-X that also includes an IdM solution. He is an expert to ask technical questions regarding the functioning of both Gaia-X and IDS.
#E5	Scientist: Innovator who develops parts of the iSHARE and IDS (includes an IdM solution) prototypes. He is an expert to ask technical questions regarding the functioning of both iSHARE and IDS.
#E6	Senior Scientist: Specialist who is an expert on IDS and Gaia-X. He develops parts of the IDS prototypes including the Eclipse Dataspace Connector and knows about the developments around Gaia-X.
#E7	Senior Project manager, Solution Architect, and subject matter expert on SSI (project such as IRMA[a]).
#E8	Senior Scientist Media Networking who is expert on Blockchain, IdM, SSI, and the Sovrin Network (Sovrin). He is part of the Networks department.

Table 5: Interviewed experts

[a] a digital wallet for storing and managing identity information. With this wallet, you can share information with municipalities while protecting your privacy *source: Privacy by Design[28]*

4 Comparing IdM Models

Research Objective 3 (RO3): Knowing what state-of-the-art IdM model best fits the requirements of the research to have guidance about the IdM model that should be pursued during the design of the artifact. Hence, the third research question is defined:

RQ3: What state-of-the-art IdM model does best fit the requirements?

To determine what state-of-the-art IdM model best fit the requirements, the requirements are evaluated based on the literature found in the systematic literature study and additional online sources. The results are shown in Table 6. Information about the decentralized trusted identity model mentioned by Bouras et al.[10] and Dunphy and Petitcolas[32] are too limited to use in the evaluation of the DIM model. Therefore, the evaluation of the DIM model is based on the literature found about the SSI model.

4.1 IIM

Users are authenticated based on a username-password combination. These do not qualify as credentials according to the applied definition, see appendix A (#RQ1). An IIM solution may use open standards but the technical specifications of the whole solution allowing developers to build their own implementation were not found (#RQ2). The SP typically maintains its own logs of the users that requested access to its resources, as well as which user was granted or denied access. These logs could be used to prove a user requested access to a resource (#RQ3). No verification is done regarding the issuer of a username-password combination. The SP issues the username-password combination itself, and stores the username and a hash of the password in its database. It checks whether the username-password combination provided by a user matches with what is known in its database (#RQ4). The SP, being one party, is in full control over the resources and policies of the IdM (#RQ5). Solutions following the IIM model are used for a specific party, not a data space (#RQ6). The SP determines and configures the authorization policies (#RQ7).

4.2 CIM

Users are authenticated based on a username-password combination. These do not qualify as credentials according to the applied definition, see appendix A (#RQ1). An CIM solution may use open standards but the technical specifications of the whole solution allowing developers to build their own implementation were not found (#RQ2). The IdP typically maintains its own logs of the users that requested access to its resources, as well as which user was granted or denied access. These logs could be used to prove a user requested access to a resource (#RQ3). No verification is done regarding the issuer of a username-password combination. The IdP issues the username-password combination, and stores the username and a hash of the password in its database. It checks whether the username-password combination provided by a user matches with what is known in its database (#RQ4). The IdP is in full control over the resources and policies of the IdM (#RQ5). Solutions following the CIM model are used for a specific party, not a data space (#RQ6). The IdP determines and configures the authorization policies (#RQ7).

4.3 FIM

Users are authenticated based on a username-password combination. These do not qualify as credentials according to the applied definition, see appendix A (#RQ1). There are FIM solutions available that have an open architecture such as the IBM Tivoli Federated Identity Manager (#RQ2). The IdP of the federation (e.g., a data space) typically maintains its own logs of the users that requested access to its resources, as well as which user was granted or denied access. These logs could be used to prove a user requested access to a resource (#RQ3). No verification is done regarding the issuer of a username-password combination. In fact, this is an issue addressed by Maler and Reed[85]. The IdP issues the username-password combination, and stores the username and a hash of the password in its database. It checks whether the username-password combination provided by a user matches with what is known in its database (#RQ4). The IdP controls the resources and policies of the IdM (#RQ5). The parties of the federation (e.g., data space) decide together whether a party can join the federation or not. This decision is

not only based on whether the party complies with the policies set but also based on trustworthiness, compatible technology, legal and business considerations, and may involve subjectivity. But FIM could be applied in a way that all parties complying with the policies set are allowed to join the data space (#RQ6). In FIM, authorization policies are typically determined by the organization that manages the IdP. In some cases, the authorization policies may be flexible and allow for a degree of customization. For example, an organization may allow certain users to have more or less access to specific systems and resources based on their job duties or level of responsibility. In other cases, the policies may be more rigid and apply uniformly to all users (#RQ7).

4.4 DIM

An SP can verify an issuer by searching for the DID in the verifiable data registry and uses the DID resolver to receive the associated DID document. The public key can be used to verify the digital signature with which the issuer signed the verifiable credential shown by the data consumer. In case the public key is the one linked to the private key with which the verifiable credential is signed, it results in a hash value that matches the original hash value meaning the signature has been verified and the credential is authentic (#RQ1). Gaia-X, an example that applies SSI, has an open architecture that provides technical specifications and guidelines of the IdM allowing developers to build their own implementation (#RQ2). Requests of users can be logged; it is possible to hash the request and log it on, for example, a blockchain. The SP can decrypt that hash and use it as evidence to prove a user requested access to its resource or service (#RQ3). A verifiable credential can be reused at as many SPs as the number of SPs that trust the issuer of that credential (#RQ4). Gaia-X is open to participation when a party agrees with the policies of the data space (#RQ6). The authorization policies are set predefined for a specific resource or service and are attribute-based (e.g., if you want to buy beer, you have to show that you are above 18 years old). When a user requests access to a resource or service, the SP determines what verifiable credential (and identity information it should contain) the user must show in order to get authorized (#RQ7).

4.5 Concluding Remarks

Table 6 shows that both the IIM and CIM model do not satisfy the must-have requirements Issuer Verification (#RQ1), Open Architecture (#RQ2), and Reusable Credentials (#RQ4) and the should-have requirements Decentralized (#RQ5) and Open Data Space (#RQ6). Neither one of those models can be used inter-organizational and hence no IdM solution of those models will be explored. The FIM model does not satisfy the must-have requirement Issuer Verification (#RQ1) and the should-have requirement Decentralized (#RQ5). Therefore, no IdM solutions of the FIM model will be explored. The DIM model satisfies all requirements and is the best fitting IdM model for the requirements of this research. As a result, the IdM4DSL will use the DIM model, and only available inter-organizational IdM treatments that use the DIM model will be investigated.

Requirement / IdM model	**Isolated**	**Centralized**	**Federated**	**Decentralized**
#RQ1 Issuer Verification	No	No	No	Yes
#RQ2 Open Architecture	No	No	Yes	Yes
#RQ3 Proof of Request	Yes	Yes	Yes	Yes
#RQ4 Reusable Credentials	No	No	No	Yes
#RQ5 Decentralized	No	No	No	Yes
#RQ6 Open Data Space	No	No	Yes	Yes
#RQ7 Authorization Policies	Yes	Yes	Yes	Yes

Table 6: Identity management model assessment based on requirements

5 Available Inter-organizational IdM Treatments

This section addresses **Research Objective 4 (RO4)**: Knowledge regarding the existing inter-organizational IdM solutions is required, how they satisfy the identified requirements, and how they can be improved. Having this knowledge enhances communication with stakeholders and experts, and aids in designing an accurate inter-organizational IdM solution that satisfies all identified requirements and improves existing IdM solutions used during data sharing in logistics. Hence, the fourth research question is defined with two sub-research questions:

RQ4: What inter-organizational IdM solutions are used during data sharing in logistics and to what extent do these solutions satisfy the identified requirements?
SRQ4.1: How do these IdM solutions satisfy the identified requirements?
SRQ4.2: How can the IdM solutions be improved?

First, the process used to select the inter-organizational IdM treatments is explained. The subsequent sections cover the design and analyses of the investigated inter-organizational IdM treatments.

5.1 Available Treatment Selection Process

Interviews were held with six experts individually in which the problem context and requirements were explained, and was asked what inter-organizational IdM treatments adopting the DIM model should be investigated. Table 7 shows the results, a 1 means that the experts suggested to investigate the respective IdM treatment and 0 if not suggested. The results row shows the sum of the number of suggestions. Experts #E6 and #E8 were not part of the selection process because they were not involved in the research at that point in time. During the research, the IdM treatments were investigated in descending order of the total number of suggestions they received and the amount of time available. The suggestions IDS, Gaia-X, the European Digital Identity Wallet (EUDI Wallet), and Sovrin, were investigated in addition to the pre-determined IdM treatment iSHARE.

Expert	IDS	Gaia-X	Sovrin	EDIW	IRMA
#E1	1	1	1	1	0
#E2	1	1	0	0	0
#E3	1	1	1	1	0
#E4	1	1	0	0	0
#E5	1	1	0	0	0
#E7	1	1	0	1	1
Result	**6**	**6**	**2**	**3**	**1**

Table 7: Suggested IdM Treatments to Investigate

5.2 Available Treatment Analyses Strategy

During the analyses of the inter-organizational IdM treatments, technical documentation available medio 2022 was used and relevant experts were interviewed to fill knowledge gaps. A mix of open and closed empirical knowledge questions was asked during the expert interviews. The questions were tailored towards answering SRQ4.1 and SRQ4.2, and to what extent the treatments satisfy the identified requirements. Table 8 shows per inter-organizational IdM treatment the experts that were interviewed with their role and the number of interviews conducted. Appendix C summarizes the results from the interviews. Due to the limited availability of expert #E3, only one interview has been conducted to fill knowledge gaps about the EUDI Wallet.

Available Treatment	Expert	Role	Number of interviews
iSHARE	#E5	Scientist: Innovator	5
IDS	#E4	Medior Scientist: Integrator	4
IDS	#E5	Scientist: Innovator	4
IDS	#E6	Senior Scientist: Specialist	1
Gaia-X	#E4	Medior Scientist: Integrator	4
EUDI Wallet	#E3	Senior Scientist	1
Sovrin	#E8	Senior Scientist Media Networking	1

Table 8: Interview Details Available Treatments

5.3 iSHARE Trust Framework

The iSHARE Trust Framework is an initiative of the Netherlands' Logistics Top Sector (partially) funded by the Dutch government. The goal of iSHARE is to enable more efficient and effective data sharing between government organizations in the Netherlands, while also protecting the privacy and security of the data involved. To do so, iSHARE specifies a set of standards and technical specifications focused on the governance and trust[1] of participants to enhance data sharing while maintaining data sovereignty. The technical specifications include how the IAA process should be set up in order to create trust between parties that want to exchange data. Instead of creating many bilateral contracts, all participants have to comply with one federated legal framework, and are registered to become digitally verifiable.

5.3.1 Design

ISHARE is based on the OAuth 2.0 protocol that is frequently used as a technique for implementing policy management, based on standard web service calls in the form of APIs, employing access tokens for the authentication of data consumers[56]. In order for a data consumer to retrieve data from a data provider, a two-step process is followed. First, an access token is received from the data provider. Secondly, the data can be retrieved from the data provider using the access token.

Based on this technology, iSHARE currently provides the following trust framework capabilities to support data spaces[55]:

- Participant trusted registration and administration
  When a party wants to become a member of a data space, it goes through an onboarding process. The iSHARE Satellite validates the onboarding process based on the iSHARE and data space policies, and guarantees trustworthy onboarding. Data administrators register participants with a unique ID, X.509 certificate (eIDAS or PKIoverheid) identification and public key, signed Terms of Use (additional terms are possible), and Chamber of Commerce papers to ensure that the contract is legally signed
- Participant discovery and status information
  Each iSHARE Satellite provides APIs that can be used to discover participants of data spaces and to retrieve data regarding specific data space participants.
- Authorization Registry
  An Authorization Registry allows data space participants to manage data access or usage rights to other data space participants. For example, the Authorization Registry allows the manager of Lisa to set specific time frames for data availability, register the data that is shared, and that it is shared with the Douane. Dave is allowed to access the data too because he is an employee of the Douane. Information regarding participants of the network such as their role, website, phone number, and Authorization Registry ID can be received via APIs.

5.3.2 Analysis

The first step in the two-step process followed in order for a data consumer to retrieve data from a data provider starts with the identification of the data consumer. If the data consumer is who he says he is, is authorized to

access the data, and is an active member of the data space, the data provider will provide him an access token. This communication goes via HTTP using the OAuth 2.0 protocol. The request for an access token consists of multiple headers and parameters. One of the parameters is the client assertion, a JSON Web Token (JWT) produced by the client application of the data consumer that qualifies as a credential. The client assertion is presented to the data provider as proof of the client's identity. The client assertion includes six mandatory fields (i.e., the payload): iss (the issuer of client assertion), sub (the subject of the client assertion), aud (the audience for which the client assertion is created), jti (a unique identifier used to prevent reuse of the client assertion), iat (the time when the client assertion was issued), and exp (the time when the client assertion expires). The issuer and subject information come from the X.509 certificate issued by a trusted CA. The identity information contained in the X.509 certificate is verified by the CA. As proof of accurate verification, the CA digitally signs an X.509 certificate with his private key. This creates a chain of trust; because the data provider trusts the CA, he trusts the identity information contained in the data provider's X.509 certificate.

Before sending the request for an access token, the header and payload are encoded and signed with the private key of the data consumer's X.509 certificate using an encryption algorithm ensuring the integrity of the message (i.e., JSON Web Signature (JWS)). JSON Web Encryption (JWE) is used to encrypt the JWS. The header of the request includes the encryption algorithm and the public key corresponding to the private key used to digitally sign the client assertion. When the data provider receives the request, he decodes the request resulting in the header, payload, and signature. The data provider verifies the signature by using the public key and the encryption algorithm located in the header. In case the public key is the one linked to the private key with which the client assertion is signed, it results in a hash value that matches the original hash value meaning the signature has been verified and the message has not been tampered with. When the signature is verified, the data provider verifies the payload by, for example, checking whether the payload complies with the iSHARE specifications, checking whether the client assertion is not received before, checking whether the client assertion is not expired, checking whether the audience contains his EORI number, and checking whether the iss and sub field contain the same EORI number (the one from the data consumer). When all checks have been successful, then the data provider checks for additional information at the iSHARE Scheme Owner, such as whether the EORI number (from the iss and sub field) is an active member in the data space and the company name. This infrastructure ensures the ability of a data provider to verify a credential shown by a data consumer (#RQ1). If the EORI number is an active member of the data space, the data provider issues an access token to the data consumer. Now the data consumer can request access to the data by providing the access token in the header and including the question (e.g. "can I have access to dataset X?") in the payload. The data provider decodes the request, verifies the access token and provides access to the data accordingly.

The technical specifications of iSHARE are openly available[70] and include the governance and technologies that should be used in order to perform the IAA process. These ensure that only authorized entities and parties are provided with access to resources. ISHARE also provides UML diagrams (both for human-to-machine and machine-to-machine interaction) that showcase the process starting with requesting access to a service until receiving access to a service. According to expert #E5, 'an implementation must be built by yourself in case you want to make use of the iSHARE Trust Framework. Some reference implementations are available such as the Authorization Registry[73]. The technical specifications provided by iSHARE are sufficient for developers to build their own implementation' (#RQ2).

A JWT is exchanged between the data consumer and data provider consisting of claims about the data consumer. The JWT is signed with the JWS standard to ensure the integrity of these claims. The JWT is exchanged when a data consumer request access to a resource of the data provider[69]. If the data consumer would store the received JWT, it could be used as proof of request. But iSHARE has nothing specified about storing the JWT to be able to prove that the data consumer requested access to a resource at a specific time or any other service that stores this kind of information (#RQ3).

Regardless of the type of interaction (human-to-machine or machine-to-machine), a client assertion is generated specifically for the data provider that the data consumer wants to interact with (i.e., the aud field is set to a specific data provider). As a result, a client assertion cannot be used at other data providers (#RQ4).

The credentials used by a data consumer are issued by the data consumer itself. The iSHARE Scheme Owner is a central role fulfilled by the iSHARE Foundation and is responsible for the admission of the iSHARE Satellites, and the maintenance of the iSHARE Schema. Every iSHARE participant should be connected to the Scheme Owner via the legal framework (Terms of Use and the corresponding contract). They can contact the Scheme Owner to check participant information such as whether a party is a member of a specific data space. The iSHARE Foundation is a centralized party in control of the policies (e.g. credential, data privacy, and security policies) and specifications.

Each data space has an iSHARE Satellite responsible for the trustworthy onboarding of parties that want to become a member of the data space. It is operated by coordinating organizations[68] (#RQ5).

When a party wants to become a member of a data space, it goes through an onboarding process. The iSHARE Satellite validates the onboarding process based on the iSHARE and data space policies, and guarantees trustworthy onboarding. The iSHARE Satellite agreed to all the rules of iSHARE that are legally binding so it is trusted that their onboarding process is impartial[68] (#RQ6).

ISHARE supports both fine and coarse-grained authorization[95]. The policies are based on XACML 3.0, a standard for representing and evaluating access control policies that support ABAC or in combination with RBAC[33]. These policies are configurable by the data provider for any resource (all, a selection, or a combination). A data provider can, for example, allow the data consumer to access all data fields of a specific dataset but only from the years 2021 and 2022 (#RQ7).

5.4 International Data Spaces

IDS is a research initiative that aims to facilitate the exchange of data between organizations in a secure and standardized manner while guaranteeing data sovereignty for data owners. It does so by leveraging existing standards, technologies, and well-accepted governance models. In IDS data sovereignty is defined as "a natural person's or corporate entity's capability of being entirely self-determined with regard to its data"[92]. For this particular capability and related aspects such as requirements for secure and trusted data exchange in business ecosystems, the IDS initiative proposes a Reference Architecture Model[92]. The initiative focuses on three activities: research activities, standardization activities, and activities for the development of products and solutions for the market. The research activities and development of products and solutions for the market are performed by organizations such as Fraunhofer and TNO. Standardization activities are performed by the International Data Spaces Association (IDSA), consisting of more than 130 member organizations working together to fulfill the shared vision of a secure, trustworthy, and equal world in which all businesses are empowered to choose their own usage policies and realize the full value of their data. Their goal is to establish a global standard for IDS and interfaces, as well as support the associated technologies and business models that will drive the next industrial data economy. The IDS standard can be used by actors in the market to provide software services and technology (in compliance with the IDS standard) that form the operational IDS ecosystem.

5.4.1 Design

The central component of IDS is the IDS Connector that allows data owners and data providers to exchange and share their data with other participants in the IDS ecosystem while ensuring data sovereignty at any time. At the beginning of a connector configuration and provisioning sub-process, the participant has to write a Self-Description for the connector. The Self-Description consists of information about the respective organization, who maintains the connector, and the content and type of data that the connector offers or requests. These Self-Descriptions are published by the Broker services such that other IDS participants can read them. A connector can also deploy a Data App to enrich or transform the data, or to improve the quality, and can be integrated into the data exchange workflow between IDS participants.

To participate in IDS, the participant must have a participant certificate and operate a certified connector. Each connector taking part in IDS needs to have a unique identifier, a valid X.509 certificate, and must be able to verify the identity of other connectors. An IdM solution is required to enable access control-related decisions based on reliable identities and properties of participants. A central role in the IdM is the Identity Provider consisting of a CA, Dynamic Attribute Provisioning Service (DAPS), and a service named Dynamic Trust Monitoring (DTM). The Identity Provider offers a service to create, maintain, manage, and validate identity information of and for IDS participants. A CA manages digital certificates for the IDS participants. The DAPS manages dynamic attributes linked to participants' identities and issues Dynamic Attribute Tokens (DATs). The DTM continuously monitors the security and behavior of the network.

Each connector is provided with an X.509 certificate by a CA and is used for authentication and encryption between connectors. An identity may have several attributes linked to its identity that change over time. Instead of creating a new X.509 certificate for every change made to these attributes, IDS makes use of DATs: a JWT containing signed

dynamic attributes such as issuer, subject, audience, scope, and security profile[4]. Every time a connector (now a data consumer) wants data from a data provider, it requests a Dynamic Attribute Token (DAT) from the DAPS. The DAPS demands identification done with the OAuth 2.0 protocol communicating via an encrypted tunnel (e.g., TLS). When the authentication by the DAPS was successful, the data consumer is provided with the requested DAT which it can use in order to request access to the data of the data provider. The data consumer uses the DAT to identify itself at the data provider. When the authentication by the data provider was successful, the data provider performs authorization. The authorization policies can be negotiated between the data consumer and the data provider. The usage restrictions can be specified with the policy editor Policy Administration Point and created based on the Open Digital Rights Language (ODRL), a standard for representing and exchanging digital rights information and understood by any usage control technology. Once finalized, the authorization policies are defined in a contract forming a unique, binding agreement between the data consumer and data provider. The contract is the foundation for clearing and configuring access control policies. After the data has been exchanged, IDS uses data provenance tracking to find out when, how, and by whom data was modified, and what other data influenced the process of creating new data items. The information collected during the provenance tracking is stored at the Clearing House which also stores data that has been successfully sent or received i.e., the Clearing House allows logging of data transfer information.

5.4.2 Analysis

For a data consumer to receive access to the data of a data provider a two-step process is followed. First, the data consumer identifies himself to the DAPS. If the data consumer is who he says he is, the DAPS provides a DAT. With the DAT, the data consumer can request access to the data of the data provider which is the second step in the process. During this two-step process, the verification of the issuer of the credential takes place.

The whole process starts with the data consumer requesting a DAT from the DAPS using OAuth 2.0 communicating via an encrypted tunnel (e.g., TLS). The request consists of multiple headers and parameters. One of the parameters is the client assertion, a JWT produced by the client application of the data consumer. The client assertion is presented to the DAPS as proof of the client's identity. The client assertion includes six mandatory fields (i.e., the payload): iss (the issuer of client assertion), sub (the subject of the client assertion), aud (the audience for which the client assertion is created), jti (a unique identifier used to prevent reuse of the client assertion), iat (the time when the client assertion was issued), and exp (the time when the client assertion expires). The issuer and subject information come from the X.509 certificate issued by a trusted CA. The identity information contained in the X.509 certificate is verified by the CA. As proof of accurate verification, the CA digitally signs an X.509 certificate with his private key. This creates a chain of trust; because the DAPS trusts the verification of the CA, it trusts the identity information contained in the data provider's X.509 certificate.

Before sending the request, the header, and payload are encoded and signed with the private key of the data consumer's X.509 certificate using an encryption algorithm ensuring the integrity of the message (i.e., JWS). JWE is used to encrypt the JWS. The header of the request includes the encryption algorithm and the public key corresponding to the private key used to digitally sign the client assertion. When the DAPS receives the request, it decodes the request resulting in the header, payload, and signature. The DAPS verifies the signature by using the public key and the encryption algorithm located in the header. In case the public key is the one linked to the private key with which the client assertion is signed, it results in a hash value that matches the original hash value meaning the signature has been verified and the message has not been tampered with. When the signature is verified, the DAPS verifies the payload by, for example, checking whether the credential is not expired. When all checks have been successful, the DAPS issues a DAT to the data consumer.

The DAT is the credential with which the data consumer requests access to the data provider hence, is the credential that requires issuer verification. Before sending the request, the header and payload are encoded and signed with the private key of the DAPS X.509 certificate using an encryption algorithm ensuring the integrity of the message. Again, the X.509 certificate is issued by a trusted CA that verifies the DAPS. As proof of accurate verification, the CA digitally signs an X.509 certificate with his private key. The data provider trusts the verification of the CA and hence trusts the information contained in the DAPS's X.509 certificate. Also, because the data provider trusts the verification of the data consumer performed by the DAPS, it trusts the information contained in the DAT.

The header of the request includes the encryption algorithm used, the JSON Web Key Set (JWKS) containing the public keys corresponding to the private keys used to digitally sign the DAT, and the kid field that holds a key identifier indicating which key was used to secure the JWS. When the data provider receives the request, he decodes

the request resulting in the header, payload, and signature. The data provider verifies the signature by using the public key indicated by the kid field and the cryptographic algorithm located in the header. In case the public key is the one linked to the private key with which the DAT is signed, it results in a hash value that matches the original hash value meaning the signature has been verified and the message has not been tampered with. This infrastructure ensures the ability of a data provider to verify a credential shown by a data consumer (#RQ1).

The IDSA has published a reference architecture[92] providing technical specifications and guidelines including the technical specifications about the IdM. This reference architecture is used by several parties, including Fraunhofer and TNO, who help develop the reference architecture but also develop prototypes and implementations (#RQ2).

In IDS there is a component called the Clearing House that monitors and logs data transactions and data value chains, and monitors policy enforcement. It can be informed at any time about transactions being done in a data space including failed transactions, for example, caused by unauthorized requests (#RQ3).

A DAT expires in one hour but can be reused within that period of time. The DAT consists of an audience field that can be set to a data space as a whole or specific data providers. If the DAT is not expired, it can be reused at other data providers that are part of the audience (#RQ4).

Policies are set by the IDSA and the data space. The IDSA consists of more than 130 member organizations that use a democratic voting system to decide about policy changes[5]. In the reference architecture, nothing is specified about the amount of DAPS that should be implemented in a data space and by whom it should be controlled. According to experts #E5 and #E6 current implementations have a single DAPS for each data space controlled by one party. When the DAPS is controlled by one party it can decide to stop issuing credentials to a member of the data space as long as it is in line with the credential revocation policies set by the data space. This makes the DAPS centralized i.e. one party controls the issuance of credentials. Expert #E4 pointed out that a connector is now able to configure which DAPS they trust credentials from. This already takes the control away from the DAPS determining which DAPS it trusts and which not. Nevertheless, this does not take away control entirely. According to experts #E4, #E5, and #E6 federated DAPS, where DAPS can be trusted by multiple data spaces, is currently being designed. This implies that a data space is not reliant on one single DAPS making IDS decentralized. Nothing regarding this development is publicly available. The IDS Participant Information Service (ParIS) collects and provides attributes about participants in a consistent and standard manner. It can be used by other participants to retrieve a business partner's VAT, legal representatives, or organizational structure. The information the ParIS contains is provided and maintained by one party (i.e., the Support Organization) of the data space. It verifies the correctness of the claims and equips the dedicated ParIS with the new IDS participant instance[3] (#RQ5).

Although not specified in the reference architecture, each data space has a data space administrator that accepts or rejects a party from participation based on its compliance with the policies of the data space. According to experts #E4, #E5, and #E6 a data space is open to join by any party that complies with the data space policies. But, the data administrator has control over the compliance assessment and could deviate from objectively evaluating compliance with the policies of the data space. The data space administrator could be a single party or a foundation as the case with the Smart Connected Supplier Network (SCSN)[88]. It is unlikely that a data administrator will include subjectivity or deviate from the data space policies especially when a foundation performs the compliance assessment. But, it remains unknown whether the compliance assessment is only based on facts (#RQ6).

The policies for authorization can be configured by the data owner and data provider. My Data Control Technologies (MDCT), developed by Fraunhofer, enforces data sovereignty by interfering with security-related data flows in the connector enabling the enforcement of partial filtering and masking of data, limitations on the intended use, and context and situation-specific restrictions. MDCT allows to create new data usage restrictions at run-time[40]. TNO also provides an implementation allowing a data consumer and data provider to negotiate about the authorization policies[123]. The authorization policies can be specified with ODRL (#FRQ7).

5.5 Gaia-X

Gaia-X is a European initiative that aims to create a data infrastructure providing a trusted, interoperable, and sovereign data exchange platform for organizations in Europe. The goal of Gaia-X is to give European organizations greater control over their data and how it is used, while also enabling them to take advantage of the latest technologies such as artificial intelligence and the Internet of Things. The initiative was launched in 2020 by a consortium of companies, research institutions, and government agencies from across Europe, with the support of the European

Commission. The key objectives of Gaia-X are to create a data infrastructure that is transparent, secure, and open, and to promote the development of a European data ecosystem that is competitive and innovative. The Gaia-X Architecture document[2] describes the top-level Gaia-X Architecture model and focuses on conceptual modeling and important operational model aspects, and is vendor- and technology-agnostic. It describes the concepts required to establish the Gaia-X Data and Infrastructure Ecosystem and serves as the foundation for the Gaia-X Architecture's continued development, specification, and application. The policies for Gaia-X such as credential, privacy, and security policies are set by the Gaia-X Association, a combination of more than 340 companies. They use a democratic voting system in order to determine the policies of Gaia-X.

5.5.1 Design

The Gaia-X Architecture provides an infrastructure where organizations are able to offer services such as cloud offerings and data provisioning. This infrastructure requires Federation Services to enable and facilitate interoperability and portability of resources within and across Gaia-X-based Ecosystems while providing data sovereignty. Each participant has an Organizational Credential Manager (OCM) (a cloud SSI wallet app to store, present, and issue verifiable credentials) and a unique identifier (DID) that is published in the Gaia-X Registry (in SSI referred to as the verifiable data registry), a public distributed, immutable, permissionless database with a decentralized infrastructure and the capacity to automatically execute code. The backbone of the ecosystem governance stores information such as the list of issuers, the results of the issuer validation processes, and potential revocation of issuers' identity. It enables, among others, verification about whether the verifiable credential issuer's identity is from a Gaia-X complaint issuer, and verification about whether any signature was revoked.

Participants can create Self-Descriptions with the Self-Description Wizard consisting of claims that describe an entity in a machine-readable manner. Self-Descriptions can describe a participant itself, a resource, or a service offering. In order to receive proof that the claims made in the Self-Description are valid, the Self-Description can be validated by a Trust Anchor (i.e., issuer). When the claims are valid, the issuer digitally signs the Self-Description (with the private key of its X.509 certificate) which is then referred to as a verifiable credential. Additionally, the Gaia-X Association specifies certain rules that participants have to comply with in order to become a participant in the Gaia-X Ecosystem. These rules are set out in the Gaia-X Trust Framework[9] and include rules about the Self-Descriptions and verifiable credentials. Users maintain full control over their decisions thanks to these rules, which offer common management and fundamental levels of interoperability between different ecosystems. The Gaia-X Trust Framework is defined to create trust between participants of the ecosystem. An additional step is required to verify whether a verifiable credential is compliant with the Gaia-X Trust Framework and potential additional policies set by the data space. A compliance tool is hosted by the data space or Gaia-X Association that verifies the compliance of verifiable credentials. If the verifiable credential is compliant, evidence and the digital signature of the compliance tool (set with the private key of its X.509 certificate) host is added to the credential. The verifiable credentials are stored in the OCM of the respective party. When a Self-Description is about a service offering it is also stored in the Federated Catalogue that can be used by consumers to discover and select providers and their service offerings.

Gaia-X applies a different standard: the verifiable credential W3C specification that makes use of DIDs, DID documents, and verifiable credentials. Verifiable credentials about a participant itself can be used when the participant wants to consume services, for example, a data provisioning service. When reaching out to the service provider, the service consumer can show the verifiable credential accepted by the service provider. When the service consumer shows a verifiable credential issued by an issuer that the service provider trusts, verification of the digital signature will be performed by the service provider. Verification of the issuer is done via the Gaia-X Registry. The Gaia-X Registry provides an interface where a data provider can verify a verifiable credential. The data provider inputs the verifiable credential it wants to verify. The Gaia-X Registry extracts the issuer DID from the verifiable credential and searches for the associated DID document consisting of information such as (company) name, verification method, and the public key(s). The public key is used to verify the digital signature with which the issuer signed the verifiable credential shown by the data consumer. In case the public key is the one linked to the private key with which the verifiable credential is signed, it results in a hash value that matches the original hash value meaning the signature has been verified. To prevent the occurrence of replay attacks (an attacker intercepts the data and re-transmits it to get access to the service), the service consumer also digitally signs the verifiable credential with its private key. The same process for verifying the digital signature of the issuer is followed but now the service consumer's DID is used.

Parties are also able to provide principal verifiable credentials to individuals (e.g., employees) that are registered in their internal IdM. This demands a connection between the internal IdM and its OCM. Inside the OCM roles about

an employee can be defined and a registration link (e.g., a QR code) can be generated to provide to the respective employee. The employee is then able to scan the QR code with its Personal Credential Manager (PCM), a personal SSI wallet app to store and present verifiable credentials. The employer is then able to issue verifiable credentials via its OCM to the PCM allowing employees to act on behalf of the employer.

5.5.2 Analysis

For a service consumer (now a data consumer) to receive access to the data of a service provider (now a data provider) a two-step process is followed. First, the data consumer needs to acquire a verifiable credential issued by an issuer trusted by the data provider. When a data consumer request access to data, the data provider returns a trusted list of issuers and the verifiable credentials it accepts. In the second step, the data consumer identifies himself to the data provider with the verifiable credential. If the data consumer is who he says he is, and the verifiable credential is issued by a trusted issuer, authorization is performed and access is granted accordingly. During the second step of this process, the verification of the issuer of the credential takes place.

Assuming the data consumer holds an accepted verifiable credential, a verifiable presentation is created that is shown to the data provider. A verifiable presentation can hold multiple verifiable credentials and allows a data consumer to selectively choose which parts of a verifiable credential are included. A data consumer can, for example, only show the company name and ID, and exclude the address and postal code. When the data provider shows a verifiable credential issued by an issuer that the data provider trusts, verification of the digital signature is performed by the data provider. The Gaia-X Registry provides an interface where the data provider can verify a verifiable credential. The data provider inputs the verifiable credential it wants to verify. The Gaia-X Registry extracts the issuer DID from the verifiable credential and searches for the associated DID document consisting of information such as (company) name, verification method, and the public key(s). The public key is used to verify the digital signature with which the issuer signed the verifiable credential shown by the data consumer. In case the public key is the one linked to the private key with which the verifiable credential is signed, it results in a hash value that matches the original hash value meaning the signature has been verified. To ensure the integrity of the message and prevent the occurrence of a replay attack, the data consumer also digitally signs the verifiable credential with its private key. The same process for verifying the digital signature of the issuer is followed but now the data consumer's DID is used. This infrastructure enables data providers to verify whether the verifiable credential is issued by an issuer they trust (#RQ1).

The Gaia-X Association published an architecture[2] providing technical specifications and guidelines also covering the IdM. This reference architecture is used by multiple parties among which the Gaia-X Federation Services (GXFS)[5] who helps develop the architecture and built an implementation (#RQ2).

Gaia-X specifies a Contract Logging Service that receives logging messages (data provided, data received, policy enforced, and policy-violating messages) during a transaction to trace each event. In case a service consumer has no rights to access a resource, the transaction will fail and a policy-violating message (including the details) will be stored at the Contract Logging Service. The participants of the transaction and, if necessary, a third eligible party may query the stored information both during and after the transaction. Furthermore, the Ocean Protocol[23][87] has been proposed during a hackathon. It is a marketplace where data providers can offer their datasets and data consumers can buy them. A dataset is tokenized such that the Gaia-X Data Exchange Logging Service[6] can include the token in the transaction, allowing the data provider to prove that the data consumer requested access to the dataset (#RQ3). The Gaia-X Data Exchange Logging Service is an implementation (developed by GXFS) of the Contract Logging Service specified by Gaia-X[26].

Verifiable credentials can be reused by all parties that trust the issuer of that credential. Verifiable credentials are stored in the wallet of the holder. When the holder (now data consumer) requests access to a dataset, the data provider indicates the verifiable credentials it accepts along with the issuers of the credentials. If the data consumer already has a credential issued by an issuer that matches the ones indicated by the data provider, the data consumer can reuse it (#RQ4).

The Gaia-X Association is a combination of more than 340 companies that set the policies (e.g. credential, privacy,

[5] an initiative funded by the German Federal Ministry for Economic Affairs and Climate Action. It was formed to launch the creation of GXFS, as determined by a collaborative community approach. Gaia-X is the owner of the specifications and open-source code produced by this funded project source: Gaia-X Federation Services[54]

[6] a smart contract-based logging service that records transactions to a data service on the ledger with distributed ledger technology

and security policies)[27]. They use a democratic voting system in order to determine the policies of Gaia-X. The data space determines additional policies they find relevant to apply. The issuers that are supported depend on the trusted list of the data provider. The data consumer is limited to verifiable credentials issued by these trusted issuers but has the ability to choose which issuer to request a verifiable credential from. This way there is not one party in control of the issuance of credentials. In case an issuer is not willing to issue a credential to the data consumer, he is able to request the credential at a different issuer. According to expert #E4 a data space will likely set a policy that requires a participant to have a compliance and a membership credential. This puts significant control in the hands of the issuer of compliance and membership credentials. It is unknown whether a single party, multiple parties, or the whole data space will control these issuers. Participant information can be retrieved from the Federated Catalogue, a publicly distributed, immutable, permissionless database with a decentralized infrastructure and the capacity to automatically execute code. Although not specified in the architecture, the data space together decides whether a party may join the data space based on its compliance with the data space policies (#RQ5). According to expert #E4: "it will be a service that checks whether a party complies with the policies set by the Gaia-X Association and the data space. It is unlikely that a party complying with the policies will be denied but not impossible". It remains unknown whether the compliance assessment is only based on facts (#RQ6).

The Gaia-X Ontology Specification Draft specifies that the policies will be defined in a domain-specific language (e.g., ODRL and Rego)[14]. Rego is a policy language that allows users to define policies in the form of rules, which consist of a condition and an associated action. When a policy is evaluated, the system checks the conditions of each rule to determine whether the associated action should be taken. The authorization policies can be self-configured and negotiated between the data provider and data consumer via the Service Agreement Service. This agreement includes usage policies and the required measures to implement them[2] (#RQ7).

5.6 European Digital Identity Wallet

The EUDI Wallet is a digital wallet for storing and managing digital identity documents, such as eID cards, ePassports, and other types of identification. The goal of the European Commission is to have a secure European e-Identity that any citizen can use anywhere in Europe with technology to control what data and how data is used. All EU Member States must provide their citizens with an EUDI Wallet by 2024[19]. The EUDI Wallet is being developed by the EU as a way to improve the security and convenience of digital identity verification and authentication. With the intention to be a secure and easy-to-use platform for storing and accessing digital identity documents, as well as for performing online identity verification and authentication. It is being designed to work across the EU and to be compatible with a wide range of national and international identity systems. The EUDI Wallet is being developed as part of the EU's Digital Single Market strategy[18], which aims to create a single market for digital goods and services within the EU. The primary objective of the EUDI Wallet is to promote trusted digital identities for all Europeans allowing users to be in control of their own online interactions and presence. The eIDAS expert group[16] published the European Digital Identity Architecture and Reference Framework (ARF)[50] providing the objectives, roles of the actors of the ecosystem, wallet's functional and non-functional requirements, and the potential building blocks. Their intention is to update this outline during the process and develop it to a full architecture and reference framework including all technical specifications of the IdM.

5.6.1 Design

The EUDI Wallet is in an early stage; the first meetings about the EUDI Wallet started on 30 September 2021. At present, there is not much information available regarding the design except for the potential roles of the EUDI Wallet ecosystem, and the wallet's functional and non-functional requirements.

The EUDI Wallet can be viewed as a collection of various products and trust services that give the user complete control over how their Personal Identifiable Data (PID), Qualified Electronic Attestation of Attributes (Qualified EAA), Non-qualified Electronic Attestation of Attributes (Non-qualified EAA), and any other personal data used within the EUDI Wallet. In total there are 14 potential roles of the EUDI Wallet ecosystem:

1. The end users of the EUDI Wallet
2. EUDI Wallet issuers Member States or organizations complying with the terms of conditions determined by Member States that make the EUDI Wallet available for the end users. The EUDI Wallet would be responsible

for assuring adherence to the requirements for EUDI Wallets.

3. PID providers Verify the identity of the end-user and maintain an interface to provide PID securely to the EUDI Wallet and make information accessible for relying parties to verify the PID's validity without having the ability to learn about the use of the PID.
4. Providers of registries of trusted sources Provide a registration service for relevant entities that are required to be verified in a trustworthy manner (e.g., EUDI Wallet issuers, Qualified EAA providers, and Non-qualified EAA providers).
5. Qualified EAA providers Qualified Trust Service Provider (QTSP)s would be the one providing the Qualified EAA and would need to maintain an interface for requesting and providing Qualified EAA. This interface might include a mutual authentication interface with the EUDI Wallet or an interface towards an authentic source for attribute verification. Information or the location of the service to request the validity status of the Qualified EAA must be provided without the ability to receive information about the use of the attestations.
6. Non-qualified EAA providers Any Trust Service Provider (TSP) is able to provide Non-qualified EAA and would be supervised under the eIDAS Regulation. A TSP would need to offer a way for users to request and receive Non-qualified EAA. They might have to provide valid information regarding the Non-qualified EAA without the ability to receive information about the use of the Non-qualified EAA, depending on the domain rules.
7. Qualified and non-qualified certificate for electronic signature The EUDI Wallet must enable users to sign documents and data with a qualified electronic signature or seal.
8. Providers of other trust services Other qualified or unqualified trust service providers, like those who offer timestamps, that may interact with the EUDI Wallet.
9. Authentic sources Public or private repositories or systems that contain attributes about a natural or legal person. The authentic sources must provide an interface to Qualified EAA providers allowing them to verify the authenticity of attributes. Hence, they might have to maintain a user interface to get consent from the user before giving Qualified EAA providers access to the person's data.
10. Relying parties Natural or legal persons that rely upon an electronic identification or a trust service. They would request the attributes contained in the PID, Qualified EAA, or Non-qualified EAA that they require to perform their job. The attributes that they request have to be within the limits of applicable legislation and rules. The relying parties are responsible for the authentication of the received attributes. Furthermore, the relying parties have to maintain an interface where the EUDI Wallet can request attestations to authenticate the relying party i.e., mutual authentication.
11. Conformity assessment bodies (CAB) Public or private bodies that are responsible for regularly auditing QTSPs, EUDI Wallet issuers, and TSPs on behalf of the Member States. The EUDI Wallet issuers have to be certified before the Member States will issue the EUDI Wallet. QTSPs and TSPs have to receive the qualified status before being able to provide Qualified EAA and Non-qualified EAA.
12. Supervisory bodies The supervisory bodies supervise QTSPs and take appropriate action, in relation to non-qualified TSPs who do not meet required standards or qualifications, if needed. This may involve investigating their practices, taking enforcement actions, or revoking their licenses. The Member States who designated the supervisory bodies are required to notify the European Commission of the existence and identity of these supervisory bodies.
13. Device manufacturers and related subsystems providers Specific devices or services are required for secure cryptographic material storage such as local storage, online internet access, sensors, offline communication channels, emitters, cloud service providers, and app store providers.
14. Catalogue of attributes and schemes for the attestations of attribute providers The catalogue should publish relevant information about attestations provided by Qualified EAA and Non-qualified EAA providers. It would allow other entities, such as relying parties, to discover what attributes and schemes are provided, and how to validate/verify them, as well as distinguish between types of qualified electronic attestations of attributes.

The EUDI Wallet shall provide the following functionalities (a selection of the requirements are mentioned for elaboration):

1. Store PID, Qualified EAA, and Non-qualified EAA Storage can be local on a device the user holds or remote in a cloud-based infrastructure. Storage shall either be only local storage or hybrid storage with a pointer to remote storage stored locally. Storing the PID, Qualified EAA, and Non-qualified EAA prevents requesting the PID, Qualified EAA, and Non-qualified EAA every time the information is needed reducing the ability of electronic attestation providers to track the use of the provided electronic attestations on the user's side.

2. Request an acquire PID, Qualified EAA, and Non-qualified EAA A functionality shall be integrated where the user can request and acquire PID during the onboarding, for example, through an interface with electronic identifications that provide a high level of assurance. Users shall be offered the ability to request and acquire both Qualified EAA and Non-qualified EAA via an interface of Qualified EAA and Non-qualified EAA providers. Users shall be able to delete material from the wallet, for example, PID, Qualified EAA, and Non-qualified EAA.
3. Cryptographic functions To implement most of the functionalities a set of cryptographic functions shall be provided to enable secure access. These functionalities shall be used to manage, for example, electronic identification of the user to relying parties, authentication of PID, Qualified EAA, and Non-qualified EAA when those are linked to the EUDI Wallet, and secure storage of sensitive personal data on the device. The cryptography of the supported algorithms shall be strong enough to ensure confidentiality, integrity, and authenticity.
4. Mutual authentication In order to strengthen ecosystem security and trust, the EUDI Wallet itself shall be able to prove to the relying party the origin and integrity of the EUDI Wallet being utilized. This shall prove valid certification of the EUDI Wallet and that the solution was installed on an appropriate device with sufficient security. Furthermore, the EUDI Wallet shall be able to identify and authenticate third parties it interacts with.
5. Select, combine, and share PID, Qualified EAA, and Non-qualified EAA The EUDI Wallet shall be able to perform identification with legal weight when required. To reduce technical complexity the EUDI Wallet shall use a common protocol for identification, attribute sharing, and verification of the integrity and authenticity of the information. Collection of wallet usage information and combining personal data stored on the EUDI Wallet with personal data of other services or third parties shall be prohibited. The latter with the exception when specifically requested by the user. Privacy by design and selective disclosure of attributes shall be enforced.
6. User interface for user awareness and authorization mechanism The EUDI Wallet shall provide information that enables the user to make properly informed decisions, for example, his rights for data protection under the GDPR, the reason to share the electronic attestation or attribute, and the identities of the different parties the user will interact with. An "EU Digital Identity Wallet trust mark" and events regarding the use of a user's EUDI Wallet will be displayed to the user. The EUDI Wallet shall rely on a standardized authorization mechanism designed to ensure security and privacy. The user shall be able to define authorizations regarding the actions the EUDI Wallet is allowed to perform on its behalf. The user shall be required to use two-factor authentication in a combination of at least two out of the authentication factors: a proof of knowledge, a proof of possession, and a proof of inherence.
7. Sign documents and data with a qualified electronic signature or seal A user shall be able to produce qualified and unqualified electronic signatures and seals by the wallet being a Qualified Signature Creation Device (QSCD), a local QSCD, or using an interface to a qualified remote QSCD service.
8. Interfaces with external entities The EUDI Wallet will need to interact with certain interfaces of external entities for which specific requirements, specifications, and standards apply. These interfaces include interfaces toward Member States Infrastructures, interfaces towards Member States' identity cards, interfaces towards relying parties, brokers, or proxies, trusted registry interfaces, and device interfaces.

5.6.2 Analysis

Users can request three types of credentials: PID, Qualified EAA, and Non-qualified EAA. PID is information from which the entity subject of the data can be extracted and may, for example, be issued by the same organizations that issue official identity documents that digitally sign the PID credential. Qualified EAA credentials are digitally signed and issued by a QTSP that is legally valid. Non-qualified EAA credentials are digitally signed and issued by a TSP. The PID providers shall make information accessible for relying parties to verify the PID's validity as might be the case with Qualified EAA and Non-qualified EAA providers. This implies that it is possible to verify the issuer of a credential, otherwise, relying parties will not be able to know what provider to contact to verify the credentials' validity. However, the process of how relying parties can verify the validity of these credentials is unknown.

The eIDAS expert group published the ARF providing the objectives, roles of the actors of the ecosystem, the wallet's functional and non-functional requirements (technical specifications), and the potential building blocks. The ARF will be updated during the process and developed to a full architecture and reference framework including all technical specifications of the IdM (#RQ2).

Nothing is specified in the ARF that indicates the possibility to prove a request has been made. Since the ARF is just an outline we cannot assume that this implies that Proof of Request will not be provided. There is insufficient information to assess whether Proof of Request will be provided or not (#RQ3).

One of the functionalities the EUDI Wallet shall provide is select, combine, and share credentials with relying parties. The requirement of this functionality states that privacy by design and selective disclosure of attributes shall be enforced. This indicates that a user can select and share a subset of the attributes from a credential. A functionality to combine a credential indicates that a user can combine attributes from different credentials to prove its identity which implies that credentials are reusable (#RQ4).

Credentials are issued by multiple parties (e.g., universities, governmental parties, QTSP, and TSPs). The policies are at this point determined by the eIDAS expert group which consists of Member States' experts. But it is unknown whether the service providing information regarding members of a data space including the services that they offer, and granting access to a data space is done in a decentralized manner (#RQ5).

Nothing is specified in the ARF regarding data spaces. Since the ARF is just an outline we cannot assume that this implies that a data space will not be open. There is insufficient information to assess whether a data space is open or not (#RQ6).

Nothing is specified in the ARF regarding authorization policies that can be assigned to a user regarding data accessibility. But, according to expert #E3 authorization will be done ABAC and will be determined by the data provider (#RQ7).

5.7 Sovrin Network

Sovrin is a decentralized identity layer for the internet that is open-source and based on a permissioned DLT allowing for secure, verifiable, and transparent record-keeping and data exchange. Sovrin is open to the public, but permissioned because only stewards (trusted organizations e.g., banks, colleges, and governments) can operate nodes that participate in the consensus protocol. The Sovrin ledger serves as the foundation and provides a decentralized global public utility for SSI. Sovrin is a deployment of Hyperledger Indy that gives developers the tools, libraries, and reusable components to provide digital identities rooted on distributed ledgers, supporting interoperability across applications, administrative domains, and any other silo[36]. Sovrin is operated by a global consortium of organizations called the Sovrin Foundation responsible for the governance of the stewards and ensuring that the network operates in a transparent and secure manner. One of the main goals of Sovrin is to provide individuals and organizations with a secure, private, and convenient way to manage and use their digital identities. It is intended to enable people to easily and securely prove their identity online, and to allow organizations to verify the identities of their customers and employees.

5.7.1 Design

Sovrin consists of three key components: the Sovrin ledger, Sovrin agents, and Sovrin clients[21]. The Sovrin architecture can be summarized as shown in Figure 6.

The Sovrin ledger is a globally distributed ledger of root identity records maintained by the stewards running on the Plenum protocol[65], a consensus protocol optimized for security and scale. The Sovrin ledger is governed by the Board of Trustees of the Sovrin Foundation that approves the policies governing stewards. The ledger is operated by the validator and observer nodes. All write operations to the Sovrin ledger are sent to a validator node running the Plenum protocol for validation of new Sovrin transactions. Observer nodes fulfill three functions required to scale the network: offloading read requests so that the performance of validator nodes are not affected, hot standbys: switching towards a validator node in case of failure or compromise of another validator node, and pushing subscriptions to actors subscribed to specific Sovrin ledger events without increasing load on the validator nodes. The Sovrin ledger stores:

- DIDs and their associated DID documents consisting of the public keys and communication endpoints
- Schema and credential definitions

- Revocation registry
- Agent authorization policies

Every DID resolves to a DID document, a JSON-LD file containing all the metadata required to demonstrate control and ownership of a DID as well as to exchange the cryptographic keys and resource endpoints required to start trustworthy peer interactions between Sovrin entities.

Sovrin makes a distinction between public and private DIDs to support both its Privacy by Design architecture and the ability to scale. The public DIDs are directly stored on the ledger and are primarily required by issuers of credentials. They are stored on the Sovrin ledger so that a verifier who obtains proof of a verifiable credential can check the issuer's public key and verify the proof. The private DIDs on the other hand are stored off-ledger by the Sovrin agents for two identity holders. No one else needs to know about the relationship between the two identity holders and the two DIDs used. As a result of storing the private DIDs and DID documents off-ledger, the transaction can take place entirely off-ledger reducing the load on the Sovrin ledger while maintaining the privacy and integrity of each identity holder.

To establish trust between two identity holders verifiable credentials can be exchanged. The Sovrin ledger stores schema definitions and credential definitions in order to support the interoperable exchange of verifiable credentials. A schema definition defines a set of attributes and formats to define claims on a credential in a machine-readable manner. These schema definitions can be used by credential issuers to create an issuer-specific credential. Both are stored on the Sovrin ledger allowing issuers to re-use existing schema, and enabling verifiers to look up the issuer's credential definition, obtain their public key, and verify the origin and integrity of a verifiable credential.

The Sovrin ledger stores a revocation registry that is written to by the issuer for issuers that require to revoke their issued verifiable credential. The revocation registry consists of references to the credential definition containing a cryptographic accumulator (a single number) that can be checked by any verifier when it needs to ensure that a verifiable credential has not been revoked by the issuer. All this is done while ensuring security and privacy.

To further protect identity holders' security, the Sovrin client provides them with the ability to authorize agents (and to revoke that authorization) to present credential proofs on their behalf. The identity holder can specify agent authorization policies enabling them to prove to a verifier that an agent is authorized to act on their behalf.

Sovrin agents are not just a client but also a service that has an addressable network endpoint that will be as highly available as other important network infrastructure such as routers, DNS, and email. It provides permanent, privacy-protecting methods for identity and data management transactions to Sovrin identity owners. The Sovrin agents provide four key functionalities:

1. Persistent P2P messaging endpoints.
   Make clients addressable in the same way that IP routers and domain name servers do
2. Coordination endpoints for multiple clients.
   Coordinate messages and state across multiple Sovrin clients operating on edge devices such as smartphones and laptops
3. Encrypted backup of Sovrin keyrings.
   Maintain an encrypted backup of the identity owner's Sovrin keys to simplify key recovery
4. Encrypted data storage and sharing.
   Simplify and automate the process of storing and sharing data enabling identity owners to encrypt and manage data using their Sovrin key.

An identity owner can either host a Sovrin agent himself or choose a third party to host the service. Regardless of the type of host chosen, the Sovrin identities are portable, the identity owner always controls the agent endpoints, and the Sovrin keys and keychains are portable across any device. The Sovrin data graphs should be maintained in a system-independent semantic graph format (e.g., JSON-LD or OASIS XDI). To keep contextual separation, an identity owner may have distinct Sovrin identities registered on the Sovrin ledger. Linkability between the Sovrin identities is avoided by providing a separate agent endpoint for each Sovrin identity.

Sovrin clients are apps (typically on local devices like smartphones and laptops) that Sovrin identity owners use to interact with Sovrin agents and the Sovrin ledger, and to carry out all kinds of identity transactions. The most important task of a Sovrin client is to manage and protect the identity owner's keychain. The Sovrin keychain is

designed from the ground up to be fully self-sovereign i.e., independent of any specific Operating System, device, application, or network even the Sovrin Network. A client must maintain a copy of all or a portion of the identity owner's keychain and the Sovrin data container, and maintain the security and privacy of the identity owner's Sovrin transactions. For an identity owner to register his first Sovrin identity, he must connect with a trust anchor (an existing Sovrin identity owner) that has permission to add new identities to the network. Sovrin serves as a Decentralized Public Key Infrastructure (DPKI)

Sovrin defined a Trust Framework with the primary goal to ensure sufficient geographic diversity, jurisdictional diversity, industry diversity, and more. It defines the rights and responsibilities of each of the participants in the network explaining how the "permissioning" of a public permissioned ledger works and serves as the common contract between all Sovrin participants. The Trust Framework applies the web of trust model where no single authority or hierarchy decides who can trust whom but rather everyone decides who he trusts, and then use that relationship to decide who else to trust. The web of trust model applies to the active Sovrin stewards and trust anchors. A sustainable, organic web of trust is maintained by regularly verifying Sovrin stewards and trust anchors based on the defined criteria in the Sovrin Trust Framework. Additionally, guidelines are defined on how newly provisioned identity owners can earn their trust anchor status.

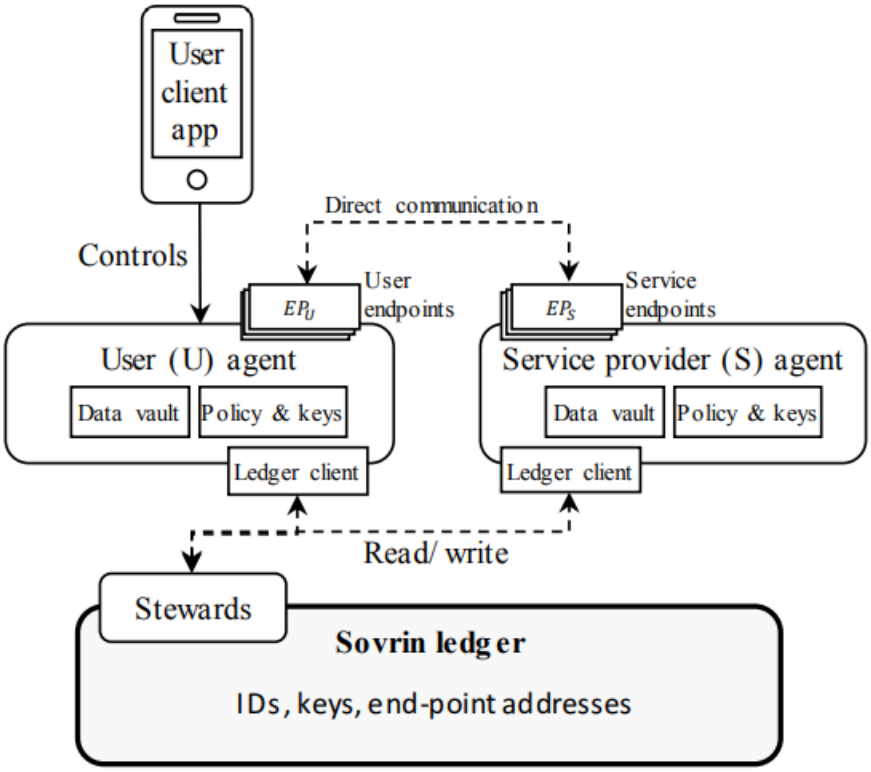

Figure 6: The Sovrin architecture with as foundation the permissioned Sovrin ledger. DIDs and their associated DID documents containing public keys, and communication endpoints are stored on the ledger. Written to by stewards that legally adhere to the Sovrin Trust Framework. Users and organizations can communicate via agents that are addressable network points. Source of the figure: Dunphy and Petitcolas[32].

5.7.2 Analysis

Sovrin supports the verification of verifiable credentials. The issuer of the verifiable credential digitally signs it with the private key of its X.509 certificate as proof that the claims of the verifiable credential can be trusted. The data consumer digitally signs the verifiable credential with the private key of its X.509 certificate to ensure integrity i.e., ensure that the authentication cannot be replayed.

Verifying the digital signature of the issuer is done in two steps. In the first step, the data provider extracts the issuer DID from the verifiable credential shown by the data consumer and searches for it on the Sovrin ledger. Resolving the DID results in a DID document consisting of information such as id, verification method, and the public keys. The public key can be used to verify the digital signature. In case the public key is the one linked to the private key with which the verifiable credential is signed, it results in a hash value that matches the original hash value meaning the signature has been verified.

The DIDs and DID documents stored on the Sovrin ledger are anonymized. As a result, Sovrin does not provide the ability for a data consumer to verify whether the public key really belongs to an issuer trusted by the data provider. This happens during the second step. The data provider must maintain its own trusted list of issuers consisting of autonym information about the trusted issuers including the DID that is stored on the Sovrin ledger. The data provider knows the verifiable credential was issued by an issuer it trusts when the issuer DID retrieved from the verifiable credential corresponds with one of the Sovrin DIDs listed in the data provider's trusted list (#RQ1). The same process is followed for verifying the data consumer's digital signature but this time the data consumer's DID is used.

The Sovrin Foundation has published multiple documents, for example, the Inevitable Rise of Self-Sovereign Identity[35] which covers the purpose of Sovrin and the major roles that are intrinsic to the Sovrin infrastructure. The Technical Foundation of Sovrin[21] explains the technical foundation of Sovrin for internet architects, analysts, and developers. The Sovrin DID method specifications are openly available and conform to the requirements specified in the DID method specification published by the W3C Credentials Community Group[37]. The Sovrin Governance Framework Working Group published the Sovrin Trust Framework[51]. Sovrin is a deployment of Hyperledger Indy[36] that provides documentation about their repositories and consist of technical specifications and guidelines[60][61][62][63][64] (#RQ2).

Sovrin only focuses on the identification and authentication of entities and does not implement authorization regarding data accessibility of data consumers (it does provide the ability to define authorization policies for agents to present specific credentials on behalf of a holder) (#RQ7). Proof of request can thus not be provided (#RQ3). Sovrin does support a proof of existence claim type, a hash of a digital object enabling data providers to prove that a digital object existed at a given point in time. This type of claim is especially useful for proving consent under privacy regulations such as the GDPR. When the data consumer receives data from a data provider, the data consumer can provide the data provider with a consent receipt (e.g. permission that the data consumer will use the data provider's data for 30 days and attributes will only be used for shipment). A proof of existence claim can then hash the link to the consent receipt and write it to the ledger in case proof is required. This proof of existence claim is only relevant for identity information that has been exchanged and not as proof that access to data has been requested.

Sovrin itself does not use or issue credentials. However, Sovrin enables issuers to define the schema and define issuer-specific credentials, allows verifiers to verify verifiable credentials, and enables the exchange of verifiable credentials between identity owners. Verifiable credentials can be reused at as many parties as the number of parties who trust the issuer of the respective credential (#RQ4).

The Sovrin Foundation is a not-for-profit global consortium that is only committed to the governance (i.e. provides business, legal, and technical support) of the Sovrin Network. The Sovrin Foundation uses a democratic system to determine governance policies. The governance only consists of policies that are required for the cooperating nodes of Sovrin to store and serve SSI records securely with respect to the privacy of the holders[38]. In the event of technical changes, the nodes have to implement the changes or will be excluded from the network. Hyperledger Indy is controlled by the Hyperledger Foundation consisting of more than 140 organizations, all having the ability to propose contributions to the Hyperledger Foundation's technical codebase. If any decision needs to be approved by a vote, the members of the Governing Board, Technical Oversight Committee, and Marketing Committee shall cast one vote per voting representative[34]. Credential issuance, the service providing information regarding members of a data space including the services that they offer, and granting access to a data space is all not part of Sovrin (#RQ5).

Anyone is able to register a DID by paying 10 euros. Whether that DID is able to participate in a data space depends on the policies applied in the data space and is not part of Sovrin (#RQ6).

6 Comparing Available Inter-organizational IdM Treatments

The results of the analyses on iSHARE, IDS, Gaia-X, the EUDI Wallet, and Sovrin are summarized in Table 9. The scales used are Yes (when the treatment satisfies the requirement), No (when the treatment does not satisfy the requirement), Not Applicable (N.A. when the requirement is out of scope for the treatment), and ? (when the information is not found or too limited for evaluation).

Both iSHARE and IDS use the OAuth 2.0 standard and an X.509 certificate signed by a CA consisting of a related public key and private key. The requests sent with the OAuth 2.0 protocol include a client assertion produced by the data consumer's client application, signed with its private key, and presented as proof of the client's identity. In iSHARE the client assertion is included in the request to get an access token from the data provider. The request is encrypted with the private key of the data consumer using an encryption algorithm. The data provider is able to verify the client assertion by verifying the digital signature with the public key using the encryption algorithm located in the header of the request. In case the public key is the one linked to the private key with which the client assertion is signed, it results in a hash value that matches the original hash value meaning the signature has been verified and the message has not been tampered with. Additionally, the data provider verifies the payload, and checks at the iSHARE Satellite whether the EORI number (from the iss and sub field) is an active member in the data space. In IDS the client assertion is used differently: it is used to request a DAT at the DAPS. The DAPS is able to verify the client assertion in the same way as the data provider in iSHARE. Another distinction is that the DAPS does not communicate with a third party (e.g., the iSHARE Satellite) to check if the data consumer is an active member in the data space; this information is known by the DAPS. When verification is successful, the DAPS issues a DAT that it digitally signs with the private key of its X.509 certificate using an encryption algorithm. The DAT is then used to request access to data at the data provider. In iSHARE this is done with the access token which does not qualify as a credential following our definition. The data provider is able to verify the DAT by verifying the digital signature with the public key from the JWKS indicated by the kid field and the encryption algorithm located in the header of the request. In case the public key is the one linked to the private key with which the DAT is signed, it results in a hash value that matches the original hash value meaning the signature has been verified and the message has not been tampered with.

Gaia-X applies a different standard: the verifiable credential W3C specification that makes use of DIDs, DID documents, and verifiable credentials. The verifiable credentials are used by data consumers to request access to the data of the data provider. Verifiable credentials are issued by the so-called issuers that digitally sign the credential with the private key of their X.509 certificate. A data provider indicates to the data consumer what verifiable credentials they accept and the issuers they trust. When a data consumer shows a verifiable credential issued by an issuer that the data provider trusts, verification of the digital signature will be performed by the data provider. The Gaia-X Registry provides an interface where the data provider can verify a verifiable credential. The data provider inputs the verifiable credential it wants to verify. The Gaia-X Registry extracts the issuer DID from the verifiable credential and searches for the associated DID document consisting of information such as (company) name, verification method, and the public key(s). The public key is used to verify the digital signature with which the issuer signed the verifiable credential shown by the data consumer. In case the public key is the one linked to the private key with which the verifiable credential is signed, it results in a hash value that matches the original hash value meaning the signature has been verified. To enable the data provider to prove the integrity of the message and prevent the occurrence of a replay attack, the data consumer also digitally signs the verifiable credential with its private key. The same process for verifying the digital signature of the issuer is followed but not the data consumer's DID is used.

The EUDI Wallet knows three types of credentials: PID, Qualified EAA, and Non-qualified EAA. All three are digitally signed by the issuer. The PID providers shall make information accessible for relying parties to verify the PID's validity as might be the case with Qualified EAA and Non-qualified EAA providers. This implies that it is possible to verify the issuer of a credential, otherwise, relying parties will not be able to know what provider to contact to verify the credentials' validity. However, the process of how relying parties can verify the validity of these credentials is unknown.

The Sovrin ledger functions as a verifiable data registry that stores DIDs and DID documents, verifiable credential schema, verifiable credential definitions, and verifiable credential revocations. With the verifiable credential schema, issuers are able to create issuer-specific credential definitions. The verifiable credentials are used by data consumers to request access to the data of the data provider. Verifiable credentials are issued by so-called issuers that digitally sign the credential with their private key. The data provider maintains a trusted list of issuers and the verifiable credentials it accepts. This trusted list is maintained off-ledger because it consists of autonym information. The

DIDs and DID documents stored on the Sovrin ledger are anonymized to ensure privacy and scaling purposes. A data provider is able to verify a verifiable credential by extracting the issuer DID from the verifiable credential and searches for it on the Sovrin ledger. Resolving the DID results in a DID document consisting of information such as id, verification method, and the public keys. The public key can be used to verify the digital signature but at this point, the data provider is still unaware whether the verifiable credential is issued by an issuer it trusts. This verification should be done by the data provider off-ledger and is not supported by Sovrin. Although this is inherited from Sovrin ensuring privacy it results in only partly realizing issuer verification. Sovrin is significantly different than iSHARE and IDS: it supports a different standard, uses DLT, and does not involve itself in the issuance of credentials. Sovrin and Gaia-X both apply the SSI model and the Sovrin ledger is comparable to the functioning of the Federated Catalogue.

The specifications of iSHARE are very specific in terms of the technology that should be applied and are defined by a central party, the iSHARE Foundation. IDS and Gaia-X take a different approach, they have a reference architecture describing the functioning of the IdM but support freedom about the technology to use. Additionally, both reference architectures are developed by a consortium of organizations that all have influence in the decision-making. The eIDAS expert group published the ARF providing the objectives, roles of the actors of the ecosystem, the wallet's functional and non-functional requirements (technical specifications), and the potential building blocks. The ARF will be updated during the process and developed to a full architecture and reference framework including all technical specifications of the IdM. The Sovrin Foundation published several papers about the functioning of Sovrin and include technical specifications. The technology that Sovrin applies is Hyperledger Indy of which the technical specifications, guidelines, and source code are openly available.

The specifications of iSHARE contain nothing that indicates support of Proof of Request. IDS has a Clearing House which is similar to the Contract Logging Service of Gaia-X. Both monitor and log data transactions, enforce policies, and can be informed by a data provider to prove an (unauthorized) request was done. There is insufficient information to assess whether the EUDI Wallet will provide Proof of Request or not. Sovrin does not provide a service that supports Proof of Request.

Within iSHARE the client assertion qualifies as a credential and is generated specifically for a data provider. Hence, the credential is not reusable. The relevant credential in IDS is the DAT which has an expiration time of one hour but can be reused during that period at data providers that are part of the audience. Gaia-X and Sovrin both make use of verifiable credentials that can be reused at as many parties that trust the issuer of the verifiable credential. One of the functionalities the EUDI Wallet shall provide is select, combine, and share credentials with relying parties. A functionality to combine a credential indicates that a user can combine attributes from different credentials to prove its identity which implies that credentials are reusable.

The iSHARE Foundation is in control of the policies and specifications, and fulfills the role of Scheme Owner which manages participant information. Participants can contact the Scheme Owner to check participant information such as whether a party is a member of a specific data space. Also, iSHARE credentials are issued by the data consumer self. Hence, iSHARE is evaluated as a centralized treatment. According to experts #E5 and #E6 the DAPS is managed by one party, hence centralized. According to expert #E4, IDS is less decentralized than Gaia-X but still is. His evaluation is based on information unknown to the public: the development of a federated DAPS, where a DAPS can be trusted by multiple data spaces. This implies that a data space is not reliant on one single DAPS making IDS decentralized. But no information is available regarding the federated DAPS and is still in development. In IDS the ParIS collects and provides attributes about participants in a consistent and standard manner. It can be used by other participants to retrieve a business partner's VAT, legal representatives, or organizational structure. The information the ParIS contains is provided and maintained by one party (i.e., the Support Organization) of the data space. It verifies the correctness of the claims and equips the dedicated ParIS with the new IDS participant instance[3]. Hence, IDS is evaluated as a centralized treatment. In Gaia-X policies are set by the Gaia-X Association, a combination of more than 340 companies that use a democratic voting system in order to determine the policies of Gaia-X. The data space determines additional policies they find relevant to apply. Credential issuance is done by issuers trusted by the data providers. In case an issuer is not willing to issue a credential to the data consumer, he is able to request the credential at a different issuer. Participant information is available at the Federated Catalogue, a publicly distributed, immutable, permissionless database with a decentralized infrastructure and the capacity to automatically execute code. The data space together decides whether a party may join the data space based on its compliance with the data space policies. In Gaia-X the degree of decentralization can be questioned in terms of the compliance and membership credential issuers but both can be controlled by the data space. Hence, Gaia-X is evaluated as decentralized. The Sovrin Foundation, a not-for-profit consortium, determines the policies of the Sovrin Network using a democratic system to determine the policies. Hyperledger Indy is controlled by the Hyperledger Foundation consisting of more

than 140 organizations, all having the ability to propose contributions to the Hyperledger Foundation's technical codebase. Decisions are made using a democratic voting system[34]. Credential issuance, the service providing information regarding members of a data space including the services that they offer, and granting access to a data space is all not part of Sovrin. Hence, Sovrin is evaluated as decentralized. There is insufficient information to assess whether the EUDI Wallet is decentralized or not.

ISHARE, and IDS have a data space administrator that accepts or rejects a party from participation. In iSHARE the data administrator is a centralized party, the iSHARE Satellite that validates a party based on the iSHARE and data space policies. It guarantees trustworthy onboarding and agreed to all the rules of iSHARE implying that their validation is objectively performed. In IDS the data administrator could be a centralized party or a foundation managed by the data space that validates compliance of a party based on the data space policies. It is unknown whether the validation is purely objectively performed. In Gaia-X the members of the data space together decide whether a party may join the data space based on the policies of their data space. It is unknown whether the validation of compliance with data space policies is purely objectively performed. There is insufficient information to assess whether a data space using the EUDI Wallet is open or not. Sovrin does not specify anything about the onboarding of parties that want to join a data space and, according to expert #E8, is not part of Sovrin.

The authorization policies within iSHARE, IDS, and Gaia-X are self-configurable. ISHARE supports fine and coarse-grained authorization where policies are defined based on XACML 3.0. In IDS the authorization policies are specified with ODRL similar to Gaia-X which also allows authorization policies to be specified with Rego. According to expert #E3 authorization will be done ABAC and will be configured by the data provider. Sovrin does not implement authorization regarding data accessibility of data consumers.

IdM Treatment / Requirement	iSHARE	IDS	Gaia-X	EUDI Wallet	Sovrin
#RQ1 Issuer Verification	Yes	Yes	Yes	Yes	No
#RQ2 Open Architecture	Yes	Yes	Yes	Yes	Yes
#RQ3 Proof of Request	No	Yes	Yes	?	N.A.
#RQ4 Reusable Credentials	No	Yes	Yes	Yes	Yes
#RQ5 Decentralized	No	No	Yes	?	Yes
#RQ6 Open Data Space	Yes	?	?	?	N.A.
#RQ7 Authorization Policies	Yes	Yes	Yes	Yes	N.A.

Table 9: Comparison available inter-organizational Identity Management Treatments Performed Medio 2022

6.1 Shortcomings and Improvements Available Inter-organizational IdM Treatments

The shortcomings addressed in this section are based on requirement satisfaction. There is insufficient information available regarding the EUDI Wallet to identify accurate shortcomings and improvements. To do so the EUDI Wallet must provide detailed information on the EUDI Wallet design and functionalities.

ISHARE falls short on Proof of Request, Reusable Credentials, and Decentralized. Proof of Request could be supported within the Authorization Registry. The Authorization Registry allows data providers to authorize specific entities or parties to access a specific (part of a) dataset. When a data consumer request access to data, the data provider request information via the respective API about whether that specific data consumer is authorized to access the respective data. In case the data consumer is not authorized, the Authorization Registry would return a message indicating that access should be denied. To support Proof of Request two things should be done. First, the Authorization Registry should store unauthorized requests with information regarding the requestor and the data it requested access to. Secondly, the Authorization Registry should provide an API that enables a data provider to request information regarding the unauthorized request allowing it to use that information as proof for unauthorized

requests.

To support Reusable Credentials, iSHARE should replace the access token with a token that does qualify as a credential and is reusable at different parties. They can implement something similar to what IDS does with the DAPS and the DAT. Currently, the iSHARE Satellite holds additional information about participants such as company name and whether a party is an active member of the data space. A data provider checks this information at the iSHARE Satellite before providing the data consumer with an access token. Instead, the iSHARE Satellite should function as a credential issuer similar to the DAPS. Allowing a data consumer to request a credential containing the information required by the data provider. The iSHARE Satellite should digitally sign the credential allowing the data provider to verify the issuer of the credential. The data consumer should digitally sign the credential allowing the data provider to verify the integrity of the message.

For iSHARE to become decentralized, the iSHARE Foundation has to give up its control over policies. Instead, it should enable participants of the network to propose policies and use a democratic voting system to propose proposals.

For IDS to become decentralized, the DAPS and ParIS should be controlled by more than one party. Both could be controlled by the data space; the data space can establish a data space foundation and have that foundation operate and control the DAPS and ParIS. Another possibility to make the DAPS decentralized is to have multiple (at least two) in a data space that are controlled by different parties. The DAPS itself will still be controlled by one party, but when multiple DAPS operate in a data space, credential issuance becomes decentralized; credential issuance is controlled by more than one party.

It remains unknown whether a data space in IDS is open or not. To ensure that the compliance assessment is conducted only objectively, the assessment process must be fully transparent. This allows both data space members and the audited party to review the process and verify whether the compliance review was validly performed. Compliance assessment should be conducted in a decentralized manner. The most appropriate way is for a data space foundation to conduct the compliance assessment.

Another possible improvement for IDS regards the DAT. The DAT are reusable for one hour and only for parties that are mentioned in the audience of the credential. In addition to a DAT token that has an expiration date, another type of DAT could be supported that has unlimited validity until the DAPS revokes the credential. This requires the DAPS to manage a revoked list of all revoked credentials or manage an active credential list of all active credentials. Which one to use depends on scalability and should be explored; a revoked list might eventually become greater than an active credential list. Instead of having an audience field for the credential, the data provider can specify a trusted DAPS list: a list of DAPS from which they trust issued credentials and, for example, what information these credentials should contain. As a result, the DAPS has less control over the issuance of credentials having a positive effect on the requirement Decentralized. In case a DAPS does not want to provide a credential to a data consumer, the data consumer can request the credential at a different DAPS (given that the data provider trusts multiple DAPS for the issuance of the respective credential).

It remains unknown whether a Gaia-X-based data space is open or not. A Gaia-X-based data space could build a service that automatically checks a party's compliance with the policies of the data space. The service demands specific verifiable credentials that reflect compliance with certain policies and provides a list of trusted issuers where those verifiable credentials can be requested. Once the verifiable credentials are acquired, the service can verify the verifiable credentials and issue a membership token accordingly. The compliance process performed by the service must be fully transparent allowing both data space members and the audited party to review the process and verify whether the compliance review was validly performed (i.e., objectively based). This service is referred to as the Transparent Data Space Compliance Assessment Service.

The degree of decentralization of Gaia-X demands attention. Both the issuer of compliance credentials and the issuer of membership credentials should be managed by the data space or a data space should have multiple issuers for the two types of credentials. For the former, the data space can establish a data space foundation and have that foundation issue the compliance and membership credentials. The latter still means that the issuer of compliance credentials and the issuer of membership credentials are controlled by one party, but when the compliance and membership credentials in a data space are issued by multiple issuers, credential issuance becomes decentralized; credential issuance is controlled by more than one party.

Sovrin lacks Issuer Verification, Proof of Request, Open Data Space, and Authorization Policies. Issuer Verification is not supported due to privacy reasons which is out of necessity. Proof of Request cannot be provided due to the

lack of authorization policies. Authorization is not part of Sovrin and would require to be supported in order to satisfy the requirement Authorization Policies. No literature was found to determine whether a data space is open and accepting or rejecting a party from joining the data space is not part of Sovrin according to expert #E8. Sovrin requires additional treatments to fulfill the requirements Proof of Request and Authorization Policies.

7 Discussion

7.1 Reflection on Chosen Research Methodology

A systematic literature review was conducted to ensure that existing scientific research is used and built upon. Five Identity Management (IdM) models were identified: Isolated Identity Management (IIM), Centralized Identity Management (CIM), Federated Identity Management (FIM), user-centric, and Decentralized Identity Management (DIM). The IdM models were compared based on the identified requirements defined with TNO experts. Resulting in DIM being the best fitting model that should be pursued in order to improve existing IdM solutions used during data sharing in logistics. Interviews were held with TNO experts to collect the inter-organizational IdM treatments that should be investigated. In addition to iSHARE Trust Framework (iSHARE), the predetermined IdM treatment to investigate, International Data Spaces (IDS), Gaia-X, the European Digital Identity Wallet (EUDI Wallet), and Sovrin Network (Sovrin) were investigated.

During the analyses and comparison of the available treatments, and interviews with experts, it became clear that the Gaia-X IdM already satisfies the requirements except for Open Data Space; it is unknown whether a Gaia-X-based data space is open or not. Within the available period of time, no design could be realized that guides how the requirement Open Data Space can be satisfied. Nevertheless, to satisfy the requirement Open Data Space, a Gaia-X-based data space could build a service checking the policies automatically. The service demands specific verifiable credentials that reflect compliance with certain policies and provides a list of trusted issuers where those verifiable credentials can be requested. Once the verifiable credentials are acquired, the service can verify the verifiable credentials and issue a membership token accordingly. To ensure that the compliance assessment is conducted objectively, the assessment process performed by the service must be fully transparent. The service should return all steps performed during the assessment process allowing both data space members and the audited party to review the process and verify whether the compliance review was validly performed.

The degree of decentralization of the Gaia-X IdM demands attention. Both the issuer of compliance credentials and the issuer of membership credentials should either be controlled by the data space or a data space should have multiple issuers for the two types of credentials. For the former, the data space can establish a data space foundation and have that foundation issue the compliance and membership credentials. The latter still means that the issuer of compliance credentials and the issuer of membership credentials are controlled by one party, but when the compliance and membership credentials in a data space are issued by multiple issuers, credential issuance becomes decentralized i.e., credential issuance is controlled by more than one party.

Experts #E6 and #E7 had observations regarding the requirement Open Data Space (#RQ6). They explained that a policy could be set that, for example, only businesses having more than 1 Billion dollars in revenue or only Dutch companies are allowed to join the data space. Following the definition of the requirement, this would mean that the data space is open but this is questioned by experts #E6 and #E7. Despite that, the requirement remained unchanged. Expert #E1 wants to prevent everyone from being able to join a data space and remove the ability to reject entities from joining the data space based on subjective reasoning. For example, DHL is part of a data space and is the one responsible for assessing whether a party may join the data space. PostNL requests to become a member of the data space but is a big competitor of DHL so DHL rejects PostNL from joining the data space. Therefore, the boundary of the ability to join a data space is set on the policies defined by the data space and the compliance assessment that must be assessed objectively.

Requirement #RQ6 can also be set for the data space instead of the IdM. It is chosen to be a requirement for the IdM because accepted parties require a data space identity and credentials (e.g., a credential representing membership of the data space).

The challenge of the DIM model is that users being in control over their identity information can still be questioned because of two reasons: 1) a verifier can store the identity information shared by the holder and 2) a holder retrieves a verifiable credential containing claims about its identity from an issuer meaning that issuer holds identity information about the holder. Selective disclosure and Zero-knowledge proof overcome the first reason. Selective disclosure allows a holder to select the attributes that he wants to share with a verifier. Zero-knowledge proof allows a holder to prove that he is above 18 years old without sharing his date of birth. The only thing that is being shared with the verifier is "yes, I am above 18 years old". The verifiable credential W3C specification applied by the Gaia-X IdM does support selective disclosure and Zero-knowledge proof. A possibility to address the second reason would be to enforce issuers by law to delete identity information after the holder requested the verifiable credential containing the

identity information.

7.2 Recommendations for Stakeholders

According to W. Hofman, research is kickstarted regarding the implementation of verifiable credentials and Decentralized Identifiers (DIDs) in the context of Basic Data Infrastructure (BDI). The United States Homeland Security Customs Border Protection Agency is already researching this area and has funded a portion of the specification published by W3C focused on "a linked data vocabulary for asserting verifiable credentials related to supply chain and other traceability information, similar to what is often referred to as 'provenance', including country of origin, chemical properties, mechanical properties, and other attributes of products and materials"[129]. Gaia-X already uses verifiable credentials and DIDs and thus aligns with that research. Also, the European Union (EU) has several initiatives that foster the Self-sovereign Identity (SSI) model: the European Blockchain Services Infrastructure, the European Self-Sovereign Identity Framework, and the EUDI Wallet. The goal of the European Commission is to have a secure European e-Identity that any citizen can use anywhere in Europe with technology to control what data and how data is used. All EU Member States must provide their citizens with an EUDI Wallet by 2024[19]. The EUDI Wallet is being developed by the EU as a way to improve the security and convenience of digital identity verification and authentication. With the intention to be a secure and easy-to-use platform for storing and accessing digital identity documents (e.g., eID cards, ePassports), as well as for performing online identity verification and authentication. It is being designed to work across the EU and to be compatible with a wide range of national and international identity systems, and part of the EU's Digital Single Market strategy[18] which aims to create a single market for digital goods and services within the EU. The goal of the EUDI Wallet is to make it easier for individuals and businesses to verify and authenticate identities online and to reduce the risk of identity fraud and other types of online crime[52]. As previously addressed, there is insufficient information available to assess whether the EUDI Wallet will provide for the stakeholder's needs. Implementing an inter-organizational IdM treatment, Gaia-X, that adopts the same IdM model and provides for the stakeholder's needs would be the best option for the stakeholder. Since Gaia-X is funded by the EU, it is likely that they will make the Gaia-X IdM interoperable with the EUDI Wallet. We see the possibility that the EUDI Wallet will be used as Organizational Credential Manager (OCM) and Personal Credential Manager (PCM) in Gaia-X.

The treatments investigated are continuously evolving especially IDS, Gaia-X, and the EUDI Wallet. The analyses in this research are performed in medio 2022. New changes are coming in the near future which impacts the results. The stakeholders should keep this in mind, follow the latest updates, and take them into account in their decision-making. As pointed out by experts #E4 and #E7, IDS is likely to follow and adopt the SSI model in the near future (1 to 1.5 years).

The following steps can be taken by the stakeholder:

- Follow the directions mentioned in section 8.4.
- Interview the end users Douane and Logistic Service Providers (LSPs) to collect new insights into the context and collect additional stakeholder goals. Based upon which new requirements could be defined that can be added to the analyses and comparison of the inter-organizational IdM treatments.
- Design the Transparent Data Space Compliance Assessment Service and demonstrate the design to experts in the field of logistics data spaces and IdM. Then collect expert opinions during interviews regarding the predicted effects of the interaction with the Transparent Data Space Compliance Assessment Service and the problem context, and verify whether the design satisfies the requirement Open Data Space.

Once the stakeholders made a decision about which treatment should be built and integrated with BDI, we suggest applying the TOGAF Architecture Development Method (ADM). The goal of TOGAF ADM is to create an enterprise architecture in line with the organization's vision and needs. Each phase in the model identifies the key activities and information to perform, to acquire the required knowledge to develop the enterprise architecture. The model consists of 9 phases that can be performed incrementally and iteratively. Following the steps results in enterprise architectures of the current situation (BDI implementing iSHARE), the future situation (BDI implementing the chosen treatment), and a migration plan broken down into phases where parts of the future situation are developed ensuring a seamless transition. The current and future situations both consist of a business architecture, an application architecture, and a technology architecture. The resulting models can be used for project management, as a frame of reference during the development of the future situation. Changes to the future situation and migration plan can be made during

the development if necessary. During the development of the future situation, it is recommended to apply the agile software management methodology. Applying agile increases the visibility of the work that is being done, when, and how increasing the transparency of project managers while improving accountability. It is easier to adapt to changes; after each sprint, a working proof of concept is shown to the end user which provides feedback that TNO can use to change the requirements and start a new sprint accordingly. The alignment within the team developing the treatment and between the team and the end user will be enhanced. This in turn will lead to high product quality increasing the value created for the end user and their satisfaction.

8 Conclusion

8.1 Main Conclusions

8.1.1 State-of-the-art Identity Management (IdM) Models

Research Objective 1 (RO1): Knowledge regarding IdM must be acquired to get an understanding about the state-of-the art, challenges, and causes of those challenges. Having this knowledge enhances communication with stakeholders and experts, and aids in designing an accurate inter-organizational IdM solution to achieve the primary research objective. Hence, the first research question is defined with two sub-research questions:

RQ1: What are the challenges posed by the state-of-the art IdM models?
SRQ1.1: What are the state-of-the-art IdM models?

There are five state-of-the-art IdM models. Isolated Identity Management (IIM), the first model, where the Service Provider (SP) performs Identification, Authentication, and Authorization (IAA), the deletion and modification of identities, stores the identity information of users, and provides the service. Then came Centralized Identity Management (CIM), which delegates SP functions to the Identity Provider (IdP). The IdP takes care of the IAA, the deletion and modification of identities, and stores the identity information of users. Federated Identity Management (FIM) extends CIM by linking different user identifiers within a federation and allows user authentication across Service Providers (SPs) in the federation using the same username-password combination. The user-centric IdM aims to give users control over their identity information by providing users with the ability to store their identity information in one place and control who can access what attributes of their identity information. According to Bouras et al.[10], the model failed because users did not control who could access what attributes of their identity information. The organizations where their identity information was stored had access to their identity information and could decide to delete a user's account. Decentralized Identity Management (DIM) aims to give users full control over their identity; the user owns his identity, has control over where it is kept, and decides for himself with whom he shares that identity (or parts of it). There are no parties other than the person himself who has exclusive control over the exchange of identity information and related data. The person himself takes the necessary information and decides whether or not to provide it to an SP.

SRQ1.2: What are the challenges experienced regarding the IdM models and what is causing them?

One of the biggest challenges is the single point of failure: the identity information is stored centrally at the SP or IdP. Users have no control over their identity information but the SP or IdP do. They manage the identity information at their central storage, allowing them to deactivate or delete a user account when they want. Users have to manage multiple username-password combinations for every SP or IdP. Although FIM reduces the number significantly, a user can be part of different data spaces for which he has to manage username-password combinations. Many users reuse their username-password combination, increasing the probability of security breaches. The IdP in FIM are still single points of failure. In case a IdP gets hacked, the identity information of all federated partners is exposed. Establishing trust between the user and IdP or SP is a challenge; the user does not know how the IdP or SP manages their identity information, whether this is done in a safe and secure way, and in compliance with General Data Protection Regulation (GDPR). Establishing trust between parties in a data space is challenging: parties may, for example, provide no assurance that the FIM process is controlled and processes are followed sufficiently, or no clear data ownership agreements are made resulting in inappropriate use of data and misunderstandings.

Although DIM tries to overcome most of these challenges, users being in control over their identity information can still be questioned because 1) a verifier can store the identity information shared by the holder and 2) a holder retrieves a verifiable credential containing claims about its identity from an issuer meaning that issuer holds identity information about the holder.

8.1.2 Inter-organizational IdM Requirements

Research Objective 2 (RO2): The collection of requirements from stakeholders regarding the inter-organizational IdM solution that improves existing IdM solutions used during data sharing in logistics. Hence, the second research question is defined:

RQ2: What are the requirements for an inter-organizational IdM solution that improves the existing solutions used during data sharing in logistics?

The requirements of the inter-organizational IdM solution that improves existing solutions used in logistics data spaces are:

- **Issuer Verification** (#RQ1, must have): a mechanism, system, or infrastructure should be available that allows a data provider (i.e., Logistic Service Provider (LSP)) to verify the issuer of a credential shown by a data consumer (i.e., an agent of Douane).
- **Open Architecture** (#RQ2, must have): the architecture of the IdM should be openly available to anyone who wants to use the IdM. The architecture should include technical specifications or guidelines of the IdM, allowing developers to build their own implementations.
- **Proof of Request** (#RQ3, must have): a data provider (i.e., LSP) must be able to prove that a user requested access to a resource at a specific time. In the event that a data consumer (i.e., an agent of Douane) submits an unauthorized request for data, the data provider is able to prove this event to the principal of the data consumer.
- **Reusable Credentials** (#RQ4, must have): if party A receives a credential from party B, this credential can also be used for identification at other parties such as party C, D, and E. This credential can be reused at other parties because parties C, D, and E trust the issuer of this credential, namely, party B.
- **Decentralized** (#RQ5, should have): more than one party can control a specific resource and the policies that govern the system's operations. If there are two parties in the data space, they should both supervise the resources and have a vote in the determination of policies for the functioning of the system. More parties mean a higher level of decentralization. The resources of the IdM that should be decentralized: credential issuance, the service providing information regarding members of a data space including the services that they offer, and granting access to a data space. Policies of the IdM include credentials, authentication, access control, data privacy, and security policies.
- **Open Data Space** (#RQ6, should have): the data space is open for participation by anyone complying with its policies. The compliance assessment must be assessed objectively.
- **Authorization Policies** (#RQ7, would have): during the process of assigning access rules to authenticated entities, a data provider is able to self-configure authorization policies and assign them to a data consumer regarding data accessibility. Authorization policies could, for example, be provided by Role-based Access Control (RBAC).

8.1.3 Best Fitting State-of-the-art IdM Model

Research Objective 3 (RO3): Knowing what state-of-the-art IdM model best fits the requirements of the research to have guidance about the IdM model that should be pursued during the design of the artifact. Hence, the third research question is defined:

RQ3: What state-of-the-art IdM model does best fit the requirements?

The IIM and CIM models cannot be used inter-organizational and do not satisfy the must-have requirements Issuer Verification (#RQ1), Open Architecture (#RQ2), and Reusable Credentials (#RQ4) and the should have requirements Decentralized (#RQ5) and Open Data Space (#RQ6). The FIM model does not satisfy the must-have requirement Issuer Verification (#RQ1) and the should-have requirement Decentralized (#RQ5). The DIM model satisfies all requirements and is the best fitting IdM model for the requirements of this research. As a result, the to-be-designed artifact, inter-organizational Identity Management for Data Sharing in Logistics (IdM4DSL), uses the DIM model, and only available inter-organizational IdM treatments that use the DIM model were investigated.

8.1.4 Existing Inter-organizational IdM Used in Logistics Data Spaces

Research Objective 4 (RO4): Knowledge regarding the existing inter-organizational IdM solutions is required, how they satisfy the identified requirements, and how they can be improved. Having this knowledge enhances communication with stakeholders and experts, and aids in designing an accurate inter-organizational IdM solution

that satisfies all identified requirements and improves existing IdM solutions used during data sharing in logistics. Hence, the fourth research question is defined with two sub-research questions:

RQ4: What inter-organizational IdM solutions are used during data sharing in logistics and to what extent do these solutions satisfy the identified requirements?

The inter-organizational IdM treatments suggested by experts that use the DIM model are International Data Spaces (IDS), Gaia-X, the European Digital Identity Wallet (EUDI Wallet), Sovrin Network (Sovrin), and IRMA. This research only covers the iSHARE Trust Framework (iSHARE) (pre-determined), IDS, Gaia-X, the EUDI Wallet, and Sovrin given the time period. Table 10 answers RQ4.

IdM Treatment / Requirement	Priority	iSHARE	IDS	Gaia-X	EUDI Wallet	Sovrin
#RQ1 Issuer Verification	Must	Yes	Yes	Yes	Yes	No
#RQ2 Open Architecture	Must	Yes	Yes	Yes	Yes	Yes
#RQ3 Proof of Request	Must	No	Yes	Yes	?	N.A.
#RQ4 Reusable Credentials	Must	No	Yes	Yes	Yes	Yes
#RQ5 Decentralized	Should	No	No	Yes	?	Yes
#RQ6 Open Data Space	Should	Yes	?	?	?	N.A.
#RQ7 Authorization Policies	Could	Yes	Yes	Yes	Yes	N.A.

Table 10: Answer to RQ4

SRQ4.1: How do these inter-organizational IdM solutions satisfy the identified requirements?

In iSHARE the client assertion qualifies as a credential and is created by the data consumer self. The issuer of the client assertion is verified by verifying the digital signature set with the private key of the issuer's X.509 certificate with the associated public key and an encryption algorithm (#RQ1). The technical specifications of iSHARE are openly available[70] and include the governance and technologies that should be used in order to perform the IAA process (#RQ2). The iSHARE Satellite validates the onboarding process based on the iSHARE and data space policies and guarantees trustworthy onboarding. The iSHARE Satellite agreed to all the rules of iSHARE which are legally binding so it is trusted that their onboarding process is impartial[68] (#RQ6). ISHARE supports both fine and coarse-grained authorization[95]. The policies are based on XACML 3.0, a standard for representing and evaluating access control policies that support Attribute-based Access Control (ABAC) or in combination with RBAC[33]. These policies are self-configurable by the data provider for any resource (all, a selection, or a combination) (#RQ7).

In IDS the client assertion and Dynamic Attribute Token (DAT) qualify as credentials. The client assertion is created by the data consumer self. The issuer of the client assertion is verified by verifying the digital signature set with the private key of the issuer's X.509 certificate with the associated public key and an encryption algorithm. The DAT is issued by the Dynamic Attribute Provisioning Service (DAPS) if the client assertion is accurate. The issuer of the DAT is verified by verifying the digital signature set with the private key of the issuer's X.509 certificate with the associated public key and an encryption algorithm (#RQ1). The International Data Spaces Association (IDSA) has published a reference architecture[92] providing technical specifications and guidelines including the technical specifications about the IdM (#RQ2). IDS specifies a Clearing House that monitors and logs data transactions and data value chains, and monitors policy enforcement. It can be informed at any time about transactions being done in a data space including failed transactions, for example, caused by unauthorized requests (#RQ3). A DAT expires in one hour but can be reused within that period of time. The DAT consists of an audience field that can be set to a data space as a whole or specific data providers. If the DAT is not expired, it can be reused at other data providers that are part of the audience (#RQ4). The policies for authorization can be configured by the data owner and data provider, and specified with Open Digital Rights Language (ODRL) (#FRQ7).

In Gaia-X the verifiable credential qualifies as a credential. The issuer of a verifiable credential is verified by verifying

the digital signature set with the private key of the issuer's X.509 certificate with the associated public key and an encryption algorithm (#RQ1). The Gaia-X Association published an architecture[2] providing technical specifications and guidelines also covering the IdM (#RQ2). Gaia-X specifies a Contract Logging Service that receives logging messages (data provided, data received, policy enforced, and policy-violating messages) during a transaction to trace each event. In case a service consumer has no rights to access a resource, the transaction will fail and a policy-violating message (including the details) will be stored at the Contract Logging Service. The participants of the transaction and, if necessary, a third eligible party may query the stored information both during and after the transaction (#RQ3). Verifiable credentials can be reused by all parties that trust the issuer of that credential (#RQ4). Policies are set by the Gaia-X Association, a combination of more than 340 companies that use a democratic voting system in order to determine the policies of Gaia-X. The data space determines additional policies they find relevant to apply. Credential issuance is done by issuers trusted by the data providers. In case an issuer is not willing to issue a credential to the data consumer, he is able to request the credential at a different issuer. Participant information is available at the Federated Catalogue, a publicly distributed, immutable, permissionless database with a decentralized infrastructure and the capacity to automatically execute code. The data space together decides whether a party may join the data space based on its compliance with the data space policies (#RQ5). Authorization policies will be defined in a domain-specific language (e.g., ODRL and Rego)[14]. The authorization policies can be self-configured and negotiated between the data provider and data consumer via the Service Agreement Service (#RQ7).

In the EUDI Wallet the Personal Identifiable Data (PID), Qualified Electronic Attestation of Attributes (Qualified EAA), and Non-qualified Electronic Attestation of Attributes (Non-qualified EAA) qualify as credentials. The PID providers shall make information accessible for relying parties to verify the PID's validity as might be the case with Qualified EAA and Non-qualified EAA providers. This implies that it is possible to verify the issuer of a credential, otherwise, relying parties will not be able to know what provider to contact to verify the credentials' validity (#RQ1). The eIDAS expert group published the European Digital Identity Architecture and Reference Framework (ARF)[50] providing the objectives, roles of the actors of the ecosystem, the wallet's functional and non-functional requirements (technical specifications), and the potential building blocks. The ARF will be updated during the process and developed to a full architecture and reference framework including all technical specifications of the IdM (#RQ2). One of the functionalities the EUDI Wallet shall provide is select, combine, and share credentials with relying parties. A functionality to combine a credential indicates that a user can combine attributes from different credentials to prove its identity which implies that credentials are reusable (#RQ4).

The Sovrin Foundation has published multiple documents consisting of specifications and guidelines[21][35][37]. The Sovrin Governance Framework Working Group published the Sovrin Trust Framework[51]. Sovrin is a deployment of Hyperledger Indy[36] that provides documentation about their repositories and consist of technical specifications and guidelines[60][61][62][63][64] (#RQ2). Sovrin itself does not use or issue credentials. However, Sovrin enables issuers to define the schema and define issuer-specific credentials, allows verifiers to verify a verifiable credential, and enables the exchange of verifiable credentials between identity owners. Verifiable credentials can be reused at as many parties as the number of parties who trust the issuer of the respective credential (#RQ4). The Sovrin Foundation, a not-for-profit consortium, determines the policies of the Sovrin Network using a democratic system to determine the policies. Hyperledger Indy is controlled by the Hyperledger Foundation consisting of more than 140 organizations, all having the ability to propose contributions to the Hyperledger Foundation's technical codebase. Decisions are made using a democratic voting system[34]. Credential issuance, the service providing information regarding members of a data space including the services that they offer, and granting access to a data space is all not part of Sovrin (#RQ5).

SRQ4.2: How can the inter-organizational IdM solutions be improved?

ISHARE can support Proof of Request by adopting two things. First, the Authorization Registry should store unauthorized requests with information regarding the requestor and the data it requested access to. Secondly, the Authorization Registry should provide an Application Programming Interface (API) that enables a data provider to request information regarding the unauthorized request that he can use as proof for unauthorized requests. To support Reusable Credentials, the iSHARE Satellite should function as a credential issuer allowing a data consumer to request a credential containing the information required by the data provider. The iSHARE Satellite should digitally sign the credential allowing the data provider to verify the issuer of the credential. The data consumer should digitally sign the credential allowing the data provider to verify the integrity of the message. For iSHARE to become decentralized, it should enable participants of the network to propose policies and use a democratic voting system to propose proposals.

IDS can become decentralized when the DAPS is controlled by the data space or a data space has multiple DAPS. Another possible improvement for IDS is the reusability of the DATs. IDS should support a second type of DAT that

allows for unlimited validity until the DAPS revokes the credential. This requires the DAPS to manage a revoked list of all revoked credentials or manage an active credential list of all active credentials. Which one to use depends on scalability and should be explored; a revoked list might eventually become greater than an active credential list. Furthermore, the data providers should specify a trusted DAPS list. This way the reusability of the DAT is not dependent on the audience field specified by the DAPS but it can be reused at as many data providers that accept the credential. As a result, the DAPS has less control over the issuance of credentials having a positive effect on the requirement Decentralized (#RQ5). In case a DAPS does not want to provide a credential to a data consumer, the data consumer can request the credential at a different DAPS (given that the data provider trusts multiple DAPS for the issuance of the respective credential).

For IDS to ensure that compliance assessment is conducted only objectively, the assessment process must be fully transparent. This allows both data space members and the audited party to review the process and verify whether the compliance review was validly performed. Compliance assessment should be conducted in a decentralized manner. The most appropriate way is for a data space foundation to conduct the compliance assessment.

The degree of decentralization of Gaia-X demands attention. Both the issuer of compliance credentials and the issuer of membership credentials should be managed by either the data space or a data space should have multiple issuers for the two types of credentials. For the former, the data space can establish a data space foundation and have that foundation issue the compliance and membership credentials. The latter still means that the issuer of compliance credentials and the issuer of membership credentials are controlled by one party, but when the compliance and membership credentials in a data space are issued by multiple issuers, credential issuance becomes decentralized; credential issuance is controlled by more than one party.

It remains unknown whether a Gaia-X-based data space is open or not. A Gaia-X-based data space could build a Transparent Data Space Compliance Assessment Service, a service that automatically checks a party's compliance with the policies of the data space. The service demands specific verifiable credentials that reflect compliance with certain policies and provides a list of trusted issuers where those verifiable credentials can be requested. Once the verifiable credentials are acquired, the service can verify the verifiable credentials and issue a membership token accordingly. The compliance process performed by the service must be fully transparent allowing both data space members and the audited party to review the process and verify whether the compliance review was validly performed (i.e., objectively based).

Issuer Verification is not supported by Sovrin due to privacy reasons which do not require improvement; it is a necessity resulting from using Distributed Ledger Technology (DLT). Authorization is not part of Sovrin which is why both the requirement Proof of Request and Authorization Policies are not satisfied. Accepting or rejecting a party from joining the data space is not part of Sovrin. Sovrin requires additional treatments to fulfill the requirements Proof of Request, Authorization Policies, and Open Data Space.

8.1.5 Design IdM4DSL

Research Objective 5 (RO5): Design an accurate inter-organizational IdM solution in line with the stakeholder's requirements that improves the existing IdM solutions used during data sharing in logistics. To do so we need to know what an infrastructure of an inter-organizational IdM solution consists of that satisfies the identified requirements. RO4 can be used to identify reusable characteristics and components and provides the improvements that require to be solved. Hence, the fifth research question has been defined:

RQ5: What does the infrastructure of an inter-organizational IdM solution consists of that satisfies all requirements?

The infrastructure of an inter-organizational IdM solution that satisfies all requirements, the IdM4DSL, consists of:

- Applied by Gaia-X based on W3C standards:
  - **Decentralized Identifier (DID)** - a permanent identifier that never changes and represents an entity known as the subject. A DID refers to a DID document[127].
  - **DID documents** - contains information about the subject such as creation time, the public key, method on how to resolve the DID to receive the associated DID document, and more. The DID document is immutable, persistent, and fully owned and controlled by the DID owner[127].

  - **DID resolver** - a service that is able to look up and return the associated DID document for a given DID named the verifiable data registry[127].
  - **Self-Descriptions** - claims that describe a participant, resource, or service offering in a machine-readable manner[128].
  - **Verifiable credentials** - claims that describe a participant, resource, or service offering in a machine-readable manner and are verified by a third party (i.e., issuer). Verifiable credentials can also be verified against the Gaia-X Trust Framework to receive proof of compliance. Then the digital signature of the party hosting the compliance tool is added to the verifiable credential[128].

- Specified by Gaia-X:
  - **Self-Description Wizard** - an interactive web form that guides non-technical end users in writing valid Self-Descriptions.
  - **Gaia-X Trust Framework** - the set of rules that specify the requirements for participating in the Gaia-X Ecosystem including rules about Self-Descriptions and verifiable credentials. Users maintain full control over their decisions thanks to these rules, which offer common management and fundamental levels of interoperability between different ecosystems. The Gaia-X Trust Framework is defined to create trust between participants of the ecosystem[9].
  - **Compliance tool** - a tool that automatically checks whether a verifiable credential complies with the Gaia-X Trust Framework. When a verifiable credential complies with the Gaia-X Trust Framework it adds its digital signature to the verifiable credential.
  - **Organizational Credential Manager (OCM)** - a cloud Self-sovereign Identity (SSI) wallet used by organizations to store, present, and issue verifiable credentials.
  - **Personal Credential Manager (PCM)** - a personal SSI wallet app to store and present verifiable credentials.
  - **Federated Catalogue** - a public distributed, immutable, permissionless database with a decentralized infrastructure and the capacity to automatically execute code. Each data space runs its own Federated Catalogue and is used to store Self-Descriptions enabling participants to discover and select providers and their service offerings.
  - **Gaia-X Registry** - a public distributed, immutable, permissionless database with a decentralized infrastructure and the capacity to automatically execute code. The backbone of the ecosystem governance stores information such as the list of issuers, the results of the issuer validation processes, and potential revocation of issuers' identity. It enables, among others, verification of whether the verifiable credential issuer's identity is from a Gaia-X complaint issuer and verification of whether any signature was revoked.
  - **Data Exchange Logging Service** - a service that provides evidence that data has been submitted and received, and data usage policies were enforced or violated[25].
  - **Data Contract Service** - a service that enables data transactions in a secure, trusted, and auditable manner. The service allows the data consumer and data provider to negotiate the data asset usage policies for the planned data exchange. Offers and counteroffers can be sent resulting in agreement or rejection. The contract is packaged in a human and a machine-readable format based on ODRL and can be retrieved for later reference[24].
  - **Trust Anchors (i.e., Issuers)** - third parties that verify the claims made in a Self-Description and issue a verifiable credential if accurate. Examples of issuers are membership issuers and compliance issuers. Both the membership issuer and compliance issuer should be managed by the data space or a data space should have multiple issuers for the two types of credentials.
  - **Trusted list** - a list of trusted issuers defined by a data provider. The list indicates to a data consumer from what issuers the data provider accepts (i.e., trusts) verifiable credentials.

- One additional component is required that should be designed by the stakeholder in future research:
  - **Transparent Data Space Compliance Assessment Service** - a service that automatically checks a party's compliance with the policies of the data space. The service demands specific verifiable credentials that reflect compliance with certain policies and provides a list of trusted issuers where those verifiable credentials can be requested. Once the verifiable credentials are acquired, the service can verify the verifiable credentials and issue a membership token accordingly. The compliance process performed by the service must be fully transparent allowing both data space members and the audited party to review the process and verify whether the compliance review was validly performed (i.e., objectively based).

8.1.6 Concluding Remarks

What inter-organizational IdM solution is open, decentralized, and flexible in the use of standards and improves existing IdM solutions used during data sharing in logistics?

The existing IdM solutions used during data sharing in logistics can be improved by implementing the Gaia-X IdM treatment of which the infrastructure is explained in the previous section. One addition to the current design is needed, namely, the Transparent Data Space Compliance Assessment Service: a service that automatically checks a party's compliance with the policies of the data space. The service demands specific verifiable credentials that reflect compliance with certain policies and provides a list of trusted issuers where those verifiable credentials can be requested. Once the verifiable credentials are acquired, the service can verify the verifiable credentials and issue a membership token accordingly. The compliance process performed by the service must be fully transparent allowing both data space members and the audited party to review the process and verify whether the compliance review was validly performed (i.e., objectively based). Adding the Transparent Data Space Compliance Assessment Service to the current design makes sure the Open Data Space requirement is also satisfied contributing to the stakeholder's goal of having an open IdM solution.

A data provider can specify a trusted issuer list from which he accepts verifiable credentials. When shown a verifiable credential by a data consumer, the data provider can verify the issuer via an interface of the Gaia-X Registry. If the verification is accurate, this means that a verifiable credential is issued by an issuer that the data provider trusts. This creates trust that the data consumer is who he claims to be. The verifiable credentials are reusable by as many parties that trust the issuer of the respective credential contributing to the stakeholder's goal to reduce the number of credentials a user has to manage.

Gaia-X has an Open Architecture that is technology agnostic and contributes to the stakeholder's goal of having an IdM solution that is open and flexible in the use of standards.

The resources and policies of the Gaia-X IdM can be controlled by more than one party; the policies are determined by the Gaia-X Association and additional policies can be set by the data space. The issuers that are supported depend on the trusted list of the data provider. The data consumer is limited to verifiable credentials issued by these trusted issuers but has the ability to choose which issuer to request a verifiable credential from. This way there is not one party in control of the issuance of credentials. In case an issuer is not willing to issue a credential to the data consumer, he is able to request the credential at a different issuer. The Federated Catalogue can be used by consumers to discover and select providers and their service offerings. The policies and resources of the IdM can be controlled by the data space contributing to the stakeholder's goal of having a decentralized IdM solution and the goal to remove the single point of failure.

The Gaia-X IdM treatment supports an Open Data Space; the data space membership credential issuer issues a membership token if a party complies with the policies set by the data space. Contributing to the stakeholder's goal of having an open IdM solution.

The Data Exchange Logging Service stores events happening during transactions between a data consumer and data provider including policy-violating messages such as an unauthorized data consumer requesting access to data. It provides evidence that data has been submitted and received, and that data usage policies were enforced or violated. The participants of the transaction and, if necessary, a third eligible party may query the stored information both during and after the transaction. A LSP can use this evidence to prove to Douane that someone in their organization submitted unauthorized requests. If Douane is able to identify who this person is and take appropriate actions to prevent the unauthorized request from happening it will contribute to the stakeholder's goal of creating trust.

The Gaia-X Trust Framework constitutes of rules that specify the requirements for participating in the Gaia-X Ecosystem including rules about Self-Descriptions and verifiable credentials. Users maintain full control over their decisions thanks to these rules, which offer common management and fundamental levels of interoperability between different ecosystems. The Gaia-X Compliance service implements the Gaia-X Trust Framework. Thus, the participants follow the same rules, creating mutual trust.

8.2 Contributions

8.2.1 Scientific Contribution

This research provides a view on the state-of-the-art IdM models with their challenges and satisfaction of technical requirements contributing to openness, decentralization, and flexibility. IdM researchers can use it to assess which IdM model is relevant for their use case or to use it as a basis for in-depth research about the evolution of IdM models or a specific model. Research gaps have been identified that can be addressed. A systematic approach is provided to define requirement terminologies. The technical requirements are defined with experts in the field of Basic Data Infrastructure (BDI), data spaces, freight transport, logistics, and IdM showcasing their vision on what IdM solution improves the existing IdM solutions used during data sharing in logistics. The research provides a detailed overview of existing inter-organizational IdM solutions used during data sharing in logistics which is useful for researchers in data spaces, logistics, data spaces, and IdM. Furthermore, it provides an analyses and comparison of inter-organizational IdM treatments that are open, decentralized, and flexible in the use of standards based on technical documentation available medio 2022 and interviews with experts of BDI, data spaces, freight transport, logistics, IdM, and the IdM treatments. Lastly, the study on IDS and Gaia-X can be used as a basis for research on data-sharing infrastructures.

8.2.2 Practical Contribution

This research provides a view on the state-of-the-art IdM models with their challenges and satisfaction of technical requirements contributing to openness, decentralization, and flexibility. The challenges and causes of each IdM model make practitioners aware and help them in decision-making, for example, to assess which IdM model is relevant for their use case or to identify challenges they need to address in their own IdM solution.

Within freight transport and logistics documents, such as the bill of lading are becoming increasingly digital and need to be verified by authorities. But there is no inter-organizational IdM solution in place that is open, decentralized, and flexible in the use of standards. The comparison of IdM models helps the stakeholder to understand what and why the DIM model should be pursued. The analyses of inter-organizational IdM treatments helps the stakeholder understand their design, characteristics, and/or components that satisfy the requirements, and how the treatments can be improved. The comparison of inter-organizational IdM treatments helps the stakeholder to understand the similarities and differences between the treatments. Lastly, this thesis makes tangible what inter-organizational IdM treatment does best fit the stakeholder needs, why, how it contributes to their goals, how it should be improved, and provides guidelines regarding the next steps. Lastly, the study on IDS and Gaia-X can be used as a basis for research on data-sharing infrastructures.

8.3 Limitations

The study on IdM models originates from a qualitative methodology. The literature collected was analyzed with a systematic selection process inspired by the guidelines of Kitchenman et al.[76]. However, as only one researcher was involved in the selection, extraction, and summarization processes, they are subject to the researcher.

It was not possible to conduct interviews with the intended end users of BDI; the Douane and Logistic Service Providers (LSPs). There is a possibility that the end users have different requirements than those collected from the TNO experts.

It was not possible to conduct interviews outside TNO with experts on the investigated inter-organizational IdM treatments. This could have led to useful additional information.

Due to the limited availability of expert #E3 only one interview has been conducted to fill knowledge gaps about the EUDI Wallet. Useful additional information could have been gathered to further deepen the analyses on the EUDI Wallet.

No measurements have been defined to measure the degree of requirement satisfaction. This research only provides insight into whether an IdM treatment satisfies a requirement or not based on the criteria used in their terminology. Having measurements provides better insight into the degree of satisfaction, enhances the comparison between the

treatments, and could be used to provide a better argument for the conclusion.

8.4 Directions for Future Work

The directions for future research are as follows:

- The research gaps mentioned in section 2.2.5 can be addressed.
- This research only provides insight into whether an IdM treatment satisfies a requirement or not. To have insight into the degree of requirement satisfaction, measurements can be defined to measure, for example, the degree of decentralization or openness.
- Within Europe, interaction will take place between customs and LSPs around the globe. Hence, multiple legal frameworks should be taken into account which was out of the scope of this research.
- EIDAS is a regulation adopted by the European Union (EU) that aims to create a single market for electronic identification and trust services within the EU. It establishes a legal framework for the use of electronic identification and authentication systems, such as electronic signatures and seals, in order to facilitate the cross-border provision of online services within the EU. It applies to both natural persons and legal entities, and is intended to facilitate electronic interactions between citizens, businesses, and public authorities within the EU. Hence, knowing whether the IdM treatments are eIDAS compliant, the degree of compliance, or what changes should be made to the IdM treatments to become eIDAS compliant would be of great value.
- The IdM treatment IRMA can be investigated and included in the comparison.
- The position papers such as Technical Convergence Discussion Document[31], Data Space Business Committee - Position Papers[43], and Gaia-X and IDS[93] could be explored and addressed to gain insight into the vision of other experts regarding the positioning of Gaia-X and IDS.
- The treatments investigated are continuously changing especially, IDS, Gaia-X, and the EUDI Wallet. The analyses in this research were performed in medio 2022. New changes are coming in the near future which might impact the analysis performed.
- During researching existing inter-organizational IdM treatments another possible treatment was encountered, Keyrock from Fiware. This treatment could be explored in future research.
- Explore selective disclosure and Zero-knowledge proof and how this solves the ability of a verifier to store the identity information shared by the holder.
- Explore the impact of and how to deal with the fact that an issuer holds identity information about the holder in order for a holder to retrieve a verifiable credential containing claims about its identity from an issuer, going against users being in control over their identity information.

References

[1] Aldosary, Maha and Alqahtani, Norah. "A Survey on Federated Identity Management Systems Limitation and Solutions". In: *International Journal of Network Security & Its Applications* 13 (2021), pp. 43–59. DOI: 10.5121/ijnsa.2021.13304.

[2] Association, Gaia-X. *Gaia-X - Architecture Document - 22.04 Release*. Tech. rep. 2022. URL: https://gaia-x.eu/wp-content/uploads/2022/06/Gaia-x-Architecture-Document-22.04-Release.pdf (visited on 10/17/2022).

[3] Association, International Data Spaces. *IDS Participant Information Service (IDS-ParIS)*. 2021. URL: https://github.com/International-Data-Spaces-Association/IDS-G/tree/main/Components/IdentityProvider/ParIS (visited on 02/15/2023).

[4] Association, International Data Spaces. *Dynamic Attribute Provisioning Service (DAPS)*. 2022. URL: https://github.com/International-Data-Spaces-Association/IDS-G/blob/main/Components/IdentityProvider/DAPS/README.md (visited on 10/11/2022).

[5] Association, International Data Spaces. *The Association*. 2022. URL: https://internationaldataspaces.org/we/the-association/ (visited on 10/10/2022).

[6] Balasubramaniam, Sriram et al. "Identity management and its impact on federation in a system-of-systems context". In: 2009, pp. 179–182. DOI: 10.1109/SYSTEMS.2009.4815794.

[7] Bartolomeu, Paulo C. et al. "Self-Sovereign Identity: Use-cases, Technologies, and Challenges for Industrial IoT". In: 2019, pp. 1173–1180. DOI: 10.1109/ETFA.2019.8869262.

[8] Bertino, Paci and Shang. "Keynote 2: Digital Identity Protection - Concepts and Issues". In: 2009. DOI: 10.1109/ARES.2009.176.

[9] Binzer, M. et al. *GXFS - IDM & Trust Architecture Overview*. Tech. rep. 2021.

[10] Bouras, Mohammed Amine et al. "Distributed Ledger Technology for eHealth Identity Privacy: State of The Art and Future Perspective". In: *Sensors* 20 (2020). DOI: 10.3390/s20020483.

[11] Brunner, Clemens et al. "DID and VC:Untangling Decentralized Identifiers and Verifiable Credentials for the Web of Trust". In: *2020 the 3rd International Conference on Blockchain Technology and Applications*. Association for Computing Machinery, 2020, pp. 61–66. DOI: 10.1145/3446983.3446992.

[12] C. Majaski D. Clemon, S. Clarine. *Distributed Ledgers: Definition, How They're Used, and Potential*. 2021. URL: https://www.investopedia.com/terms/d/distributed-ledgers.asp (visited on 10/16/2022).

[13] Cao, Yuan and Yang, Lin. "A survey of Identity Management technology". In: 2010, pp. 287–293. DOI: 10.1109/ICITIS.2010.5689468.

[14] Characteristics, Working Group Service. *Gaia-X Service Ontology*. 2022. URL: https://gaia-x.gitlab.io/technical-committee/service-characteristics/widoco/service/service.html (visited on 01/05/2023).

[15] Cocco, Luisanna, Tonelli, Roberto, and Marchesi, Michele. "Blockchain and Self Sovereign Identity to Support Quality in the Food Supply Chain". In: *Future Internet* 13.12 (2021). ISSN: 1999-5903. DOI: 10.3390/fi13120301.

[16] Commission, European. *Register of Commission expert groups and other similar entities — eIDAS Expert Group*. 2022. URL: https://ec.europa.eu/transparency/expert-groups-register/screen/expert-groups/consult?lang=en (visited on 02/11/2023).

[17] Corda. *Corda — Leading DLT Platform for Regulated Industries*. 2022. URL: https://www.corda.net/ (visited on 10/10/2022).

[18] Council, European. *Digital single market for Europe*. 2020. URL: https://www.consilium.europa.eu/en/policies/digital-single-market/ (visited on 02/01/2023).

[19] Council, European. *European digital identity (eID): Council makes headway towards EU digital wallet, a paradigm shift for digital identity in Europe*. 2022. URL: https://www.consilium.europa.eu/en/press/press-releases/2022/12/06/european-digital-identity-eid-council-adopts-its-position-on-a-new-regulation-for-a-digital-wallet-at-eu-level/ (visited on 02/01/2023).

[20] Cycle, Fuel. *A Quick Guide to Semi-Structured Interviews*. 2022. URL: https://fuelcycle.com/blog/a-quick-guide-to-semi-structured-interviews/ (visited on 12/19/2022).

[21] D. Reed J. Law, D. Hardman. *The Technical Foundations of Sovrin*. Tech. rep. 2016. URL: https://sovrin.org/wp-content/uploads/2018/03/The-Technical-Foundations-of-Sovrin.pdf (visited on 11/01/2022).

[22] Damienbod, D. *Challenges to Self Sovereign Identity*. 2021. URL: https://damienbod.com/2021/10/11/challenges-to-self-sovereign-identity/ (visited on 10/19/2022).

[23] Dao, Delta. *Result Presentation Gaia-X Hackathon Ocean*. Gaia-X Hackathon. 2021. URL: https://gitlab.com/gaia-x/gaia-x-community/gx-hackathon/gx-hackathon-1/-/wikis/uploads/f5cf7944935d85d4ba0ae6ba7aae27df/1st_Hackathon_Results_Compute_to_Data_and_DLT.pdf (visited on 10/28/2022).

[24] Data, Gaia-X European Association for and AISBL, Cloud. *Software Requirements Specification for Gaia-X Federation Service — Sovereign Data Exchange Data Contract Service SDE.DELS*. Tech. rep. 2021.

[25] Data, Gaia-X European Association for and AISBL, Cloud. *Software Requirements Specification for Gaia-X Federation Service — Sovereign Data Exchange Data Exchange Logging Service SDE.DELS*. Tech. rep. 2021.

[26] Data, Gaia-X European Association for and AISBL, Cloud. *Software Requirements Specification for Gaia-X Federation Service Sovereign Data Exchange Data Exchange Logging Service SDE.DELS*. Tech. rep. 2021.

[27] Data, Gaia-X European Association for and AISBL, Cloud. *Association - Gaia-X: A Federated Secure Data Infrastructure*. 2022. URL: https://gaia-x.eu/who-we-are/association/ (visited on 10/16/2022).

[28] Design Foundation, Privacy by. *IRMA uitleg*. n.d. URL: https://privacybydesign.foundation/irma-uitleg/ (visited on 01/16/2023).

[29] Dictionary, Cambridge. *identification definition*. 2023. URL: https://dictionary.cambridge.org/dictionary/english/identification (visited on 01/02/2023).

[30] Doyle, Alison. *What Is a Semi-Structured Interview?* 2022. URL: https://www.thebalancemoney.com/what-is-a-semi-structured-interview-2061632 (visited on 12/19/2022).

[31] DSBA, Data Spaces Business Alliance -. *Technical Convergence Discussion Document*. Tech. rep. 2022. URL: https://data-spaces-business-alliance.eu/wp-content/uploads/dlm_uploads/DSBA-Technical-Convergence.pdf (visited on 01/31/2023).

[32] Dunphy, Paul and Petitcolas, Fabien A.P. "A First Look at Identity Management Schemes on the Blockchain". In: *IEEE Security & Privacy* 16.4 (2018), pp. 20–29. DOI: 10.1109/MSP.2018.3111247.

[33] E. van der Harst, D. Hoppenbrouwer. *XACML 3.0*. 2018. URL: https://ishareworks.atlassian.net/wiki/spaces/IS/pages/74875018/XACML+3.0 (visited on 11/28/2022).

[34] foundation, Hyperledger. *Charter*. 2022. URL: https://www.hyperledger.org/about/charter (visited on 11/17/2022).

[35] Foundation, Sovrin. *The Inevitable Rise of Self-Sovereign Identity*. 2017. URL: https://sovrin.org/wp-content/uploads/2018/03/The-Inevitable-Rise-of-Self-Sovereign-Identity.pdf (visited on 11/22/2022).

[36] Foundation, Sovrin. *Developers*. 2022. URL: https://sovrin.org/developers/ (visited on 01/05/2023).

[37] Foundation, Sovrin. *Sovrin DID Method Specification*. 2022. URL: https://sovrin-foundation.github.io/sovrin/spec/did-method-spec-template.html (visited on 11/16/2022).

[38] Foundation, Sovrin. *The Sovrin Foundation*. 2022. URL: https://github.com/sovrin-foundation/sovrin (visited on 11/16/2022).

[39] Fraunhofer. *Challenges and Potentials of a Logistic Data Space*. 2019. URL: https://internationaldataspaces.org/wp-content/uploads/IDSA-LC-position_paper.pdf (visited on 02/13/2023).

[40] Fraunhofer. *MY DATA Control Technologies*. 2022. URL: https://www.dataspaces.fraunhofer.de/en/software/usage-control/mydata.html (visited on 01/05/2023).

[41] Frederiksen, Tore Kasper et al. "Identity management: State of the art, challenges and perspectives". In: *IFIP Advances in Information and Communication Technology* 576 LNCS (2020), pp. 45–62. DOI: 10.1007/978-3-030-42504-3_4.

[42] Fuchs, Christian. "Industry 4.0: The Digital German Ideology". In: *TripleC* 16 (2018), pp. 280–289. DOI: 10.31269/vol16iss1pp280-289.

[43] Gaia-X. *Data Space Business Committee - Position Papers*. Tech. rep. 2021. URL: https://gaia-x.eu/wp-content/uploads/files/2021-08/Gaia-X_DSBC_PositionPaper.pdf (visited on 01/31/2023).

[44] Gataca. *Self-Sovereign Identity (SSI) 101: Decentralized Identifiers (DIDs) & Verifiable Credentials (VCs)*. 2021. URL: https://gataca.io/blog/self-sovereign-identity-ssi-101-decentralized-identifiers-dids-verifiable-credentials-vcs/ (visited on 08/12/2022).

[45] Gataca. *SSI Essentials: Everything you need to know about Decentralized Identities*. 2021. URL: https://gataca.io/blog/ssi-essentials-everything-you-need-to-know-about-decentralized-identity/ (visited on 08/16/2022).

[46] Gilani, Komal et al. "Self-sovereign Identity Management Framework using Smart Contracts". In: 2022. DOI: 10.1109/NOMS54207.2022.9789831.

[47] Goodell, Geoff and Aste, Tomaso. "A Decentralized Digital Identity Architecture". In: *Frontiers in Blockchain* 2 (2019). ISSN: 2624-7852. DOI: 10.3389/fbloc.2019.00017. URL: https://www.frontiersin.org/articles/10.3389/fbloc.2019.00017.

[48] Griffith, J. *What you Need to Know about Centralized vs Decentralized Identity Management*. 2021. URL: https://www.pingidentity.com/en/resources/blog/post/centralized-decentralized-identity-management.html (visited on 05/27/2022).

[49] Group, Alpega. *The types of trucks circulating on our roads*. 2023. URL: https://teleroute.com/en-en/ (visited on 02/13/2023).

[50] group, eIDAS expert. *European Digital Identity Architecture and Reference Framework*. Tech. rep. 2022. URL: https://ec.europa.eu/newsroom/dae/redirection/document/83643 (visited on 11/28/2022).

[51] Group, Sovrin Trust Framework Working. *Sovrin Provisional Trust Framework*. Tech. rep. 2017. URL: https://www.evernym.com/wp-content/uploads/2017/07/SovrinProvisionalTrustFramework2017-03-22.pdf (visited on 11/02/2022).

[52] Group, Thales. *All your ID credentials at hand: welcome to the Digital ID Wallet*. 2022. URL: https://www.thalesgroup.com/en/markets/digital-identity-and-security/government/identity/digital-identity-services/digital-id-wallet (visited on 02/01/2023).

[53] Guo, Chunfang and Wang, Ying. "Application of federated identity management in ERP system". In: *2008 IEEE International Conference on Service Operations and Logistics, and Informatics*. Vol. 2. 2008, pp. 1971–1974. DOI: 10.1109/SOLI.2008.4682855.

[54] GXFS. *Gaia-X Federation Services*. 2022. URL: https://www.gxfs.eu/ (visited on 01/09/2023).

[55] H.J.M. (Harrie) Bastiaansen, TNO. *ISHARE as generic trust framework capability*. Tech. rep. 2022. URL: https://topsectorlogistiek.nl/wp-content/uploads/2022/07/TNO-2022-R11094-Report-iSHARE-as-generic-capability-1.pdf (visited on 09/16/2022).

[56] Hardt, D. *RFC 6749: The OAuth 2.0 Authorization Framework*. 2022. URL: https://www.rfc-editor.org/rfc/rfc6749 (visited on 10/10/2022).

[57] Hoeven, G. *Overzicht Internationale Ontwikkelingen Federatief Zakelijk Data Delen*. 2022. URL: https://www.slideshare.net/gerardvanderhoeven/220222federatiefdatadelentopsectorlogistiekpdf (visited on 09/20/2022).

[58] Identity, Ping. *Identity as a Service — What is IDaas?* n.d. URL: https://www.pingidentity.com/en/resources/content-library/articles/identity-as-a-service-idaas.html (visited on 02/11/2023).

[59] Imageware. *Identification, Authentication, Authorization - What's The Difference*. 2021. URL: https://imageware.io/identification-authentication-authorization-difference/ (visited on 09/07/2022).

[60] Indy, Hyperledger. *How tos*. 2018. URL: https://github.com/hyperledger/indy-sdk/tree/main/docs/how-tos (visited on 11/16/2022).

[61] Indy, Hyperledger. *Indy HIPE — Hyperledger Indy HIPE documentation*. 2018. URL: https://hyperledger-indy.readthedocs.io/projects/hipe/en/latest/text/index.html (visited on 11/16/2022).

[62] Indy, Hyperledger. *Indy SDK — Hyperledger Indy SDK documentation*. 2018. URL: https://hyperledger-indy.readthedocs.io/projects/sdk/en/latest/docs/index.html (visited on 11/15/2022).

[63] Indy, Hyperledger. *Welcome to Hyperledger Indy Node's documentation! — Hyperledger Indy Node documentation*. 2018. URL: https://hyperledger-indy.readthedocs.io/projects/node/en/latest/index.html (visited on 11/15/2022).

[64] Indy, Hyperledger. *Welcome to Indy Plenum's documentation! — Indy Plenum documentation*. 2018. URL: https://hyperledger-indy.readthedocs.io/projects/plenum/en/latest/index.html (visited on 11/16/2022).

[65] Indy, Hyperledger. *Plenum Byzantine Fault Tolerant Protocol*. 2022. URL: https://github.com/hyperledger/indy-plenum (visited on 11/17/2022).

[66] Institute, Corporate Finance. *Distributed Ledgers*. 2023. URL: https://corporatefinanceinstitute.com/resources/cryptocurrency/distributed-ledgers/ (visited on 01/14/2023).

[67] Internet, Next Generation. *Identifier — eSSIF-Lab*. 2022. URL: https://essif-lab.github.io/framework/docs/terms/identifier/ (visited on 10/11/2022).

[68] iSHARE. *Satellite explained*. 2020. URL: https://ishare.eu/ishare-satellite-explained/ (visited on 11/10/2022).

[69] iSHARE. *Technical Standards — dev.ishareworks.org 1.0.0 documentation*. 2020. URL: https://dev.ishare.eu/introduction/standards.html (visited on 11/14/2022).

[70] iSHARE. *Welcome to iSHARE — dev.ishareworks.org 1.0.0 documentation*. 2020. URL: https://dev.ishare.eu/index.html (visited on 11/14/2022).

[71] iSHARE. *Data of parties in Network*. 2022. URL: https://ishare.eu/data-of-parties-in-ishare-network/ (visited on 06/15/2022).

[72] iSHARE. *For Data Spaces*. 2022. URL: https://ishare.eu/ishare/benefits/for-data-spaces/ (visited on 06/15/2022).

[73] iSHARE. *GitHub - iSHAREScheme/AuthorizationRegistry: Reference implementation of the Authorization Registry as specified in the iSHARE Scheme*. 2022. URL: https://github.com/iSHAREScheme/AuthorizationRegistry (visited on 01/02/2023).

[74] Jensen, Jostein. "Benefits of federated identity management - A survey from an integrated operations viewpoint". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 6908 LNCS (2011), pp. 1–12. DOI: 10.1007/978-3-642-23300-5_1.

[75] Jensen, Jostein. "Federated Identity Management Challenges". In: 2012, pp. 230–235. DOI: 10.1109/ARES.2012.68.

[76] Kitchenham, Barbara et al. "Systematic literature reviews in software engineering – A systematic literature review". In: *Information and Software Technology* 51.1 (2009), pp. 7–15. DOI: https://doi.org/10.1016/j.infsof.2008.09.009.

[77] Lågbu, Amund. "A concept for improving ICT tools as support for morning meetings in the oil and gas industry". In: 2015.

[78] Lim, Don and Palvia, Prashant C. "EDI in strategic supply chain: Impact on customer service". In: *International Journal of Information Management* 21.3 (2001), pp. 193–211. DOI: 10.1016/S0268-4012(01)00010-X.

[79] Liu, Yang et al. "Blockchain-based identity management systems: A review". In: *Journal of Network and Computer Applications* 166 (2020), p. 102731. ISSN: 1084-8045. DOI: https://doi.org/10.1016/j.jnca.2020.102731. URL: https://www.sciencedirect.com/science/article/pii/S1084804520302058.

[80] LoginTC. *What is Biometric Authentication and How Does It Work?* 2022. URL: https://www.logintc.com/types-of-authentication/biometric-authentication/ (visited on 02/11/2023).

[81] Lutkevich, B. *What is an Identity Provider? Definition from SearchSecurity*. 2021. URL: https://www.techtarget.com/searchsecurity/definition/identity-provider (visited on 05/27/2022).

[82] Lutkevich, B. *Access control*. 2022. URL: https://www.techtarget.com/searchsecurity/definition/access-control (visited on 02/11/2023).

[83] M.C.G. Keijzer R.W. Knops, F.B.J. Grapperhaus. *Nederlandse digitaliseringsstrategie*. 2018. URL: https://open.overheid.nl/repository/ronl-c2025613-f72f-4ef7-92e8-8df626140ae0/1/pdf/kamerbrief-over-nederlandse-digitaliseringsstrategie.pdf (visited on 09/26/2022).

[84] Maler, E. *To Succeed In Decentralizing Digital Identity, Focus On Relationships First*. 2021. URL: https://www.forbes.com/sites/forbestechcouncil/2021/06/16/to-succeed-in-decentralizing-digital-identity-focus-on-relationships-first/?sh=473489c61824 (visited on 09/19/2022).

[85] Maler, Eve and Reed, Drummond. "The Venn of Identity: Options and Issues in Federated Identity Management". In: *Security & Privacy, IEEE* 6 (2008), pp. 16–23. DOI: 10.1109/MSP.2008.50.

[86] McLaren, Tim, Head, Milena, and Yuan, Yufei. "Supply chain collaboration alternatives: Understanding the expected costs and benefits". In: *Internet Research* 12 (2002), pp. 348–364. DOI: 10.1108/10662240210438416.

[87] Meinke, k. *3rd Gaia-X hackathon Bootstrapping Gaia-X Ecosystem*. Gaia-X Hackathon. 2022. URL: https://gitlab.com/gaia-x/gaia-x-community/gx-hackathon/gx-hackathon-3/-/wikis/uploads/7007172e1971fac6f532e24a4c52008e/20220328_Bootstrapping_Decentralized_Gaia-X_Ecosystem.pdf (visited on 10/28/2022).

[88] Network, Smart Connected Supplier. *Organisatie van de Stichting SCSN*. n.d. URL: https://smart-connected.nl/nl/over-scsn/stichting-scsn/organisatie-van-de-stichting-scsn (visited on 01/09/2023).

[89] Nieuwenhuizen Wijbenga, C. van. *Kamerstuk 26643, nr. 581 — Overheid.nl*. 2018. URL: https://zoek.officielebekendmakingen.nl/kst-26643-581.html (visited on 09/26/2022).

[90] Nieuwenhuizen Wijbenga, C. van. *Kamerstuk 31409, nr. 186 — Overheid.nl*. 2018. URL: https://zoek.officielebekendmakingen.nl/kst-31409-186.pdf (visited on 09/26/2022).

[91] Nieuwenhuizen Wijbenga, C. van. *Kamerstuk 26643, nr. 719 — Overheid.nl*. 2020. URL: https://zoek.officielebekendmakingen.nl/kst-26643-719.html (visited on 09/26/2022).

[92] Otto, B. et al. *Reference Architecture Model*. Tech. rep. 2019. URL: https://internationaldataspaces.org//wp-content/uploads/IDS-Reference-Architecture-Model-3.0-2019.pdf (visited on 10/10/2022).

[93] Otto, B. et al. *Gaia-X and IDS*. Tech. rep. 2021. URL: https://internationaldataspaces.org/wp-content/uploads/dlm_uploads/IDSA-Position-Paper-GAIA-X-and-IDS.pdf (visited on 01/31/2023).

[94] P. Leijnse, M. Scheers. *Domeinarchitectuur Identiteiten & Toegang*. 2022. URL: https://www.surf.nl/files/2023-01/hosa-domeinarchitectuur-iam-versie-1.0.pdf (visited on 01/16/2023).

[95] P. Nijs, R. Westenberg. *Use case: M2M interaction (with fine-grained authorization)*. 2019. URL: https://ishareworks.atlassian.net/wiki/spaces/IS/pages/70222246/Use+case+M2M+interaction+with+fine-grained+authorization (visited on 11/28/2022).

[96] Pauer, A. et al. *Data exchange as a first step towards data economy*. 2018. URL: https://www.pwc.de/en/digitale-transformation/data-exchange-as-a-first-step-towards-data-economy.pdf (visited on 02/13/2023).

[97] Plan, Product. *MoSCoW Prioritization*. 2022. URL: https://www.productplan.com/glossary/moscow-prioritization/ (visited on 01/14/2023).

[98] Platforms, FEDeRATED Network of. "Vision Document". In: (Dec. 2019), p. 116. URL: http://federatedplatforms.eu/index.php/library/item/visionreport-milestone1.

[99] Platforms, FEDeRATED Network of. "Pilots/LivingLabs Scoping". In: (Oct. 2020), p. 112. URL: http://www.federatedplatforms.eu/index.php/library/item/federated-milestone-4-report-pilots-livinglab-scoping.

[100] R.Awati. *Federated identity management (FIM)*. 2021. URL: https://www.techtarget.com/searchsecurity/definition/federated-identity-management (visited on 05/27/2022).

[101] Rieks, J. *Actor — eSSIF-Lab*. 2022. URL: https://essif-lab.github.io/framework/docs/terms/actor/ (visited on 10/27/2022).

[102] Rieks, J. *Agent — eSSIF-Lab*. 2022. URL: https://essif-lab.github.io/framework/docs/terms/agent/ (visited on 10/27/2022).

[103] Rieks, J. *Authority (Centralized or Decentralized) — eSSIF-Lab*. 2022. URL: https://essif-lab.github.io/framework/docs/terms/authority/ (visited on 10/27/2022).

[104] Rieks, J. *Credential — eSSIF-Lab*. 2022. URL: https://essif-lab.github.io/framework/docs/terms/credential/ (visited on 10/27/2022).

[105] Rieks, J. *Holder — eSSIF-Lab*. 2022. URL: https://essif-lab.github.io/framework/docs/terms/holder/ (visited on 10/27/2022).

[106] Rieks, J. *Identification Pattern — eSSIF-Lab*. 2022. URL: https://essif-lab.github.io/framework/docs/terms/pattern-identification/ (visited on 10/27/2022).

[107] Rieks, J. *Identify — eSSIF-Lab*. 2022. URL: https://essif-lab.github.io/framework/docs/terms/identify/ (visited on 10/27/2022).

[108] Rieks, J. *Identity — eSSIF-Lab*. 2022. URL: https://essif-lab.github.io/framework/docs/terms/identity/ (visited on 10/27/2022).

[109] Rieks, J. *Issuer — eSSIF-Lab*. 2022. URL: https://essif-lab.github.io/framework/docs/terms/issuer/ (visited on 10/27/2022).

[110] Rieks, J. *Partial identity — eSSIF-Lab*. 2022. URL: https://essif-lab.github.io/framework/docs/terms/partial-identity/ (visited on 10/27/2022).

[111] Rieks, J. *Principal — eSSIF-Lab*. 2022. URL: https://essif-lab.github.io/framework/docs/terms/principal/ (visited on 10/27/2022).

[112] Rieks, J. *Transaction — eSSIF-Lab*. 2022. URL: https://essif-lab.github.io/framework/docs/terms/transaction/ (visited on 10/27/2022).

[113] Rieks, J. *Trust — eSSIF-Lab*. 2022. URL: https://essif-lab.github.io/framework/docs/terms/trust/ (visited on 10/27/2022).

[114] Rieks, J. *Verifier — eSSIF-Lab*. 2022. URL: https://essif-lab.github.io/framework/docs/terms/verifier/ (visited on 10/27/2022).

[115] Rieks, J. *Verify — eSSIF-Lab*. 2022. URL: https://essif-lab.github.io/framework/docs/terms/verify/ (visited on 10/27/2022).

[116] Rosencrance, L. *identity management (ID management)*. 2021. URL: https://www.techtarget.com/searchsecurity/definition/identity-management-ID-management (visited on 10/20/2022).

[117] Security, RSI. *Challenges of Managing Personally Identifiable Information*. 2019. URL: https://blog.rsisecurity.com/challenges-of-managing-personally-identifiable-information/ (visited on 05/27/2022).

[118] Smith, Don. "The challenge of federated identity management". In: *Network Security* 2008.4 (2008), pp. 7–9. DOI: https://doi.org/10.1016/S1353-4858(08)70051-5.

[119] Smith, Heather A. and McKeen, James D. "The identity management challenge". In: *Communications of the Association for Information Systems* 28.1 (2011), pp. 169–180. DOI: 10.17705/1cais.02811.

[120] Snyder, Hannah. "Literature review as a research methodology: An overview and guidelines". In: *Journal of Business Research* 104 (2019), pp. 333–339. ISSN: 0148-2963. DOI: https://doi.org/10.1016/j.jbusres.2019.07.039.

[121] Stad, H. *ministerie-wil-nu-echt-doorpakken-met-digitalisering*. 2023. URL: https://www.logistiek.nl/189514/ministerie-wil-nu-echt-doorpakken-met-digitalisering (visited on 02/13/2023).

[122] Standards, National Institute of and Technology. *non-repudiation - Glossary — CSRC*. 2006. URL: https://csrc.nist.gov/glossary/term/non_repudiation (visited on 11/10/2022).

[123] TNO. *Provide Artifact*. 2022. URL: https://tno-tsg.gitlab.io/playground/provider/artifact/ (visited on 01/05/2023).

[124] TNO. *Mission and Strategy*. URL: https://www.tno.nl/en/about-tno/mission-and-strategy/ (visited on 06/15/2022).

[125] TNO. *Organisation*. URL: https://www.tno.nl/en/about-tno/organisation/ (visited on 06/15/2022).

[126] VMware. *What is Identity Management? — VMware Glossary*. 2022. URL: https://www.vmware.com/topics/glossary/content/identity-management.html (visited on 10/01/2022).

[127] w3c. *Decentralized Identifiers (DIDs) v1.0*. 2022. URL: https://www.w3.org/TR/did-core/ (visited on 11/29/2022).

[128] w3c. *Verifiable Credentials Data Model v1.1*. 2022. URL: https://www.w3.org/TR/vc-data-model/ (visited on 11/28/2022).

[129] w3c. *Traceability Vocabulary v0.0*. 2023. URL: https://w3c-ccg.github.io/traceability-vocab/ (visited on 02/01/2023).

[130] Wang, Yingli, Potter, Andrew, and Naim, Mohamed. "An exploratory study of electronic logistics marketplaces and its impact on customised logistics". In: (Jan. 2007). URL: https://www.poms.org/conferences/poms2007/CDProgram/Topics/full_length_papers_files/007-0265.pdf.

[131] Wieringa, Roel J. In: (2014), p. 332. DOI: 10.1007/978-3-662-43839-8.

[132] Wohlin, Claes. "Guidelines for Snowballing in Systematic Literature Studies and a Replication in Software Engineering". In: Association for Computing Machinery, 2014. DOI: 10.1145/2601248.2601268.

Appendices

A Terms

Actor
Actors are the ones involved in an interaction to negotiate and execute transactions on behalf of a party. For the negotiation and execution actors make use of their parties' knowledge as main guidance and may use knowledge of other parties to fill in gaps or provide additional information, as required. *For the elaborate description see [101]*

Agent
An actor who performs an activity on behalf of another party is said to be that party's agent. In this context, we say that the actor fulfils the role of an agent for that party. *For the elaborate description see [102]*

Authority
An authority is a party that other parties (have to) obey when it comes to making decisions, promoting ideas, enforcing regulations, etc. A distinction is made between two kinds of authority:

- centralized authority is the capacity or right to issue commands, make binding judgments, and compel compliance. This form of power disregards other people's inherent autonomy
- decentralized authority is the freedom to make judgements, formulate ideas, set norms, etc. that other parties to the authority will adopt and adhere to because they believe it is in their own best interests to do so

For the elaborate description see [103]

Credential
A credential is a collection of claims (i.e., assertions) to which additional information is added (e.g., date of issuing, subject name). It also contains a number of proofs, most commonly a proof of provenance (i.e. evidence that the data was produced on behalf of a particular party) and a proof of integrity (i.e. evidence that data has not been tampered with since its issuance). *For the elaborate description see [104]*

Data Sovereignty
"A natural person's or corporate entity's capability of being entirely self-determined with regard to its data"[92]

Data Space
A data spaces is a group of collaborating parties that exchange data in order to reach a common goal. The collaborating parties determine common policies to enable secure data exchange and guaranteeing data sovereignty. A data space is a type of federation i.e., a group of companies or organizations that are united to share resources and potentially collaborate on certain projects.

Freight Transport
Freight transport is the movement of goods or cargo from one place to another by means of a vehicle, such as a truck, train, or ship. Freight transport is a crucial part of the global economy, as it allows goods to be transported over long distances, often across international borders, and enables businesses to move raw materials, finished products, and other goods efficiently and cost-effectively. Freight transport plays a vital role in the supply chain, connecting manufacturers, distributors, and retailers to customers and enabling the smooth flow of goods and materials throughout the world.

Holder
A holder is a functional component that manages presentation requests received from the verifier. This usually entails finding the requested information in the principal's wallet and using it to create a presentation (=response). If the required credential is not in the wallet, the holder can negotiate a transaction with the specified issuer in order to gain the necessary credential which can then be kept in the wallet and utilized in the presentation once it has been acquired. *For the elaborate description see [105]*

Identity
The identity (of an entity) is the union of all partial identities with which that entity is associated. It represents the knowledge that all parties have about that entity as a whole. *For the elaborate description see [108]*

Issuer
"An issuer is a (functional) component that implements the capability to construct credentials from data objects, according to the content of its principal's issuer-policy (specifically regarding the way in which the credential is to be digitally signed), and pass it to the wallet-component of its principal allowing it to be issued." *For the elaborate description see [109]*

Issuer Policy
An issuer policy is a digital policy that permits an operational issuer component to operate in line with their principal's goals.

Logistics
Logistics is the process of organizing and conducting the effective storage and transportation of goods from their point of origin to their destination (point of consumption). This sector is seen as being filled with entities (i.e., things such as people, organizations, containers) that exist. Entities differ from each other in terms of characteristics and are therefore divided into well-defined categories:

1. Party An entity that sets its objectives, maintains its knowledge, and uses that knowledge to pursue its objectives in an autonomous (sovereign) manner. More information can be found at eSSIF-lab
2. Actor Someone or something that can act, i.e. actually do things, execute actions, such as people or machines. More information can be found at eSSIF-lab
3. Jurisdiction The composition of a (non-empty) set of objectives, one scope, one legal system and one party that operates the legal system within that scope. More information can be found at eSSIF-lab

Although logistic parties are autonomous and in principle can do as they please, they live in an ecosystem with many other parties that are also autonomous. Logistic parties require to accept one another because of the interactions they have and the affect an interaction may have on the parties' knowledge and other effects.

Logistic Service Provider
A party that provides a service in logistics and can be used by other parties or actors. Examples of logistic SPs are maritime shipping companies, marine terminals, and depots.

Partial Identity
A partial identity is the knowledge that a particular party possesses about a specific entity (referred to as subject of the partial identity). *For the elaborate description see [110]*

Principal
A principal (of an actor) is a party that the actor is acting on behalf of. The party is referred to as fulfilling the position of principal for that actor in the setting in which the actor is acting. Additionally, the actor acts as the party's agent. *For the elaborate description see [111]*

Transaction
A transaction is an exchange of goods, services, funds, or data between some parties. These parties are known as the participants of the transaction. A transaction is split into three phases:

1. Negotiation phase The actors involved in the transaction exchange data in order to create a contract that details the terms of the transaction. Either a commitment decision is made by the actors or the transaction is terminated
2. Execution phase Actors of the same party work to fulfill the obligations of the contract. The result is either work completion or actors that gave up
3. Acceptance phase Actors involved in the transaction exchange data and either accept the results or escalate (e.g. start a lawsuit against the other party)

For the elaborate description see [112]

Trust
Trust is the (un)conscious determination of a party that X is actually who or what it says it is. To trust is to believe in something to the point where one can depend on it to be true. One can rely on or trust their own judgements or conscience.
In the example, the verifier needs to trust the holder otherwise it is unlikely that the verifier will share the requested data. Trust is not given but rather something that changes over time; parties (un)consciously and continuously decide about. Parties are autonomous which is why trust is highly subjective. Therefore, the concept of 'trusted registries' and 'trusted issuers' that do not take this subjectivity into account in essence are (centralized) authorities and denies the fact that parties are autonomous. Those views are acknowledged however not followed in this paper[113].

Verifier
A verifier is a functional component that implements the ability to ask peer agents to present information from credentials and to verify such responses in line with its principal's verifier policy. *For the elaborate description see [114]*

Verifier Policy
A verifier policy is a digital policy that allow operational verifier components to operate in line with their principal's goals.

Verify
Verification (of data) is the act, by or on behalf of a party, of ascertaining if the data is accurate (i.e., comes from the party who wrote it), timely (i.e., not out-of-date), and complies with other requirements that apply to its structure. It does not imply that whatever the data depicts is genuinely correct, real, justified, or truthful. *For the elaborate description see [115]*

B Process of Defining Requirements

At first, four interviews were held with expert #E1 and #E2 to collect the requirements of an IdM solution reminding them of the context and the goals collected in the previous interviews. Resulting in a list of requirements, their terminologies, and classified into functional and non-functional requirements. This classification was made because we wanted to evaluate the IdM solutions based on both the functioning of the system and their performance. In between the first four interviews an significant number of changes have been made. The terminologies were validated by expert #E3 who concluded that the terms were unclear and could be interpreted in multiple ways. He helped setting up a structure in order to define the terminologies as specified in section 3. Secondly, four interviews were held with expert #E1 and #E2 following that structure resulting in four functional (Reusable Credentials, Roles of Authorization, Non-repudiation, and Issuer Credentials Registry) and five non-functional requirements (Open-source, Decentralized, Trust by Default, Interoperable, Technology Independence). This list was presented to expert #E3, #E4 and #E5 in individual interviews with each. According to expert #E3 and #4 non-repudiation is very difficult to achieve. It is the "assurance that the sender of information is provided with proof of delivery and the recipient is provided with proof of the sender's identity, so neither can later deny having processed the information[122]". As an example, expert #E3 and #E4 explained that one could be provided with proof of delivery but this does not mean that the receiver is not able to deny that he did not receive the information, or does not mean that the information was actually received by the receiver. Expert #E3 illustrated this problem with the example of him having ordered a book at bol.com. He received a message that his book was delivered (proof of delivery) while he did not receive the book. The support service of bol.com told him that this happened more frequently and he had to wait. One day later, expert #E3 received the book. As a package deliverer I encountered a situation where I delivered the package, the person received a proof of delivery but called DHL that the package was not delivered. The person tried to get a refund from his money while also receiving the package.
According to expert #E3 the requirement Trust by Default cannot be obtained. He stated that trust does not results from using a system because every party is autonomous in who or what he trusts which is why trust is highly subjective. Therefore he suggested to remove the requirement.
According to expert #E4 and #E5, the requirement Interoperable and Technology Independence should be replaced with Open Architecture which is also in line with the goal of an open IdM solution (see section 1.2.2). They explained that systems are not interoperable by itself but rather inherit interoperability by having an open architecture providing the standards to use and technical specifications to develop an implementation. In case developers follow these standards and technical specifications their implementation becomes interoperable with other implementations that follow the same set of standards and technical specifications. The same goes for Technology Independence, systems are in general not independent of technology however, when the architecture is open architecture, developers can implement the specifications with the technology they want.

The feedback was presented and processed during two additional interviews with expert #E1 and #E2. Non-repudiation was replaced with Proof of Request with as additional reason that proof of delivery was seen as out of scope for an IdM. After authenticating, and authorizing the sender (i.e. holder), the IdM is done. Then access to the requested resource will either be granted or not. Delivering the resource is not part of the IdM but belongs to the functioning of the data space. Therefore, the IdM does not require to provide proof of delivery. Non-repudiation was replaced by Proof of Request in order to realize the goal of an SP being able to prove that an unauthorized entity request access to data as specified in section 1.2.2. During the discussion about replacing the requirements Interoperable and Technology Independence with Open Architecture it became clear that Open-source was not required either. Open-source refers to a type of software whose source code is available to the general public, meaning anyone can access, modify, and distribute the source code for any purpose. Expert #E1 and #E2 decided that the source code does not necessarily need to be open as long as long as the architecture is open allowing

developers to build an implementation. Thus the requirements Interoperable, Technology Independence, and Open-source was replaced with Open Architecture. Both experts agreed on removing Trust by Default. At the end of the two interviews doubts arose regarding the requirement Issuer Credential Registry and an idea arose about adding the requirement Open Federation. Because of that another interview with expert #E1 and #E2 was conducted resulting in changing the requirement Issuer Credential Registry into Issuer Verification and adding the requirement Open Federation. The requirement Issuer Credential Registry was defined as "a registry must be available in which members of the data sharing ecosystem are able to search for issuers and the credentials they provide. This allows members to know which issuer to contact to request a certain credential and helps verifiers to search for the public key of an issuer required during the verification process for decryption of the provided credential by an agent". It was changed into Issuer Verification because both experts were concerned about realizing the goal of an SP being able to verify the issuer of a credential. An issuer credential registry might be a solution to realize that goal but is not a requirement. The requirement Open Federation was defined as "the federation is open for participation by anyone complying with its policies. If a party complies with its policies, it cannot be rejected from joining the federation". This requirement was defined to support the goal of having an open IdM solution. Both experts referred to Bitcoin as an example where everyone can join the network if they comply with the policies of the network.

A second round of feedback was collected from expert #E3, #E4, and #E5 in individual interviews with each. The feedback given was mainly about the terminologies and partly about the requirements themselves. During this research, the view of expert #E1 is followed which was explained to the experts and they understood. Two additional experts (#E6 and #E7) were interviewed (each individually). Expert #E6 showed interest in the research so an interview was scheduled to explain my research. During that interview he provided feedback regarding the requirement Open Federation mentioning that a policy could be set that only businesses that have more than 1 Billion dollar in revenue or Dutch companies are allowed to join the federation. Following the definition of the requirement, this would mean that the federation is open about which he raised his doubts. Expert #E1 advised me to schedule an interview with expert #E7 to provide him with an explanation about my research. During that interview he provided feedback regarding the requirement Open Federation raising the exact same doubts as expert #E6. I explained both experts that expert #E1 wants to prevent everyone from being able to join a data space and the ability of rejecting entities from joining the data space because of personal reasons and his vision should be followed. They both understood why the requirements was defined but wanted to raise awareness of their doubts.

The feedback was presented to expert #E1 and #E2, accepted, and processed. The naming of the requirement Roles of Authorization was changed into Authorization Policies because the authorization policies does not have to be role-based but could also be attributed-based. Instead of Open Federation, the requirement was changed into Open Data Space since the context is specifically about data spaces and not federations.

Eventually the requirements were classified based on priority instead of functional and non-functional. The IdM treatments that are going to be explored in further research are most likely to be proof-of-concepts. As a consequence, measuring non-functional requirements might not be possible. Additionally, at this point the stakeholder wants to know whether an IdM treatment satisfies the requirement or not and is not concerned with, for example, how decentralized the IdM treatment is.

Issuer Verification is a must because the Douane (or the LSP) wants to know who issued a credential provided by a party in order to determine the validity ("Do we trust this credential? Yes it is issued by Government X). Open Architecture is a must because the architecture of IdM should be openly available such that TNO is able to build their own implementation. This also means that the architecture should include technical specifications and guidelines as specified in the requirement. Reusable Credentials is a must because it mitigates the amount of credentials a party has to manage. Decentralized and Open Data Space are should haves because the IdM solution we are looking for is not restricted to have these functionalities. However, from the vision of decentralization and openness they should actually be realized. The authorization policies requirement explicitly talks about self-configurable policies. It is prioritized as a would have because it is enough to have common authorization policies and therefore self-configurable policies have no priority.

C Expert Interviews to Fill Knowledge Gaps

C.1 iSHARE Trust Framework

What qualifies as a credential within iSHARE?
While looking at the definition of a credential that you apply both the X.509 certificate and client assertion qualify as credential. But, the X.509 certificate is not used as a credential but only used to manage an identity and to secure communication. The client assertion is used by the data consumer to identify himself to the data provider so as proof of its identity.

Who is the credential issuer?
The credential is produced by the client application of the data consumer. Thus the data consumer using the credential is also the issuer of the credential.

What does the credential contains?
The client assertion includes six mandatory fields: iss (the issuer of client assertion), sub (the subject of the client assertion), aud (the audience for which the client assertion is created), jti (a unique identifier used to prevent reuse of the client assertion), iat (the time when the client assertion was issued), and exp (the time when the client assertion expires).

What is the reusability of the credential?
The audience for which the credential is created is specifically set to the data provider that a data consumer interacts with. The credential can be reused for the audience as long as it is not expired but it cannot be reused at other data providers.

What process is followed in order for a data consumer to receive access to a resource of a data provider?
For a data consumer to receive access to data of a data provider a two-step process is followed. First the data consumer identifies himself to the data provider. If the data consumer is who he say he is, the data provider provides an access token. With that access token, the data consumer can request access to the data of the data provider which is the second step in the process.

How is the identity of the data consumer authenticated?
The whole process starts with the data consumer requesting an access token from the data provider. This communication goes via HTTP and uses the OAuth 2.0 standard. The request consist of multiple headers and parameters of which I will mention only the relevant ones. One of the parameters is the client assertion, the credential. The client assertion is presented to the data provider as proof of the client's identity. The client assertion is encoded and signed with the JSON Web Signature (JWS) and encrypted using an encryption algorithm. The header of the request includes the encryption algorithm used and the X.509 certificate public key that corresponds to the key that is used to digitally sign the client assertion. When the data provider receives the request, he decodes the request resulting in the header, payload, and some signature. The data provider verifies the signature by using the X.509 certificate public key in the header. When the signature is verified then the identity of the data consumer is authenticated. In addition, the data provider verifies the payload by, for example, checking whether the payload complies with the iSHARE specifications, checking whether the client assertion is not received before, checking time of issuance and expiration time, checking whether the audience contains his EORI number, and checking whether the iss and sub field contain the same EORI number. When all checks have been successful, then the data provider checks at the iSHARE Satellite whether the EORI number (from the iss and sub field) is an active member in the data space. If the EORI number is an active member of the data space, the data provider issues an access token to the data consumer. Now the data consumer can request access to the data by providing the access token in the header and including the actual question (e.g. "can I have access to dataset X?") in the payload. The data provider decodes the request, verifies the access token and provides access to the data accordingly.

How can a data provider prove an unauthorized request was submitted within iSHARE?
The credential is exchanged between the data consumer and data provider when the data consumer requests access to a resource of the data provider. The credential consists of claims about the data consumer and is signed with the JWS standard to ensure non-repudiation of these claims. If the data provider stores the request and received credential it can prove that an unauthorized requests was submitted. But iSHARE has nothing specified that enables a data provider to prove an unauthorized request was submitted.

Can a party complying with all policies still be denied by the iSHARE Satellite from joining the data space?
No, the iSHARE Satellite validates the onboarding process based on the iSHARE and data space policies. It agreed to all the rules of iSHARE and is legally required to guarantee a trustworthy onboarding process.

C.2 International Data Spaces

The empirical knowledge questions were asked to expert #E4, #E5, and #E6. Their answers were very similar and are therefore summarized into one answer.

What qualifies as a credential within IDS?
While looking at the definition of a credential that you apply the X.509 certificate, client assertion, and DAT qualify as credential. But, the X.509 certificate is not used as a credential but only used to manage an identity and to secure communication.

Who is the credential issuer?
The client assertion is produced by the client application of the data consumer. Thus the data consumer using the client assertion is also the issuer of the client assertion.

The DAT is issued by the DAPS that manages dynamic attributes linked to participants identities.

What does the credential contains?
The client assertion includes six mandatory fields: iss (the issuer of client assertion), sub (the subject of the client assertion), aud (the audience for which the client assertion is created), jti (a unique identifier used to prevent reuse of the client assertion), iat (the time when the client assertion was issued), and exp (the time when the client assertion expires).

The DAT includes iss (the issuer of client assertion), sub (the subject of the client assertion), aud (the audience for which the client assertion is created), iat (the time when the client assertion was issued), exp (the time when the client assertion expires), scope (determines the set of attributes to be requested from the DAPS), and security profile (define the security operations that should be performed).

What is the reusability of the credential?
The client assertion can only be used to request a DAT at the DAPS.

A DAT expires in one hour but can be reused within that period of time. The DAT consists of an audience field that can be set to a specific data provider. If the DAT is not expired which it does in one hour, it can be reused by other data providers that are part of the audience.

What process is followed in order for a data consumer to receive access to a resource of a data provider?
For a data consumer to receive access to data of a data provider a two-step process is followed. First the data consumer identifies himself to the DAPS. If the data consumer is who he say he is, the DAPS provides a DAT. With the DAT, the data consumer can request access to the data of the data provider which is the second step in the process.

How is the identity of the data consumer authenticated?
The whole process starts with the data consumer requesting a DAT from the DAPS using the OAuth 2.0 standard. The request consist of multiple headers and parameters. One of the parameters is the client assertion. The client assertion is presented to the DAPS as proof of the client's identity. The client assertion is encoded and signed with the JWS and encrypted using an encryption algorithm. The header of the request includes the encryption algorithm used and the X.509 certificate public key that corresponds to the key that is used to digitally sign the client assertion. When the DAPS receives the request, it decodes the request resulting in the header, payload, and some signature. The DAPS verifies the signature by using the X.509 certificate public key in the header. When the signature is verified, the DAPS authenticated the data consumer. In addition, the DAPS verifies the payload by, for example, checking the time of issuance and expiration time. When all checks have been successful, the DAPS issues a DAT to the data consumer. The DAT is digitally signed with the private key of the DAPS X.509 certificate and is the credential with which the data consumer requests access to the data of the data provider.

When the data provider receives the request, he decodes the request resulting in the header, payload, and some signature. The header of the request includes the encryption algorithm used, the JSON Web Key Set (JWKS) containing the public keys that corresponds to the key that is used to digitally sign the DAT, and the kid field that holds key identifier indicating which key was used to secure the JWS. The data provider verifies the signature by using the public key indicated by the kid field. Because the data provider trusts that the DAPS authenticated the data consumer correctly, he trusts the authenticity of the DAT when the signature is accurate. Then the data consumer is authenticated and access to the data is provided accordingly.

To what degree are the issuance of credentials controlled by multiple parties?
Current implementations have a single DAPS for each data space controlled by one party. When the DAPS is managed by one party, it can decide to stop issuing credentials to a member of the data space as long as it is in line with the token revocation policies set by the data space. This makes the DAPS centralized i.e. a single party controlling the issuance of credentials. IDS connectors are now able to configure which DAPS they trust credentials from and enables inter data space communication. This already takes the control away from the DAPS determining which DAPS it trust and which not. Nevertheless, this does not solve the problem entirely. Currently, there is worked on the concept of federated DAPS where a DAPS can be trusted by multiple data spaces. The goal is to make a data space not reliant on one single DAPS. However, we cannot tell you more regarding these developments.

Who accepts or rejects a party from joining the data space?
Each data space has a data space administrator that accepts or rejects a party from participation based on the policies of the data space. The data space administrator could be a centralized party or a foundation as the case with the Smart Connected Supplier Network (SCSN).

Can a party complying with all policies still be denied from joining the data space?
Yes this is possible. The data administrator controls whether a party may join the data space or not. Although, the data administrator should determine this based on the policies set by the data space, it might deviate from the policies and use subjectivity in the evaluation. This is unlikely in the scenario where the data space administrator is a foundation as the case with SCSN but might be possible in case of a centralized party.

C.3 Gaia-X

What qualifies as a credential within Gaia-X?
While looking at the definition of a credential that you apply, the verifiable credential qualifies as credential.

What process is followed in order for a data consumer to receive access to a resource of a data provider?
The process starts with the creation of a Self-Description which consists of claims that describe an entity. In your case this would be the Douane creating a Self-Description for Dave who is the data consumer. In order to receive a verifiable credential, the Self-Description is provided to an issuer that validates the claims and digitally signs the Self-Description if the claims are valid. Thus a Self-Description digitally signed by an issuer is a verifiable credential. Then, there is an additional step of checking whether the verifiable credential is compliant with the Gaia-X Trust Framework and policies set by the data space. If the verifiable credential is compliant, it is digitally signed by the third party performing the validation.

When a data consumer request access to data, the data provider indicates which verifiable credentials it accepts along with the issuers of those credentials. Assuming the data consumer holds the verifiable credential accepted, he creates a verifiable presentation that he shows to the data provider. A verifiable presentation can hold multiple verifiable credentials allowing a data consumer to selectively choose which parts of a verifiable credential is included. Take as an example a passport in the form of a verifiable credential. A data consumer can decide to only show his name and document number and exclude the rest of the information. A data provider verifies a verifiable credential based on the digital signatures included. The data provider can do so by extracting the DID of the issuer from the verifiable credential and search for it in the Federated Catalogue. The Federated Catalogue returns the associated DID document consisting of information such as (company) name, verification method, and the public key(s). The public key can be used to verify the digital signature with which the issuer signed the verifiable credential shown by the data consumer. This way the data providers is able to check whether the verifiable credential is issued by an issuer indicated before and is able to trust the information inside the verifiable credential. If the verifiable credential is accurate the data consumer and data provider will negotiate about the accessibility of the data consumer for the

data. Ones they are in agreement, the data provider provides access accordingly.

How is the identity of the data consumer authenticated?
This is done when the data provider verifies the digital signature of the issuer. Because the data provider trusts that the issuer validated the Self-Description correctly, it trust the validity of the verifiable credential. However, there is an additional step to authenticate the data consumer. To ensure integrity of the message and prevent the occurrence of a replay attack, the data consumer also digitally signs the verifiable credential with its private key. The same process as for verifying the digital signature of the issuer is followed but now the data consumer's DID is used. When this signature is also verified, the data consumer is authenticated.

To what degree are is the issuance of credentials controlled by multiple parties?
The issuers that are supported depend on the trusted list of the data provider. A data provider can trust as many issuers as he wants allowing the data provider to choose which one it request a verifiable credential from. This way there is not one party in control of the issuance of credentials. However, in practice the data space will likely require a participant to have a compliance and a membership credential. This put significant control in the hands of the issuer of compliance and membership credentials. Whether these issuers will be controlled by one party, multiple parties, or the data space is unknown. I think the most convenient way would be that the issuer will be one party however, while looking at Gaia-X that strives for decentralization they might do otherwise.

Who accepts or rejects a party from joining the data space?
The data space together decides whether a party may join the data space. How this technically works is not known to me but it will most likely be a service that checks whether a party complies with the policies set by the Gaia-X Association and the data space.

Can a party complying with all policies still be denied from joining the data space?
It is unlikely that when a party complies with the policies it will be denied however not impossible.

C.4 European Digital Identity Wallet

What qualifies as a credential within the EUDI Wallet?
The PID, Qualified EAA, and Non-qualified EAA will be issued in the form of verifiable credentials following the SSI concept.

What does the PID credential contains?
PID credentials consist of information similar to PII; information from which the entity subject of the data can be extracted. The PID credentials will be used by organizations as universities and governmental parties.

What does the Qualified EAA credential contains?
Qualified EAA are digitally signed by a trusted party that is legally valid.

What does the Non-qualified EAA credential contains?
Non-qualified EAA are claims that have no automatic legal validity but publishers cannot deny that they issued the claim.

How is authorization between holder and verifier?
Authorization will be done ABAC and will be determined by the verifier. How this exactly is going to look like is yet unknown.

C.5 Sovrin Network

Who issues the credentials in Sovrin? Sovrin itself does not use or issue credentials. Parties that want to issue verifiable credentials can define and store schema and credential definitions on the Sovrin ledger that they can use to create verifiable credentials. What Sovrin does support is the verification of verifiable credentials.

How is the identity of the data consumer authenticated?
This is done by verifying the digital signatures with which the verifiable credential is signed. The issuer of the

verifiable credential digitally signs it as proof that the claims of the verifiable credential can be trusted. The holder of the verifiable credential digitally signs it to ensure the integrity i.e., ensure that the authentication cannot be replayed.

Verifying the digital signature of the issuer is done in two steps. In the first step the verifier extracts the issuer DID from the verifiable credential shown by the holder and search for it on the Sovrin ledger. Resolving the DID results in a DID document consisting of information such as id, verification method, and the public keys. The public key can be used to verify the digital signature. The Decentralized Identifiers (DIDs) and DID documents stored on the Sovrin ledger are anonymized. As a result, Sovrin does not provide the ability for a verifier to verify whether the public key really belongs to an issuer trusted by the verifier. This happens during the second step. The verifier must maintain its own trusted list of issuers consisting of autonym information about the issuers trusted by the verifier including the DID that is stored on the Sovrin ledger. When the issuer DID extracted from the verifiable credential matches with one of the Sovrin DIDs stored on the verifier's trusted list, the verifier knows the issuer of the credential and that it can be trusted. The second step is not supported by Sovrin. The same process is followed for verifying the holder's digital signature but this time the holder's DID is used.

To what extent are the technical specification or guidelines of Sovrin open?
The Sovrin Foundation has published multiple documents in which the architecture is explained. Sovrin is open-source, if you want you can clone the code and set up a network yourself or change the code towards your liking.

How can a data provider prove an unauthorized request was submitted?
Sovrin does not implement authorization.

Who accepts or rejects a party from joining the data space?
This is not part of Sovrin.