Rheinisch-Westfälische Technische Hochschule Aachen
Informatik 5, Information Systems
Prof. Dr. Stefan Decker

Master Thesis

Towards a Data Space for Cyber Threat Intelligence

Navid Rahimidanesh

Mat. Nr. 416790

September 15, 2024

1st Advisor Mehdi Akbari Gurabi, M.Sc.
2nd Advisor Ömer Sen, M.Sc.
1st Supervisor Prof. Dr. Stefan Decker
2nd Supervisor Prof. Dr. Andreas Ulbig

Abstract

Due to the complexity of the ever-changing threat landscape and the interconnected nature of the threats, organizations can benefit from collaborative cyber defense. They can achieve collaboration by sharing cybersecurity information. We found that an effective collaboration platform requires interoperability, flexibility, trust, sovereignty and commercial potentials, which are not fully covered by existing systems. We designed a dataspace solution based on International Data Spaces (IDS) to address these requirements. We conducted a prototype implementation by reviewing existing frameworks. We evaluated the design and prototype by applying real-life scenarios and also requirement analysis. The results showed that our solution can improve trust and data sovereignty by enforcing control over the usage of the shared information, while keeping the added performance overhead insignificant.

Contents

Contents ii

List of Figures v

1 Introduction 1
- 1.1 Motivation . . . 1
- 1.2 Objectives . . . 2
- 1.3 Methodology . . . 2
- 1.4 Outline . . . 2

2 Background 5
- 2.1 Cybersecurity: A Brief Introduction . . . 5
  - 2.1.1 Cybersecurity Framework . . . 6
  - 2.1.2 CSF Framework Core . . . 6
  - 2.1.3 Cybersecurity in the Energy Sector . . . 6
- 2.2 Cyber Threat Intelligence . . . 7
  - 2.2.1 Definitions . . . 7
  - 2.2.2 Role of CTI . . . 8
  - 2.2.3 CTI Sharing Models . . . 9
  - 2.2.4 Automation in CTI Sharing . . . 10
  - 2.2.5 Regulations and Standards . . . 11
  - 2.2.6 Motivations and Concerns of CTI Sharing . . . 15
  - 2.2.7 Related Works . . . 15
  - 2.2.8 Summary . . . 18
- 2.3 Usage Control and Data Sovereignty . . . 18
  - 2.3.1 Access Control . . . 18
  - 2.3.2 Trust Management . . . 19
  - 2.3.3 Digital Rights Management (DRM) . . . 21
  - 2.3.4 Usage Control . . . 22
  - 2.3.5 Data Sovereignty . . . 23
- 2.4 Data Spaces . . . 24
  - 2.4.1 Definition . . . 24
  - 2.4.2 International Dataspaces (IDS) . . . 25

2.4.3 Related Initiatives . . . 26
2.4.4 Application of IDS in Practice . . . 27
2.5 Related Research and Presentation Methods . . . 27
2.5.1 Design Science Research (DSR) . . . 27
2.5.2 4 + 1 Architecture Model . . . 28

3 Use Case and Requirements **31**
3.1 Methodology for Requirement Analysis . . . 31
3.2 High Level Requirements . . . 31
3.2.1 Interoperability and Decentralization . . . 32
3.2.2 Flexibility and Automation . . . 33
3.2.3 Trust and Security . . . 36
3.2.4 Privacy and Sovereignty . . . 38
3.2.5 Commercial Activities . . . 39
3.3 Use Case . . . 39
3.3.1 Collaborative Cybersecurity in a Critical Infrastructure Sector . . . 39
3.3.2 Actors . . . 40
3.3.3 Scenarios . . . 42

4 Conceptual Approach **43**
4.1 Methodology for Design . . . 43
4.2 Functional View . . . 43
4.2.1 Interoperability . . . 44
4.2.2 Flexibility . . . 44
4.2.3 Security and Trust . . . 45
4.2.4 Data Privacy and Sovereignty . . . 46
4.2.5 Commercial Activities . . . 47
4.3 Business View . . . 47
4.3.1 Core Participants . . . 47
4.3.2 Intermediary Participants . . . 48
4.3.3 Supporting Roles . . . 49
4.3.4 Interaction Between Roles . . . 49
4.4 System View . . . 50
4.4.1 Connector . . . 50
4.4.2 IDS Apps . . . 53
4.4.3 Policy Engine . . . 53
4.4.4 Identity Provider . . . 54
4.5 Process View . . . 54
4.5.1 Onboarding and Certification . . . 54
4.5.2 Publishing Data Offers . . . 57
4.5.3 Contract Negotiation . . . 57
4.5.4 Data Exchange . . . 59
4.5.5 Policy Engine . . . 60

5 Realization **63**
5.1 Tool Selection . . . 63
5.1.1 IDS Connector . . . 63
5.1.2 Policy Engine . . . 64
5.1.3 Compatibility Analysis . . . 65
5.1.4 Minimum Viable Dataspace (MVDS) . . . 65
5.2 Deployment . . . 66

6 Evaluation **69**
6.1 Investigation Procedure . . . 69
6.2 Results . . . 69
6.2.1 Application to Example Use Case Scenarios . . . 69
6.2.2 Validation of the Policy Framework . . . 71
6.2.3 Technical Tests and Metrics . . . 75
6.2.4 Analytical Verification of the Architecture . . . 76
6.3 Discussion . . . 78

7 Summary and Outlook **81**

Appendices **83**
Comparison of Usage Control Technologies . . . 83
Integration of EDC and MYDATA . . . 84
Policy Source Codes for Example Scenario . . . 84
Firewall Logs . . . 87

Bibliography **89**

License **97**

List of Figures

2.1 NIST Cybersecurity Framework 7
2.2 The Intelligence Cycle 8
2.3 Information Sharing Models 9
2.4 Threat Intelligence Platform (TIP) [69]. 12
2.5 XACML Data-flow [17]. 20
2.6 XACML Policy Langauge [17]. 20
2.7 ODRL information model [40]. 22
2.8 Technical enforcement vs. Organizational/legal enforcement 23
2.9 Usage Control Scope 24
2.10 Design Science Research Cycle 28
2.11 The "4+1" view model [31]. 29

3.1 Inception Phase 32
3.2 Use Case Actors 40

4.1 Construction Phase 44
4.2 Roles and their interactions; Adopted from IDS RAM [41]. 49
4.3 Different types of Connectors. 50
4.4 Connector Functional View 51
4.5 Provider Connector 52
4.6 Consumer Connector 52
4.7 Policy Engine Components' Interaction Overview 53
4.8 Onboarding Process 55
4.9 Summary of DAPS Interaction 56
4.10 Cataloging Data Model 58
4.11 Broker Interaction 58
4.12 Contract Negotiation Process 59
4.13 Data Exchange Flow 60
4.14 Example of Usage Control Enforcement Process 61

5.1 Deployed Components 66
5.2 Connector Physical Architecture 67

6.1 Validated Real-World Scenario 70

6.2 Latency Measurement 75

Chapter 1

Introduction

1.1 Motivation

By the increasing use of information technology in many sectors, organizations are facing more cyber threats than ever [59]. The loss incurred by cybercrime is huge and is increasing every year, increased from $3 to $6 trillion annually from 2015 to 2021 [1]. Due to the importance of cybersecurity, organizations are spending more to protect their systems. It is estimated that the total spending on cybersecurity will exceed $1.75 trillion from 2021-2025 [19]

To protect against the threats, one should understand them first. To do so, one should collect, process, analyze, disseminate the information about the threats. It includes information about threat actors, their motivations, tactics, techniques, and procedures (TTPs), indicators of compromises (IOCs), the systems' vulnerabilities, incident response plans and mitigation strategies. This process is called Cyber Threat Intelligence (CTI). High quality intelligence requires collection of data from as many of resources as possible, including external sources.

This leads to a collaborative approach for CTI called CTI sharing, or Collaborative CTI (CCTI). CTI can be relevant across organizations due to the similar threat landscape they face. It is due to the common systems, procedures, and adversaries they have. CTI sharing often happens as information sharing communities, where several organizations collaborate as allies by sharing information to get a stronger collective defense.

However, sharing CTI is not a trivial task. Establishing trust, achieving interoperability and automation, dealing with sensitive or classified information, infrastructure for managing external information, validation of quality of information are examples of the challenges of CTI sharing. [26]

A concept that might alleviate these challenges is data spaces. Data space is an emerging data management approach which tackles the burden of large-scale data integration scenarios with an incremental, "pay-as-you-go" fashion. An initiative with the goal of standardizing data spaces is International Data Spaces (IDS). IDS tackles the issues of trust and data sovereignty in the context of business data exchanges. We believe that data sovereignty and trust is of paramount importance in the context of CTI

sharing. Therefore, IDS could provide significant benefits to the CTI sharing use case.

To the best of our knowledge, the concept of Dataspaces is not studied in the context of collaborative CTI.

1.2 Objectives

With the goal of tackling the barriers of cyber threat information sharing, we investigate the suitability and limitations of dataspaces in the context of CTI sharing. Through this research, we aim to fill the following gaps:

- Find specific CTI sharing use cases and scenarios where current platforms cannot adequately address the challenges and find out why.
- Find the degree to which dataspaces can alleviate the challenges in selected CTI sharing scenarios.
- Find the implementation considerations when setting up a dataspace for CTI sharing.

1.3 Methodology

Our methodology is inspired by the design science and consists of four steps:

- Firstly, we perform a requirement analysis based on a literature review to find gaps in the existing systems and potential improvement points based on dataspaces.
- Second, we do data modelling and architecture design to solve the gaps we identified in the previous stage.
- Third, we proceed with a prototype based implementation to get more insight about the design achieved in the previous stage.
- Lastly, we perform evaluation methods on both the design and the associated prototype. Our evaluations cover both validation, whether our solution is applicable to the use cases, and verification, whether the design fulfills its requirements.

1.4 Outline

The rest of this thesis is structured as follows:

- Chapter 2 provides the necessary foundation to understand the context of CTI sharing and the capabilities of International Dataspaces (IDS) and the related works to tackle the challenges of CTI sharing.
- Chapter 3, will describe the chosen use case and the rationale behind the selection. It will discuss gaps in the current CTI sharing schemes.
- Chapter 4 describes our IDS based platform for CTI sharing in a conceptual level.
- Chapter 5 will go through the implementation considerations.

- Chapter 6 contains our evaluation methods and respective results.
- Finally, Chapter 7 will summarize the findings and recommendations for future works.

Chapter 2

Background

In this section, we will review the relevant concepts and technologies that are necessary to understand the rest of the work. First, we will discuss the current state of the CTI sharing, its goals, and it challenges. Next, we will understand the concept of Data Spaces, its context and what it promises. Our goal is to set the stage for our later discussions on how the data spaces could facilitate CTI sharing.

2.1 Cybersecurity: A Brief Introduction

Cybersecurity, the practice of protecting computer systems, networks, and data, encompasses several crucial terms and concepts:

- **Threat**: Any potential danger to information systems, which could lead to unauthorized access, damage, or data loss.
- **Vulnerability**: Weaknesses or flaws in a system that can be exploited by threats to gain unauthorized access or cause harm.
- **Malware**: Malicious software designed to damage, disrupt, or gain unauthorized access to computer systems.
- **Confidentiality**: Ensuring that sensitive information is accessed only by authorized individuals.
- **Integrity**: Ensuring data and systems are not altered by unauthorized individuals or processes.

Relevant actors in the realm of cybersecurity include:

- **Threat Actors**: Entities with malicious intent, including cybercriminals, nation-state actors, and hacktivists.
- **Hackers**: Individuals or groups that exploit vulnerabilities to gain unauthorized access or cause harm to systems.
- **Security Analysts**: Professionals responsible for protecting systems by identifying, analyzing, and mitigating cybersecurity threats.

- **Vendors**: Companies that develop and provide cybersecurity solutions, tools, and services.

2.1.1 Cybersecurity Framework

The Cybersecurity Framework (CSF) from the National Institute of Standards and Technology (NIST) is a comprehensive set of guidelines designed to help organizations manage and reduce cybersecurity risk [39].

2.1.2 CSF Framework Core

It comprises the Framework Core, which organizes cybersecurity activities, outcomes, and informative references common across critical infrastructure sectors. It consists of five key steps that provide a strategic approach to managing and reducing cybersecurity risk (Figure 2.1):

- **Identify**: Develop an understanding of the organizational context, resources, and cybersecurity risks to manage them effectively. This step involves asset management, risk assessment, and governance.
- **Protect**: Implement appropriate safeguards to ensure the delivery of critical infrastructure services. This includes access control, data security, and protective technology.
- **Detect**: Develop and implement activities to identify the occurrence of a cybersecurity event. This involves continuous monitoring, detection processes, and security event analysis.
- **Respond**: Take action regarding a detected cybersecurity incident to contain its impact. This includes response planning, communications, analysis, mitigation, and improvements.
- **Recover**: Maintain plans for resilience and restore any capabilities or services impaired due to a cybersecurity incident. This step encompasses recovery planning, improvements, and communications.

2.1.3 Cybersecurity in the Energy Sector

In order to improve the efficiency of energy distribution grids, efforts are done to make them smart, i.e. smart grids. Smart grids use different information technology (IT) components to collect and process data. However, these components are susceptible to cyber threats. They are an interesting target for attackers, specially advanced state-sponsored attackers, due to the level of damage that is achievable by a successful attack in the energy sector. A threat actor for smart grids can be an advanced persistent threat (APT) supported by an enemy government, or a gang of experienced cyber criminals intended to disrupt the energy supply by attacking different actors in the supply chain. By doing so, they can reach their goal of causing a blackout, exfiltrating sensitive information, or gaining financial benefits (e.g. ransomware).

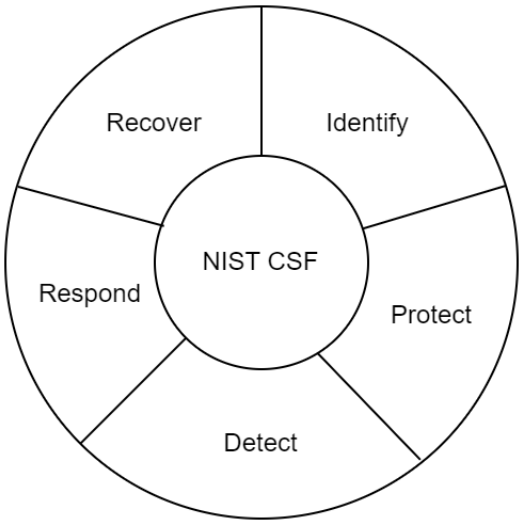

Figure 2.1: Cybersecurity activities as in NIST Cybersecurity Framework [39].

The important threats that smart grids face are listed by Wallis et al. [67]: Data injection attacks on state estimation [14], distributed denial of service (DDoS) and denial of service (DoS) attacks [68], targeted attacks, coordinated attacks, hybrid attacks, and advanced persistent threats [34]. Moreover, in recent years, ransomware campaigns have emerged as a significant risk to the sector [29].

2.2 Cyber Threat Intelligence

2.2.1 Definitions

The term "Intelligence" is often associated with a state defense, an example intelligence organization is the Central Intelligence Agency (CIA). However, it could be relevant in the private sector as well. For example, Competitive Intelligence, which is about gaining marketplace competitiveness through researching the market rivals. It could be defined as: The process and the product of the process of collecting, analyzing and disseminating the information which is helpful in decision-making [33].

To ensure the effectiveness, a systematic approach to intelligence is necessary. Therefore, a life-cycle is often mentioned in the literature to explain all the stages in the process [33]. It consists of 6 stages that follow each other in a circle, where each one builds on the previous one (Figure 2.2).

Similar to the provided definition of intelligence in the context of cybersecurity, we use this definition of Cyber Threat Intelligence (CTI) throughout this thesis: The process and the product of the process of collecting, analyzing, and disseminating information about potential or current cyber threats [33].

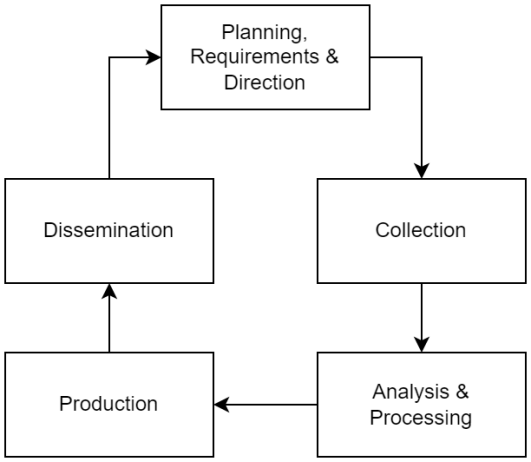

Figure 2.2: The Intelligence Cycle; Adopted from [33].

CTI Sharing is when this process is done collaboratively by multiple organizations. Organizations could improve their CTI capabilities through collaboration, by using collective resources and using findings of others [70].

2.2.2 Role of CTI

CTI is an important part of the security posture of any organization, because a constant study of cyber threats is necessary to mitigate and prevent cyber incidents. It is due to the increase in the attack surface following the digitalization and emergence of new attack vectors developed by the threat actors, hence a constant evolution of the threat landscape.

CTI relies on collecting data from diverse sources, including security tools, threat feeds, honeypots, forums, social media platforms, and other relevant online and offline sources. It could be categorized into three different levels, each concerning different aspects of the threat landscape and different stakeholders:

1. **Strategic CTI – Why?** Strategic CTI expresses high level insights such as overall threat landscape, the motivations of threat actors, and the business or political impact of the threats. It mainly benefits executive management and other decision-making departments by allowing data driven decision-making to reduce the risks of cyberattacks [33].
2. **Tactical CTI – How?** Tactical CTI is about "how" the threats can cause incidents. Examples are the tactics, techniques, and procedures (TTPs) used by threat actors, vulnerabilities in the organization's security infrastructure, and the strategies that were used to mitigate the impact of the breach. Security teams can achieve more efficiency by not repeating the work already done, leading to more agile cyber incident response [33].

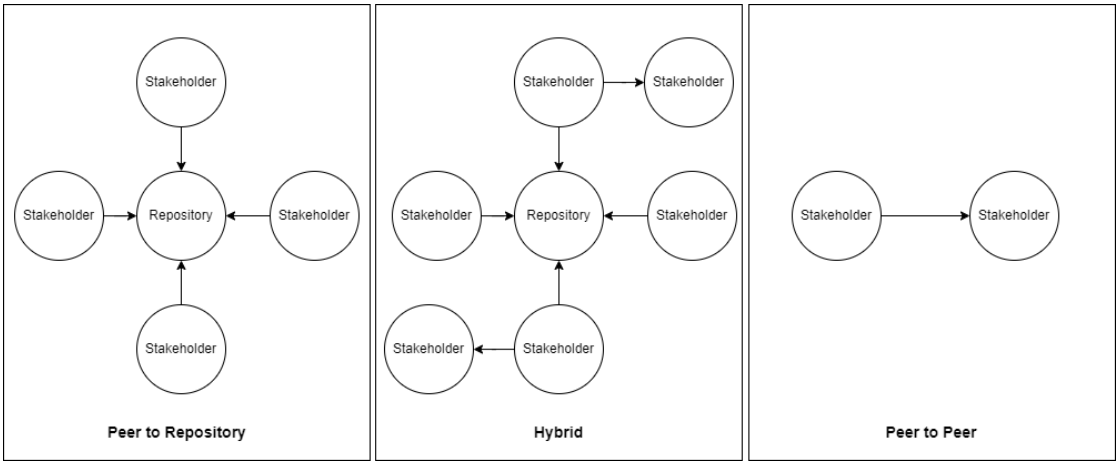

Figure 2.3: Information Sharing Models [66].

3. **Technical CTI – What?** Technical CTI concerns with the indicators of compromise, meaning concrete technological signs about the attacker or an attack, such as malware hashes or malicious IP addresses. Another term for these products is the Indicator of Compromise (IoC), which refer to any technical observable showing an undergoing attack. Security teams and system administrators can feed IoCs to the firewalls and intrusion prevention systems (IPSs) [33].

2.2.3 CTI Sharing Models

In a high level view, there are three types of CTI sharing As (Figure 2.3). First, the peer to peer model, which allows direct exchange between participants, second, peer to repository, also known as Hub and Spoke, which incorporates a mediator that manages all exchanges, and third, a hybrid model that combines both models.

The repository sharing model often happen in the form of sharing communities. The current collaboration communities could be segregated into three different types, namely peer, commercial, and government [6]:

1. **Peer and Sector-Specific Communities:** These are the most common type [6]. These collaborations happen in a community of participants with shared goals, where they know and trust each other by means of meetings and so on. An example type of existing sharing communities are Information Sharing and Analysis Centers (ISACs), which are non-profit organizations that help organizations in a specific sector, usually a critical national infrastructure, e.g. electricity, water, gas, health care, finance, etc., to share CTI with each other.

   An example ISAC would be European Energy Information Sharing and Analysis Center (EE-ISAC) which has acquired over 30 members from utilities, academia, governmental and non-governmental organizations since its foundation in 2015. Members exchange cyber threat information through plenary meetings, working

groups, and a dedicated platform (based on MISP). The information exchange is based on a trust achieved by confidentiality agreements and regular physical meetings with the same members [67].

2. **Internet Communities:** Another type of peer communities are in the form of Open Source Intelligence (OSINT). These communities are accessible via internet and accessible via everyone, public CTI feeds, online forums, or social media accounts (e.g. Twitter) are common channels of OSINT.
3. **Commercial Communities:** These are managed by a vendor, who gathers the participants and collects a fee and provides them with the shared information. It is responsible for the quality of the information and the trust between the participants. It also often anonymizes the information to protect the privacy of the participants. iDefense, Symantec, McAfee, Mandiant, Arbor Networks, and CrowdStrike are some examples of commercial communities. These communities usually use their own proprietary platforms and standards.
4. **Governmental Communities:** These communities are initiated by the government, typically in the form of a public-private partnership (PPP), with either mandatory or voluntary participation from the participants. National or regional Cybersecurity Emergency Response Teams (CERTs) are common governmental organizations that manage and/or participate in this type of community.

2.2.4 Automation in CTI Sharing

Traditional CTI exchange is a manual exchange of data through the following means [66]:

- E-mails
- Phone calls
- Web-community portals
- Shared databases
- Data feeds

This puts the burden of a lot of manual labor not only on the provider side, to prepare and send the information, but also on the consumer side to ingest and analyze the data, verify the quality and relevance of the data, and import it to local system [66]. Therefore, means for automated sharing and consumption of CTI has emerged. It is not expected in the near future to drop the security analyst from the loop completely, but it tries to speed up the whole process [66].

CTI Languages and Protocols

Traditional signature-based methods struggle to reliably detect security breaches because modern, sophisticated attacks are designed to bypass known signatures and exploit multiple vulnerabilities across various systems simultaneously. As a result, organizations must share high-level threat intelligence data to quickly adapt their systems to new threats,

Title	Description
Structured Threat Information eXpression (STIX) [9]	Structured language for CTI sharing (human and machine-readable in JSON)
Trusted Automated eXchange of Indicator Information (TAXII) [9]	Language to share CTI (open transport mechanism with native support for HTTP and HTTPS)
Malware Attribution Enumeration and Characterization (MAEC) [36]	A standardized language for sharing structured information about malware (human and machine-readable in XML)
Incident Object Description Exchange Format (IODEF) [27]	Framework for sharing computer security incidents in XML
Vocabulary for Event Recoding and Incident Sharing (VERIS) [65]	Language to describe structured security events

Table 2.1: CTI Languages and Protocols; Adopted from [66].

utilizing machine-readable formats to eliminate human delay in intelligence sharing [59]. To enable expressing CTI in a machine-readable format, several protocols have been created (Table 2.2.4)

Threat Intelligence Platforms (TIPs)

To process machine-readable CTI formats, Threat Intelligence Platforms (TIPs) has emerged. Although there is no clear definition of the scope of these platforms [54], they typically allow sharing CTI between users as well as collection of CTI from various sources (e.g. OSINT, 3rd party intelligence). However, some only focus on IOCs to be shared, allowing integration to internal security systems such as Firewall, Intrusion Prevention Systems (IPSs), and Security Information and Event Management (SIEM) for an automated response 2.4. A list of available TIPs are in mentioned in table 2.2.4. [54] provides a good survey of these platforms. These platforms are often the backbone of existing sharing communities.

2.2.5 Regulations and Standards

The landscape of Cyber Threat Intelligence (CTI) sharing is heavily influenced by a complex web of regulations and recommendations. Two important regulations are General Data Protection Regulation (GDPR) and The Network and Information Systems (NIS) Directive.

Cybersecurity Regulations Suggesting CTI Sharing

There are some security breach notification laws aiming to increase the overall security posture of its constituency. In Europe, there is NIS directive and its successor, NIS2,

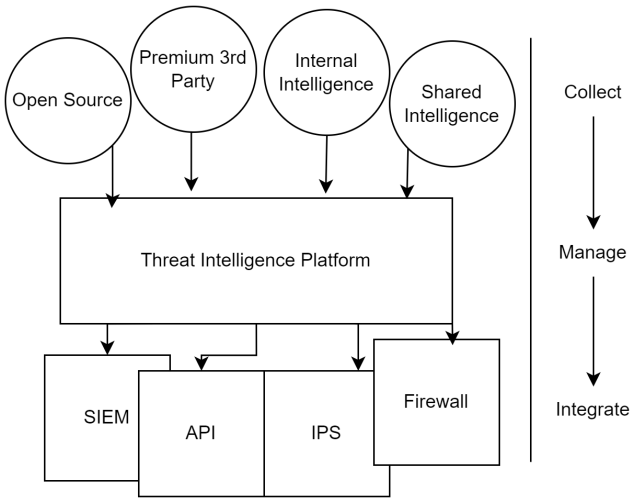

Figure 2.4: Threat Intelligence Platform (TIP) [69].

TIP	**Paid**	**Open Source**
Malware Information Sharing Platform (MISP) [a]		✓
Anomali ThreatStream [b]	✓	
ThreatConnect [c]	✓	
ThreatQ [d]	✓	
EclecticIQ Platform [e]	✓	
OpenCTI [f]	✓	✓
IBM X-Force Exchange [g]		
AlienVault Open Threat Exchange (OTX) [h]		
CrowdStrike [i]	✓	

[a] https://www.misp-project.org
[b] https://www.anomali.com/products/threatstream
[c] https://threatconnect.com
[d] https://www.threatq.com/
[e] https://www.eclecticiq.com/
[f] https://filigran.io/solutions/products/opencti-threat-intelligence/
[g] https://exchange.xforce.ibmcloud.com/
[h] https://otx.alienvault.com/
[i] https://www.crowdstrike.com/

Table 2.2: Threat Intelligence Platforms

Regulation	Jurisdiction
Electronic Communication Act	Slovenia
Cybersecurity Information Sharing Act (CISA)	US
NIS/NIS2	Europe
EO13636	US

Table 2.3: Cybersecurity Regulations Suggesting CTI Sharing; Adopted from [66].

which mandates CTI sharing. Likewise, executive orders like EO13636 promote information sharing in the US. Furthermore, there are country-specific regulations that differ in each jurisdiction. For instance, Slovenia requires communication vulnerabilities to be reported to SI-CERT, while Belgium mandates reporting to its national regulator [66]. The regulations are summarized in table 2.3.

NIS directive, adopted in Europe in 2016, aims at improving the overall level of cybersecurity across the EU [63]. It mandates the EU states to establish a national Computer Security Response Teams (CSIRTs) to monitor, detect, and respond to cyber incidents. These CSIRTs are required to collect information from private sector and share this with other national CSIRTs. It recommends the globally accepted standards and best practices for CTI sharing, such as the STIX/TAXII protocols. On December 2020, the revised directive NIS2 was proposed. NIS2, aims to improve the cooperation between the EU member states and the private sector, by increasing the scope of the directive to include more sectors, and by introducing new requirements for the incident reporting and information sharing.

Privacy Regulations Affecting CTI Sharing

Organizations should be careful when sharing information with other organizations about the privacy of natural persons. There are several privacy laws that should be considered when sharing CTI (Table 2.4).

- In Europe, GDPR is a comprehensive privacy law protecting information about EU citizens, by limiting sharing of Personally Identifiable Information (PII). It aims to protect the personal data and privacy of individuals within the EU and addresses the transfer of personal data outside the EU. GDPR is highly relevant in the context of CTI due to the potential inclusion of personal data in the threat intelligence information, such as IP addresses, email addresses, and other identifying information.
- In the US, acts like Electronic Communications Privacy Act (ECPA) and Foreign Intelligence Surveillance Act (FISA) restrict voluntary disclosure of communications content.

CTI may include information that is permissible to share in one country but prohibited in another. For example, the UK Data Protection Act does not classify IP addresses as personal information, whereas a German court ruled in 2016 that IP addresses can be

Table 2.4: Privacy Regulations affecting CTI Sharing; Adopted from [66].

Regulation	Jurisdiction
ECPA	US
DPA	UK
GDPR	Europe

personal information in certain situations. Therefore, Organizations must ensure they adhere to national privacy laws, especially when sharing CTI with international stakeholders. There are research investigating whether static and dynamic IP addresses are personal details according to the GDPR. The findings revealed, if IP addresses are shared as threat intelligence, then it can be justified in the public interest under Article 6 (1)(e) of the GDPR [66].

Standardization Efforts

Apart from the regulations, there are also recommendations from standardization bodies that try to harmonize the CTI sharing. Three of the most important standardization efforts are from European Union Agency for Cybersecurity (ENISA), International Organization for Standardization (ISO), and National Institute of Standards and Technology (NIST).

The first one is the ENISA's "Proactive Detection Of Network Security Incidents" which aims to enhance the capabilities of CERTs in detecting network security incidents by utilizing both external information sources and internal monitoring tools. This report provides guidelines and best practices to proactively identify and respond to potential cyber threats, thus improving the overall cybersecurity posture within the European Union. The second one is the ISO's "Information technology – Security techniques – information security management for inter-sector and interorganizational communications," which offers a standardized framework for managing and protecting sensitive information across different sectors and organizations. This standard ensures that information security measures are consistently applied, enabling secure and effective communication and collaboration between entities. The third one is the NIST's "Guide to Cyber Threat Information Sharing" which provides guidelines and best practices for sharing cyber threat information among organizations. This guide aims to enhance situational awareness and improve the collective defense against cyber threats by facilitating timely and actionable information exchange. It outlines the processes, policies, and technical aspects of effective cyber threat information sharing to help organizations mitigate risks and respond to incidents more efficiently.

The aspects that are addressed in each one are summarized in the Table 2.5.

Aspect	ENISA	ISO	NIST
Protection of shared information	✓	✓	
Cybersecurity risk management		✓	
Privacy preservation in information sharing	✓		
Data format, protocols and standards			✓
Data quality improvement	✓		
Incident handling process	✓	✓	

Table 2.5: Standardization Efforts Related to CTI Sharing; It describes aspects addressed by each effort. Adopted from [60].

2.2.6 Motivations and Concerns of CTI Sharing

Motivations

There are several motivations for sharing in the literature. [70] categorizes them into four categories: operational, organizational, economic, and policy related benefits.

1. **Operational**: (1) Reduce duplicate information handling, (2) Support breach detection and response.
2. **Organizational**: (1) Improving overall security posture and situational awareness. (2) Combating skills gap.(3) Cross-checking different sources. (4) Expanding professional networks.
3. **Economic**: (1) Total cost savings. (2) Allowing governmental subsidies. (3) Reducing investment uncertainties.
4. **Policy**: (1) Reinforcing the connection with the government agencies.

Barriers

Despite the benefits of CTI sharing, there are several challenges that hinder the effective sharing of cyber threat intelligence. [26] lists the following general challenges: Establishing Trust, Achieving Interoperability and Automation, Safeguarding Sensitive Information, protecting classified information, enabling information consumption and publication. [26] also mentions some challenges that are specific to the consumer side: infrastructure for accessing external information, evaluation of the quality of the information. These challenges are summarized in Table 2.6.

2.2.7 Related Works

There are several works we identified in the literature that aims at addressing the challenges of CTI sharing.

Name	Involved Actor	Category
Achieving Interoperability and Automation	Both	Operational
Evaluating the Quality of Received Information	Consumer	Operational
Safeguarding Sensitive Information (PII, Organization Secrets, Classified Information)	Both	Regulatory & Operational
Establishing Trust	Both	Organizational
Free Riding Effect	Both	Economical
Risk of Reputation Loss	Provider	Economical
Enabling Information Consumption and Publication (Infrastructure for automatic sharing of indicators)	Both	Technical
Accessing External Information (Infrastructure)	Consumer	Technical

Table 2.6: Barriers in Cyber Threat Information Sharing. Adopted from various sources [26], [70].

Leveraging Blockchain Technology

[30] provides a survey of existing literature on the use of blockchain technology in CTI sharing. They found that blockchain technology is promising in addressing the challenges of CTI sharing, such as privacy of participants by anonymous identity and at the same time ensuring quality of data and incentives for the participants to share data due to blockchain based reward systems. However, they also mention that most works are still in the proof of concept phase and the implementation in real world scenarios such as specific industry sectors is not done yet. A notable example is the work of [23] which uses Hyperledger Fabric as a backbone and its channel capabilities to share CTI with specific partners utilizing smart contracts to enforce the traffic light protocol (TLP). In another work, Pahleven et al. [44] extend the technological capacity of TAXII using Distributed Ledger Technologies (DLT) to enable data non-repudiation and a publish-subscribe middleware to enable real-time sharing.

Open Source CTI

Jesus et al. [25] investigated the state of the art of the open source CTI and found the barriers that have prevented the formation of any widely used open source CTI platform. The barriers mentioned are 1) Legal and regulatory (e.g. GDPR or intellectual property) 2) Interoperability (e.g. different formats) 3) Usefulness and return 4) Market factors (e.g. losing reputation, free-riding) 5) Trust in peers and adversarial usage 6) Confidentiality risks. After studying these barriers, as well as some technical gaps, they

present a confidentiality and privacy analysis of sharing a large sample data set of CTI, to make the claim that it is possible to manage risks of sharing using simple techniques like sanitization. Finally, they propose a set of requirements and a reference architecture for an open source threat intelligence platform.

Encryption Based CTI Sharing

De Fuentes et al. [13] presents a scheme called PRACIS to guarantee privacy in cybersecurity information sharing (CIS) networks. They leverage format-preserving and homomorphic encryption primitives to share STIX formatted incident data. They approach allow secure aggregation and forwarding of data in a publish-subscribe architecture. They present a prototype implementation and show that the costs incurred by their approach are easily affordable in real-world scenarios.

Incidents Information Sharing Platform (I2SP)

As part of the Phoneix project [49], Incidents Information Sharing Platform (I2SP), tries to secure European Electrical Power and Energy Systems (EPES) adhering to the NIS directive [58]. It uses the STIX/TAXII protocols to share CTI among the EPES stakeholders, incorporates advanced machine learning and federated learning processes to detect and mitigate coordinated attacks in the EPES. It is inspired by MeliCERTes [38] which is part of the European Strategy for Cyber Security. MeliCERTes is a network for establishing confidence and trust among the national Computer Security Incident Response Teams (CSIRTs) of the Member States and for promoting swift and effective operational cooperation. Member States CSIRTs participate on an equal footing in the MeliCERTes Core Service Platform (CSP) within verified Trust Circles for sharing and collaborating on computer security incidents. [38]

C3ISP

Chadwick et al. [5] provided a cloud-edge based data sharing infrastructure, a trust model and a deployment model to satisfy the requirements of four real-world sharing scenario. Their solution allows the data provider to specify the trust level and the desired sanitization approach. They validated their approach in four pilot projects.

Miscellaneous

- [52] presents an overarching security scheme and architecture to improve the security of service chains with intelligence centered approach towards proactive defense and autonomous response. It was the only work I found that mentions the use of Dataspaces in the context of CTI sharing, although not delving in details.
- [7] presents a framework for securing smart grids based on FIWARE platform by creating a digital twin complying with Common Information Model. It uses a SIEM system for security monitoring.

Aspect	[13]	[23]	[58]	[5]	This Work
Data Sanitization	✓	✓	✓	✓	✓
Sharing Policies			✓	✓	✓
Trust Modelling		✓		✓	✓
Energy Sector Application			✓		✓
Usage Control					✓

Table 2.7: Summary of Related Works and Aspects Addressed.

- Paice and McKeown [45] validate the utilization of MISP in the UK energy sector by testing different sharing models implemented by MISP in a simulated environment.

Summary of Related Works

We can see that the CTI sharing is a hot topic in the cybersecurity community. Especially following the NIS directive and the lack of adequate technical framework to support it, the research in this area has increased. There are several works that try to address the challenges such as privacy and automation, each leveraging a different set of tools. Despite the suitability of Dataspaces for CTI sharing, we weren't able to identify any work that delves into the challenges of Dataspaces for CTI sharing. Furthermore, we couldn't find any work that addresses the requirements of data sovereignty and usage control directly in the context of CTI sharing. A comparison of the aspects and the goals of some related works is summarized in table 2.2.7.

2.2.8 Summary

Cyber Threat Intelligence (CTI) is a young field that is rapidly evolving. In this section, we have discussed this concept and its importance in the context of cybersecurity. We have also discussed the different types of it, the benefits of collaborative CTI sharing, some regulations, and recommendations that influence the CTI sharing landscape, the current state of CTI sharing, the protocols that are used for CTI sharing, the sharing models and platforms that are used for CTI sharing, the challenges that hinder the effective sharing of cyber threat intelligence, and research efforts to overcome these challenges.

2.3 Usage Control and Data Sovereignty

Before describing the concept of Data Spaces, we need to describe its basics, namely the concept of usage control and data sovereignty. To do so, we first describe its related concepts, namely access control, trust management, and digital rights management.

2.3.1 Access Control

Access control focuses on safeguarding computational resources and digital information from unauthorized access. Its main goals are to manage these resources and information

to prevent unauthorized disclosure (confidentiality) and malicious alterations (integrity), while ensuring that authorized entities have access (availability) [32].

There are several models commonly used for access control in practice, including Discretionary Access Control (DAC), Mandatory Access Control (MAC), Role-Based Access Control (RBAC), Attribute-Based Access Control (ABAC). These models are constructed on the basis of defining a series of access control rules, known as authorizations, in the format ⟨subject, object, operation⟩. This format specifies for each entity (subject) the resources (object) it can access and the actions (operation) it is permitted to perform on those resources [32]. RBAC and ABAC are the most frequently used models.

In RBAC, access rights are assigned based on predefined business functions instead of individuals' identities or levels of seniority. The objective is to ensure users have access solely to the data necessary for them to fulfill their job responsibility, and nothing beyond that. On the other hand, in ABAC, access is granted dynamically based on a mix of attributes and contextual factors, such as time and location. ABAC provides the most detailed level of access control and helps minimize the number of role assignments.

DAC, MAC, and RBAC are centered around regulating access to computational resources and digital information within a secure and trusted environment [32], therefore, they are not suitable for open environments like a data space.

XACML

XACML, the eXtensible Access Control Markup Language, is an implementation of ABAC. It defines an architecture, a policy language, and a data flow scheme (Figure 2.5).

- The data-flow is summarized in Figure 2.5. It starts with Policy Administration Point (PAP) where policies are formulated and sent to Policy Decision Point (PDP). The flow continues upon an access request to a resource, which is intercepted by a Policy Enforcement Point (PEP). It sends the request to PDP for a decision. PDP collects attributes from the Subject, i.e., the requester entity, the environment, and the requested resource. Based on the attributes, PDP makes the decision, and given a positive decision, the resource access is granted. It supports obligations, which are the tasks that should be done before the access, e.g., logging.
- The policy language is summarized in Figure 2.6. It is structured into 3 levels of elements: PolicySet, Policy, and Rule. Each rule has a target which specifies a set of requests it applies to. The target is can be boolean function constucted with AllOf and AnyOf components. The condition further narrows the applicability of the rule based on the attributes of the subject, resource, environment.

2.3.2 Trust Management

Trust management has been introduced to cover authentication and authorization for strangers in an open environment such as the Internet [46].

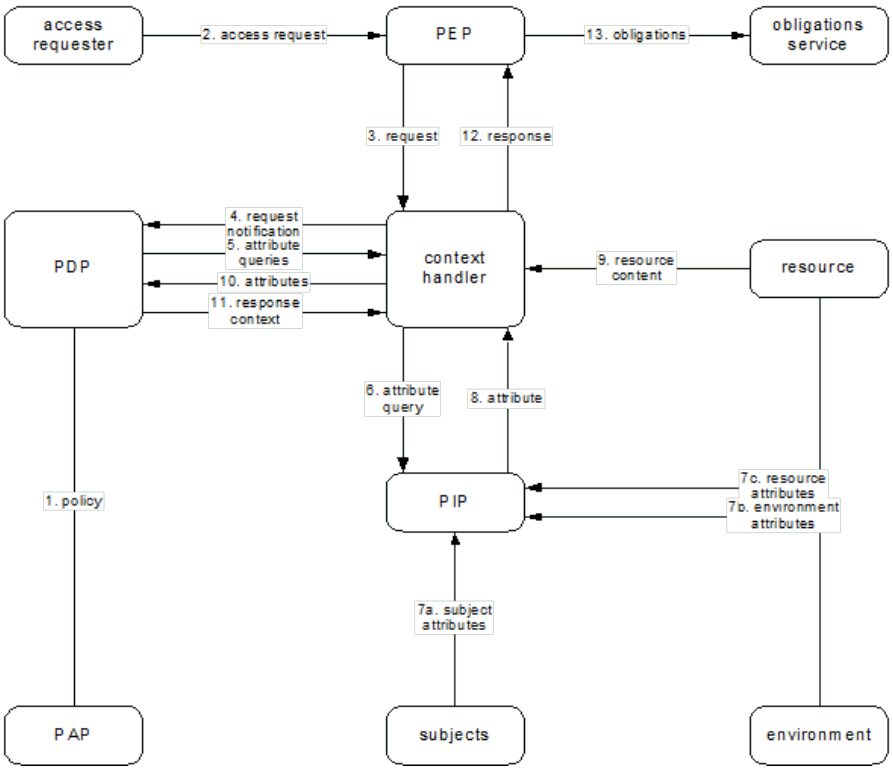

Figure 2.5: XACML Data-flow [17].

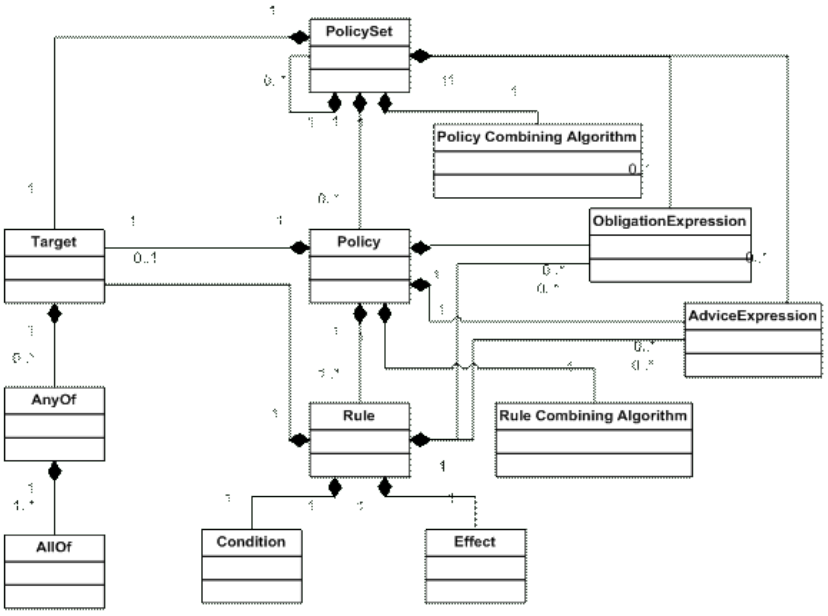

Figure 2.6: XACML Policy Langauge [17].

Public Key Cryptography

Asymmetric encryption, also known as public-key cryptography, is a cryptographic method that uses a pair of keys: a public key and a private key. The public key is distributed openly and can be used by anyone to encrypt data. However, only the corresponding private key, which is kept secret by the owner, can decrypt that data. Common algorithms include RSA, ECC (Elliptic Curve Cryptography), and DSA (Digital Signature Algorithm).

Digital Signature

A digital signature is a mathematical scheme for verifying the authenticity of digital messages or documents. A valid digital signature on a message gives a recipient confidence that the message came from a sender known to the recipient. Digital signatures are typically based on public key cryptography. The sender signs the contents, or a digest of it, using its private key, and the receiver can verify the sender of the message and integrity of its content using the known public-key of the sender.

Authentication using Digital Certificates

Authentication is about verification of the identity of the users. In secure communications using public key cryptography, authentication reduces to binding the public key of the users to some identifying attributes of it. This binding is expressed as digital certificates. A digital certificate provides information about a public-key and is digitally signed by a trusted entity. To manage this certificates, several schemes are possible. Two prominent ones are as follows:

- **Certificate authorities**: This is the scheme used by X.509, and is the backbone of TLS/SSL for authenticating servers on the internet. Here, a hierarchical set of certificate authorities issue and revoke the certificates [3].
- **Web of Trust**: This scheme is used by PGP, which is used to authenticate the email sender. Here, each user accumulates the certificates and associated public-keys it trusts and introduces them to other users. If A trusts B and B trusts C, then A can decide if it trusts C as well, which depends on the degree of trust and the number of other evidences [3].

2.3.3 Digital Rights Management (DRM)

DRM is a broad range of technological methods to ensure that the usage of a digital product is legitimate. It supports the creators and their sales, by making pirate copies substantially harder. There are many types of DRM, such as Media-Based, Product Key, Executable, and Physical DRMs.

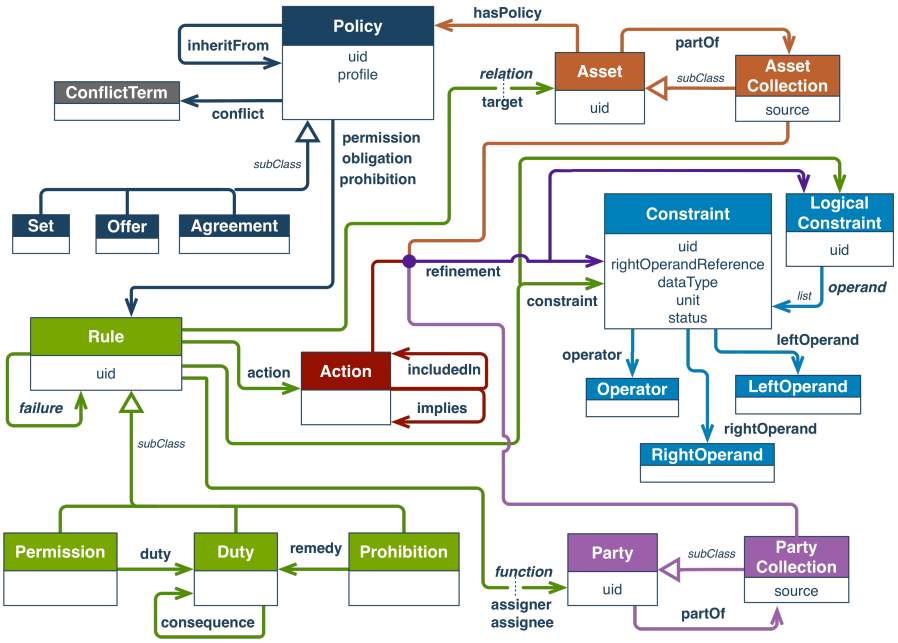

Figure 2.7: ODRL information model [40].

ODRL

To address the needs of DRM-sector, the policy expression language, Open Digital Rights Language (ODRL) has emerged. It provides a flexible and interoperable information model, vocabulary, and encoding mechanism for expressing policies about the usage of digital content. A summary of its information model is depicted in Figure 2.7.

2.3.4 Usage Control

Usage Control, as the name suggests, is about controlling the way the data is used after the access has been granted. It is the generalized version of the traditional access control which only concerns with "who" rather than "how", "where", "why". In usage control, the data owner defines the usage policies and the usage control mechanism enforces them [16]. Some of its use cases are limiting with whom the data could be (re)shared, specifying data retention policies, performing data transformations, e.g., anonymization, aggregation, etc., before distributing the data, and attaching requirements related to privacy regulations, e.g., requiring user consent, or specifying purpose of data usage. Note that Data Usage Control cannot guarantee enforcement in an untrusted environment without legal and regulatory measures. Instead, it builds upon existing trust relationship and reduce the complexity of legal and regulatory measures.

The automatic enforcement of policies requires massive implementation effort due to the plethora of infrastructure systems and different policy classes. Therefore, organiza-

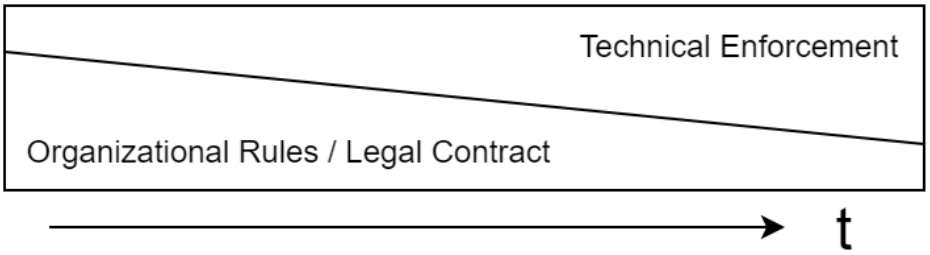

Figure 2.8: Technical enforcement vs. Organizational/legal enforcement
Organizational rules will be replaced by automatic technical enforcement in the long run [41].

tional rules and legal contracts are necessary to ensure correct policy enforcement. These rules will be replaced by technical measures in the long run (Figure 2.8).

Comparison With Related Concepts

Usage Control tries to reach a consolidated view of access control, trust management, and digital rights management (DRM) [46]. Access Control assumes a closed environment where all parties are known during policy specification. Trust management considers unknown parties as well. Both models only control the access rights on the server side, meaning there is no control after data being shared. Digital rights management (DRM) enables some control on the client side (Client-side reference monitor), e.g., prohibit the copy and use only when paid, however, it is focused only on a specific use case, i.e., does not check anything about the client apart from whether paid or not. Usage Control scope include both server-side reference monitoring and client-side reference monitor. Also, it does not focus only on payment based sharing (Figure 2.9).

UCON

Park and Ravi developed the UCON model [46]. In their model, the term "usage" means usage of rights, which include read, save, update, and delegation of these rights to another party. These rights are controlled by a so-called "Reference Monitor" in its "Control Domain". Their model has three main components: Subjects, Objects, and Rights. And additional components are rules, conditions, and obligations.

2.3.5 Data Sovereignty

By the increase of the value of data in businesses and data becoming a commodity, protecting data using laws and regulations has become a necessity. Data sovereignty is a concept that has arisen in this context. It refers to the right of the owner of the data to have control over their data. By default, if a party is processing data owned by another party, the processing party can technically do anything with the data. Data sovereignty

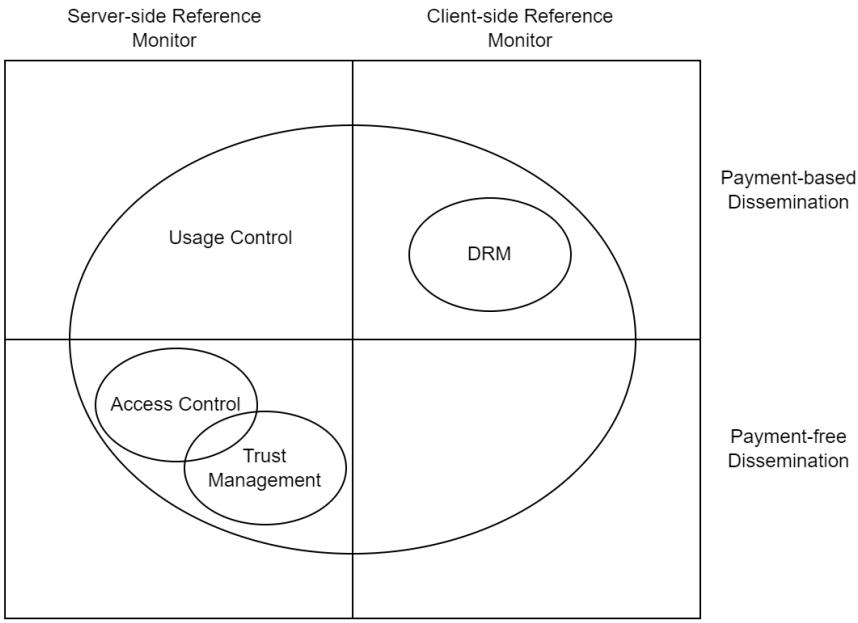

Figure 2.9: Usage Control Scope; Usage Control covers client-side and server-side reference monitoring and payment based and payment-free dissemination [46].

tries to address this issue. Data sovereignty can be achieved by using usage control because it allows introducing and enforcing restrictions on what could (not) happen to the data.

2.4 Data Spaces

In this section, we are going to understand the context of International Data Spaces, its goals, and how IDS promises to provide it. This will lay the foundation for the following chapters, where we apply the concepts of IDS in the context of CTI sharing to find out how it can facilitate CTI sharing.

2.4.1 Definition

The term dataspaces term was first coined by Franklin et al. [18] to describe a new abstraction in data management to solve the data integration problem that follows: An organization has interrelated data in diverse origins, encompassing databases, files with various formats, and web services. The task is to query or update the data. Franklin et al. proposed a Data Space Suppport Platform (DSSP) that helps developers by providing a single query language based on a unified view of the data sources. This implies a pay-as-you-go approach, where physically moving and transforming the data is done only by demand. The same concept was applied for the use case of interorganizational data

exchange and integration in newer contexts. In this context, the term Data Space refers to the platform consisted of data sources in different organizations to do data exchange defined by a set of standards and protocols to enable interoperability. [51]

2.4.2 International Dataspaces (IDS)

The International Dataspaces (IDS) is an initiative to standardize data spaces in the context of European business data exchanges.

Goals

IDS is aiming for the goal of creating a standard for a distributed software architecture for data exchange with sovereignty. These goals are described as follows:

- **Federation:** Dataspaces are usually designed to be decentralized and federated, meaning, there is no entity having direct control over all data exchanges. Different participants can interact with each other directly.
- **Openness:** Dataspaces should be open, meaning anyone complying with the policies should be able to join, which encourages a fair and non-monopolistic market. This entails an easy access, which means, anyone should be able to connect with a limited effort.
- **Data Sovereignty:** Dataspaces can fulfill data sovereignty by keeping the data in the owner's side, and only sharing the metadata publicly.
- **Interoperability:** The need for federation emphasizes the role of interoperability. Interoperability is only possible when certain open standards are established. Consequently, dataspaces complying to the same standards can be embedded inside each other, enabling cross-data-space exchange [51].
- **Governance:** In order to facilitate the cooperation of different participants, a set of policies, rules and protocols should exist. To define them, a governance body is commonly expected to exist [51].

> "Dataspaces are defined as: A federated, open infrastructure for sovereign data sharing, based on common policies, rules, and standards." [51]

History

It was launched in 2015 as a Fraunhofer research project funded by the German Federal Ministry for Education and Research [42]. In 2016, the IDS Association (IDSA) was founded as a non-profit organization to continue the research. It resulted in definition of the IDS Reference Architecture Model (IDS RAM).

IDS Reference Architecture Model (IDS RAM)

The IDS RAM is the description of IDS components and their interactions [42]. It allows anyone to implement the IDS compliant components using any technology IDS

RAM defines the following components [41]: Connector, Identity Provider, IDS Broker, Clearing House, IDS Apps, App Store, Vocabulary Provider [48]. The IDSA provides a reference implementation of these components called IDS Testbed [1]. Furthermore, it conceptualizes the following roles for the participants: Data Owner, Data Provider, Data Consumer, Data User, App Provider. [48]. Also, it defines the standards and procedures to ensure data sovereignty. It uses usage control to enforce the usage policies defined by the data owner. It uses the following components to do so: Usage Control Policy Management Point (UC PMP), Usage Control Policy Decision Point (UC PDP), Usage Control Policy Enforcement Point (UC PEP) [42]. The policies are defined in a machine-readable format, which is an extension of the Open Digital Rights Language (ODRL) [16]. These policies should be mapped to a specific policy language supported by the tool that enforces them.

2.4.3 Related Initiatives

- **Gaia-X** is an initiative, launched in 2019, that aims to foster creation of an infrastructure that allows for free and easy exchange of data and services between organizations and evade the vendor lock-in imposed by current proprietary cloud and service providers. To do so, regulations and technical specifications that are based on European values, applicable to any existing cloud and edge technology stack are going to be defined. The goal is to bring transparency, controllability, portability and interoperability across data and services. By facilitating data collection and sharing between organizations, a vibrant data ecosystem across Europe and beyond could evolve. Gaia-X Association deliverables include federation services, common policy rules and an architecture of standards. Federation services can be utilized by the ecosystem participants to achieve a global interoperability, compliance and effortless set up. This includes, "Identity and Trust", "Federated Catalog" and "Data Exchange services". [64]

  In comparison to IDSA, Gaia-X is less mature and still in the development phase, whereas IDSA is used in the industry[43]. It focuses on cloud infrastructures and businesses operating within EU, in contrast to IDSA which is more on the technicalities of the sovereign data exchange[43]. Finally, Gaia-X can use IDSA in the data exchange layer [43].

- Trusted Integrated Knowledge Dataspace For Sensitive Healthcare Data Sharing (**TIKD**) [20], aims at creating a secure collaborative knowledge graph database of potentially personal data with fine-grained access control and privacy-aware data interlinking.

- **Real-time Linked Dataspace (RLD)** is designed for the Smart Environments, supporting a pay-as-you-go data integration management system for real-time heterogeneous data sources that provides unified query interface based on linked data technologies [8].

[1] https://internationaldataspaces.org/offers/reference-testbed/

2.4.4 Application of IDS in Practice

Dataspaces concept has been applied to many use cases in many sectors [12]. IDSA publishes a list of all applications of Dataspaces in a report and tool called "Dataspaces Radar" [12]. In the report for 2024, it mentions 145 entries in different sectors. It mentions the following sectors: Mobility, Automotive, Energy, Health, Manufacturing and many more [12]. Some notable examples are:

- **Advaneo Data Marketplace (DMP)** [2] is a data marketplace that connects IDS compliant data users to providers to perform sovereign data exchange.
- **Catena-X** [3] is a data exchange framework for the automotive sector based on IDS. The organizations across the automotive supply chain can collaborate on Catena-X to tackle their unique challenges. As of 2024, they have acquired 186 members.
- **Smart Connected Supplier Network (SCSN)** [4] is a Dutch initiative with the goal of interoperability of data exchange between supply chain, which uses IDS as its backbone. They exchange trade information such as receipts and integrate with internal systems such as ERP systems.

2.5 Related Research and Presentation Methods

In order to conduct our research and effectively present it, we took inspiration from pre-existing frameworks in the literature. In this section, we will describe two of them.

2.5.1 Design Science Research (DSR)

Design Science Research (DSR) is a research methodology that aims to create and evaluate artifacts to solve real-world problems in the context of Information Systems research [22]. Hevner defines a three cycle view of DSR, which consists of relevance, rigor, and design, in an inspiring article [22]. A good DSR research should have adequate progress in each cycle. A summary of the DSR methodology is shown in Figure 2.10.

- **Relevance:** The relevance cycle tries to keep the Design Science activities relevant to its application's domain. Therefore, it involves understanding the problem domain, the stakeholders, the limitations of current systems. It should form the basis for elicitation of the requirements of the resulting artifact. After the artifact is created, it should be evaluated in terms of its relevance to the problem domain. This take place multiple rounds until the artifact is deemed relevant enough.
- **Rigor:** The rigor cycle is aimed at ensuring the scientific rigor of the research. This involves grounding the design methods in the existing literature, using appropriate scientific methods. It requires systematically reviewing the existing literature to ensure that the research is innovative. At the end of each cycle, the findings should be shared with the scientific community to get feedback.

[2] https://www.advaneo-datamarketplace.de/en/
[3] https://catena-x.net/de/
[4] https://smart-connected.nl/en

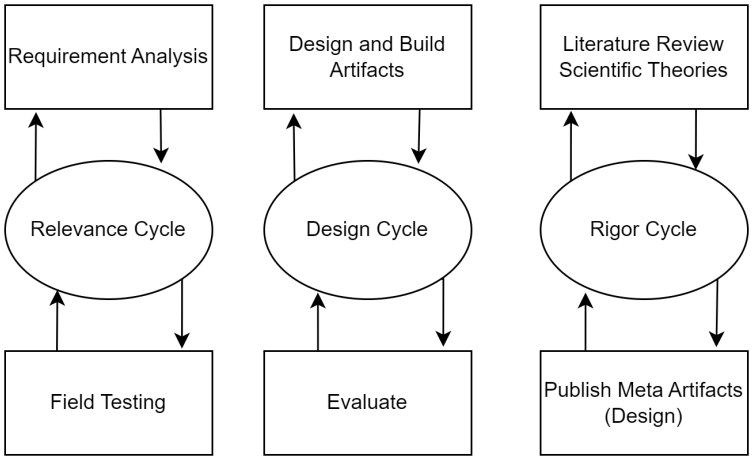

Figure 2.10: Design Science Research Cycle [22]

- **Design:** The design cycle is the core of the DSR methodology. It involves creating the artifact that solves the problem. It should be based on the input from the former two cycles, namely the requirements and the scientific methods. Each produced artifact should be evaluated appropriately to ensure that it meets the requirements.

2.5.2 4 + 1 Architecture Model

The 4 + 1 Architecture Model is used to describe the architecture of software-intensive systems based on 5 different views that are depicted in Figure 2.11. The views are as follows:

- Logical View: Focuses on system functionality from view point of the end-users.
- Process View: Deals with the interactions and concurrent behaviors.
- Physical View: Depicts software deployment on hardware
- Development View: Illustrates system structure from a programmers' point of view.
- Scenarios (+1 View): Use cases and/or scenarios.

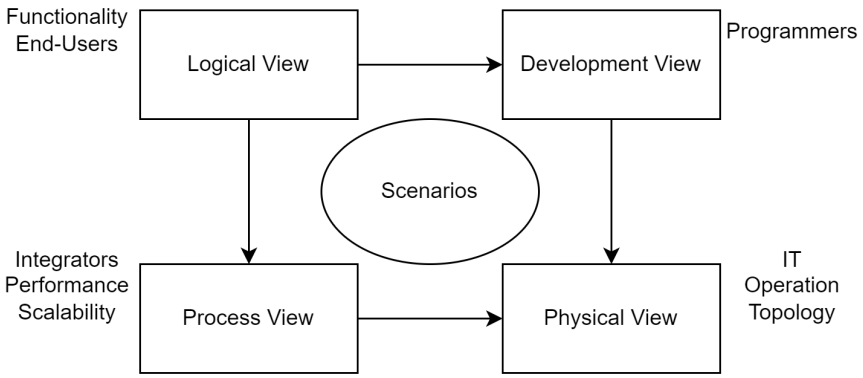

Figure 2.11: The "4+1" view model [31].

Chapter 3

Use Case and Requirements

In this chapter, we aim to find some concrete use cases that can highlight the potential of Data Spaces in the context of CTI sharing. We describe the high-level requirement of the platform, and then we describe some scenarios that will demonstrate the mentioned requirements in some concrete instances.

3.1 Methodology for Requirement Analysis

Our goal is to find some requirement for our CTI sharing platform. We try to find specific gaps in the current CTI sharing platforms that could be addressed by a Dataspace based solution. We call this the inception phase, where the goal is to understand the main challenges in designing of CTI Dataspace and a conceptual approach of the system.

To do so, we delve into the current CTI sharing scenarios by looking at the existing platforms, the scientific literature, and some guidelines and regulations. The next step in this phase is to collect as much as possible information about the Dataspace concept, its promises, and its use cases. The subsequent step is to leverage the information gathered in the previous steps to find improvement areas that could be addressed by a Dataspace based solution. After several iterations, we will have a high level understanding of the design of a Dataspace based CTI sharing platform and a real-world scenario where it could highlight its advantages over traditional CTI sharing platforms. This phase is summarized in Figure 3.1.

3.2 High Level Requirements

This section will elicit some high level requirements of a successful CTI sharing framework. We analyzed the existing literature on CTI sharing and analyzed the barriers mentioned therein, and formulated them in a high level requirement. We will discuss why they are not satisfied by existing platforms and how a framework implementing those requirements will be able to overcome that barriers.

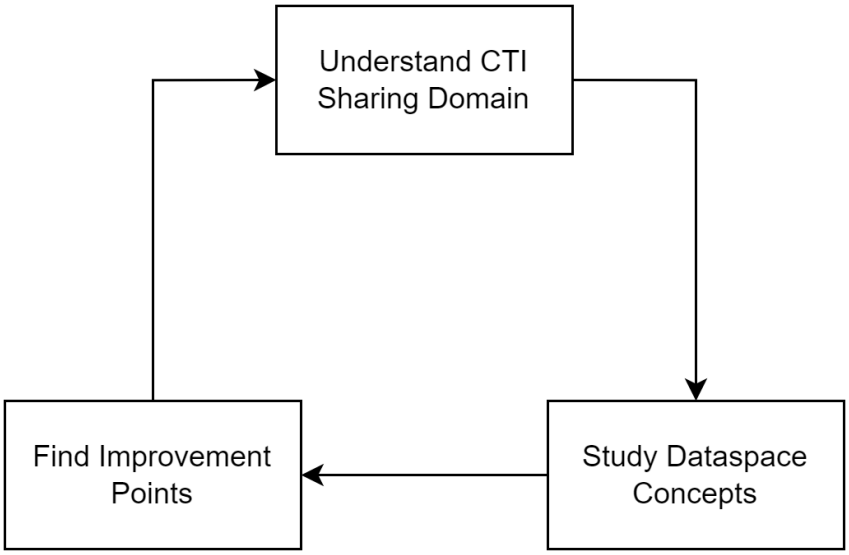

Figure 3.1: Inception Phase

3.2.1 Interoperability and Decentralization; Avoid Vendor Lock-in

We will try to understand the reason a centralized repository is not practical and why current CTI sharing is dispersed, what problems this dispersed CTI sharing pose to the participants, what is the importance of federation in CTI sharing, and what are the challenges in achieving it.

Context: Decentralized Nature of Sharing Communities

In the current state of practice in CTI sharing, as mentioned in Section 2.2.3, the sharing of CTI is done in dispersed communities. Each community has its own set of rules, policies, and standards. Organizations join these communities based on their interests and requirements. National Institute of Standards and Technology (NIST) in its guide for CTI sharing [26] provides a description of different sharing communities and a selection guide for organizations to choose the right community. There is no one-size-fits all solution for CTI sharing. There are communities around some geographic region, political boundary, industrial sector, business interest, or threat space (e.g., focused on phishing attacks) [26]. An important aspect is trust in the community. Organizations are more likely to share CTI with a community that they trust. This trust is built over time by the community's ability to protect the shared CTI and the community's ability to provide valuable CTI. It is often not permitted to share CTI across communities due to the policies or standards incompatibility, which results in a fragmented CTI sharing landscape with siloed CTI repositories.

Problem: Incompatible and Siloed CTI

To reach their intelligence goals, organizations often need to join multiple communities [26]. However, this puts a significant burden on the participants to manage and integrate the CTI from multiple sources due to several reasons. The lack of interoperability and different standards used by different communities, establishment of trust, and validating data quality are obstacles that participants should overcome [70]. Furthermore, lack of visibility across communities will hinder obtaining the full value from CTI sharing. For example, efforts to detect and mitigate a threat may be duplicated across communities [70]. Also, detection of advanced threats might not be possible because there was not enough information available to detect the threat. Another issue is the vendor lock-in problem, usually with the commercial providers, that is caused by the difficulty of moving from one CTI provider to another.

Solution: Standardization

Standardization is a solution to the interoperability requirement. Standardized allows the CTI sharing communities to connect easily and share CTI with each other while maintaining their autonomy. This requires a common set of protocols for data communication, distribution policies, and trust management that all the communities agree upon. There are some existing standardization efforts such as STIX/TAXII (more information in background section 2.2.4) and the sharing standards like ISO/IEC27010, NIST, and ENISA report (ref. 2.2.5). However, they do not cover all requirements that we identified.

3.2.2 Flexibility and Automation; Supporting Exchange of All Types of Security-Related Data

To facilitate broader adoption of a sharing framework by various communities and participants, the framework should be useful in a variety of tasks of cybersecurity and CTI lifecycle.

As current CTI processes employs many tools, systems, and processes to reduce manual work and achieve automation, it is essential for a sharing framework to integrate seamlessly with them. In the following, we will list some standards and tools that are used in practice, which should be supported by any effective CTI framework to enable greater adoption.

Cyber Threat Detection

Cyber operations personnel utilize both automated and manual mechanisms to monitor and evaluate cyber operations for the detection of specific cyber threats. This can be done by examining historical evidence, maintaining current situational awareness, or predicting threats through leading indicators. Detection usually involves identifying patterns of cyber threat indicators. For instance, in the event of a confirmed phishing attack with defined indicators, cyber operations personnel may extract observable patterns from these indicators and apply them within the operational environment to identify any

Table 3.1: Internal tools for monitoring and detection [50]

Client honeypots	Server honeypot	Firewall
Sandboxes	IDS/IPS	Antivirus programs
NetFlow	Darknet	Passive DNS monitoring
Spamtrap	Web Application Firewall	Application logs

Table 3.2: External feeds for monitoring and detection [50]

MalwareURL	Malware Domain List	Google Safe Browsing Alerts
IV	Dshield	AusCERT
EXPOSURE	HoneySpider Network	Cert.br Honeypot Project
AMaDa	Zeus/SpyEye Tracker	Team Cymru – TC Console

signs of the phishing attack [2]. ENISA published a report named "Proactive Detection of Network Security Incidents" in 2011 [50]. Proactive detection of incidents involves identifying malicious activities within a CERT's constituency using internal monitoring tools or external services that report detected incidents before the affected parties are aware. This can be seen as an early warning service, enhancing a CERT's operations, situational awareness, and incident handling efficiency, which is crucial for the core services of national or governmental CERTs [50]. Internal tools, or sensors, such as firewalls, antivirus systems, or honeypots, can be deployed by CERTs to monitor events within their network or across a larger enterprise or national constituency (Table 3.1). External services on the Internet, or feeds, provide information about detected incidents to affected parties and can be public, closed, or commercial, often requiring subscription. These services are maintained by various organizations and individuals, offering data feeds that recipients must parse to extract relevant information (Table 3.2) [50].

Correlation Tools

Correlation has been shown to be valuable for gaining better insights, eliminating false positives, and detecting duplicates [59]. Incident correlation involves comparing different events from multiple sensors and data sources to identify patterns and relationships, helping to pinpoint events related to a single attack or indicating broader malicious activities. This process enhances understanding of events, reduces the workload for handling incidents, and automates the classification and forwarding of incidents relevant to specific constituencies. Correlation is beneficial both for processing data from various tools on a monitored network and for using multiple external services that provide incident data [59]. An example type of correlation tool is SIEM (Security Information and Event Management) tools, which work on the enterprise level. SIEM tools collect, aggregate, and analyze volumes of data from an organization's applications, devices, servers, and users in real-time, so security teams can detect and block attacks. Enisa, in report [15], recommends employment of correlation methods to remove false positives and duplicates

both in the data provider and the consumer side [59]. A summary of some open-source correlation tools are summarized in table 3.3.

Table 3.3: Overview of Open-source Correlation Tools [59]

Tool	Type	Description
SEC[a]	Generic	Processes log files to detect event patterns within predefined time windows.
LogHound[b]	Generic	Employs frequent item set mining algorithm to identify patterns in event logs.
iView[c]	SIEM	Provides centralized reporting from multiple devices, enhancing visibility across network activities.
OSSIM[d]	SIEM	Integrates log management and asset discovery with security information from various controls for enhanced correlation.
Abuse Helper[e]	Incident handling	Aggregates and correlates Internet abuse information from various sources.
BGPrank[f]	Incident handling	Ranks autonomous systems based on the level of malicious activity detected.
CIF[g]	Incident handling	Compiles threat data from various sources for response, detection, and mitigation of threats.

[a] https://simple-evcorr.github.io/
[b] https://ristov.github.io/
[c] https://www.cyberoam.com/cyberoamiview.html
[d] https://cybersecurity.att.com/products/ossim
[e] https://www.cert.fi/en, https://www.cert.ee/en
[f] https://www.circl.lu/services/bgpranking/
[g] https://www.ren-isac.net

Standard Intelligence Representation

A crucial prerequisite for a mature and useful cyber threat intelligence and cyber threat information sharing is the availability of an open-standardized and structured ways to represent cyber threat information, and as such, standards like STIX has emerged [2]. STIX guiding principles are maximizing expressivity, flexibility, extensibility, automatability, and readability. There are several other standards that are used, [21] categorizes them into processes and maps them to various knowledge areas. [21] classifies them into six key knowledge areas, each associated with a specific process (indicated in parentheses): Asset definition (inventory), Configuration guidance (analysis), Vulnerability alerts (analysis), Threat alerts (analysis), risk/attack Indicators (intrusion detection), and Incident Report (management). Table 3.4 shows the correlation between the standards and these knowledge areas.

	CPE	Oval	SWID	XCCDF	CCE	OCIL	CCSS	CVE	CWE	CVSS	CAPEC	CVRF	MAEC
Asset Definition	•	•	•										
Configuration		•		•	•	•	•						
Vulnerabilities		•						•	•	•		•	
Threats								•	•	•	•		•
Incidents	•						•					•	•
Risk	•	•			•			•	•	•			•

Table 3.4: cybersecurity standards and knowledge areas [21]

Threat Intelligence Platforms

To analyze and share high-level threat intelligence data, some tools have emerged. OpenIOC is an open framework for sharing threat intelligence using an XML schema to describe technical characteristics of threats, primarily focusing on file-based Indicators of Compromise (IoCs), and is widely used within the Mandiant product community. On the other hand, the Malware Information Sharing Platform (MISP) is an open-source tool developed by Belgian Defense CERT and NATO NCIRC, providing a central database for IoCs and enabling automated sharing and integration with other systems, including support for generating various outputs like IDS and XML [59]. For a more comprehensive list, refer to table 2.2.4.

Course of Action

After analysis of the threat data, the analyst might define suggested Course of Action (CoA), or Playbooks. This could be preventive, to prevent the occurrence of the cyber incident, or responsive, to mitigate and alleviate the ongoing or past cyber incident. In order to automate the tasks of the response, some Security Orchestration, Automation, and Response (SOAR) products has emerged. Furthermore, some machine-readable formats for expressing this courses of action has been emerged. Schlette et al. [55] provide a comparative analysis of these formats. Some of these formats and products are summarized in table 3.5.

3.2.3 Security and Trust; Managing Identities

Establishing a CTI sharing collaboration requires building a comprehensive trust relationship among stakeholders. Stakeholders must establish their reputation to become trusted members of a threat-sharing community. The quality and quantity of the data shared in a sharing community is correlated to the amount of trust between participants. Trust is necessary for three aspects of CTI sharing relationships [66]:

Category	Format/Name	Inception	Maintainer / Vendor
Format	CACAO	2017	OASIS
Format	COPS	2016	DEMISTO
Format	IACD	2014	DHS / NSA / JHU
Format	OPENC2	2015	OASIS
Format	RE&CT	2019	ATC Project
Format	RECAST	2018	MITRE
SOAR	TheHive & Cortex	2014	TheHive Project
SOAR	Cortex XSOAR	2015	Palo Alto Networks
SOAR	Splunk Phantom	2014	Splunk
SOAR	ThreatConnect	2011	ThreatConnect

Table 3.5: Incident Response Formats and Products [55]

- Protecting Confidential Data: A trusted data consumer will protect the confidential and sensitive information such as Personally Identifiable Information (PII) that should only be shared with trusted stakeholders to prevent misuse and protect reputation.

- Correct Information Handling: A trusted data consumer will use the data for legitimate purposes and with proper means.

- Credibility of shared information: A trust data provider will share accurate and reliable information.

Trust Monitoring

It is often personal and achieved through time and face-to-face interactions, meaning if a key employee leaves, the trust network may be disrupted. However, trust establishment via meetings is particularly challenging among decentralized stakeholders. Trust issues can arise from stakeholders initially behaving benignly and later abusing trust. Therefore, sharing systems must implement a continuous vetting process to identify malicious peers early. It can be managed by third parties or trust managers. It should consider factors like reputation, past outcomes, and stability. Reputation is developed over time by consistently sharing high-quality, actionable threat information and adhering to threat-sharing policies. To enhance credibility, stakeholders should continuously share CTI, correlate information from various sources, and respond to community inquiries about the shared intelligence. Conversely, once a bad reputation is established, it is difficult to reverse [66]. Trustworthiness could be evaluated through direct contact and opinions of other peers.

Table 3.6: CTI Sharing Standard Policies [66]

Policy	Suggested Clauses	References
Information Exchange Policies (IEP)	purpose, scope, participants, sharing procedures, data handling, policy modification procedures, uses of data	[11], [57]
Data Sharing Agreement's (DSA)	data quality, obligations, trust domain, security infrastructure	[37]
ISO/IEC 27010:2015	–	[24]

3.2.4 Data Privacy and Sovereignty

Sharing Agreements

If an organization decides to share their CTI, a clausal for information has to be included or updated in existing policies. [11] introduces Information Exchange Policies (IEP) to which all information exchange with other stakeholders has to go through. [57] identified the following elements that must be included in the IEP: scope, purpose, sharing procedure, data handling, policy modification procedures, accepted uses of data, and intellectual property rights.

[37] introduced Data Sharing Agreement's (DSA) for interorganizational data sharing, which covers the following aspects: data quality, obligations, trust domain, security infrastructure.

The ISO/IEC standardization body proposed ISO 27010:2015 Information technology – Security techniques – Information security management for inter-sector and interorganizational communications, which provides recommendation for parties sharing information with one another [24].

Ethics in data sharing has to be part of the information sharing policy. Stakeholders should define for which purpose the CTI is used, who can access it, retention periods and destruction, and condition of publication [66].

A summary of the policies is presented in table 3.6.

Privacy Enhancing Technologies (PETs)

When sharing with semi-trusted participants, organizations use Privacy Enhancing Technologies (PETs) to protect their reputation and the privacy of their clients. Techniques like k-Anonymity, l-Diversity, t-Closeness, ϵ-Differential privacy, and Pseudonymization are used for anonymization. Encrypting CTI is crucial to prevent interception, with protocols like PRACIS designed for secure data forwarding. Anonymity is essential when stakeholders prefer not to disclose a breach due to reputation damage, but still want to share intelligence. This involves removing personally identifiable information (PII)

from content, metadata, and data transfers, potentially automated through regular expressions. Different stakeholders have varying perceptions of anonymity, necessitating adjustable masking criteria. Routing connections through networks like TOR can further enhance anonymity. Encrypting CTI ensures that sensitive information cannot be exploited before vulnerabilities are addressed, maintaining implicit privacy [66].

3.2.5 Commercial Activities; CTI Marketplace

Finding incentives for active participation in CTI sharing activities is challenging for some organizations [70]. They anticipate gaining insights from other community members without providing any valuable input. This results in a "Free riding" effect [66]. There are punishment models proposed for the free riders in a CTI exchange [66], however, the inherent heterogeneity of organizations means their ability to generate and share intelligence varies, making equal knowledge exchange unrealistic. Engaging organizations in threat-sharing collaborations can be challenging and resource draining. Further, It is believed that organizations are more likely to contribute if they expect reciprocal benefits [66]. Therefore, economic aspects are crucial in CTI sharing. Zibak [70] mentions lack of incentives and return on investments as a barrier in CTI sharing. There are obligations to share CTI such as NIS2, however, monetary incentivization is also necessary. To address this, the US Congress proposed the Cyber Information Sharing Tax Credit Act, offering a tax credit as a financial incentive for organizations that share CTI with other stakeholders [66]. Danurand et al. tried to address CTI sharing barriers in [11] by designing an infrastructure for sharing activities. He mentions "Enable Commercial Activities" as one of the high level requirements of a successful sharing infrastructure. He goes on to add, the private sector will be more inclined to engagement if accounting models and functionalities for selling data or data-related services. This will enhance the quality of data available. Therefore, diverse accounting models for data usage and mechanisms for controlling the dissemination of exchanged data under commercial contract terms are helpful. Organizations must have the capabilities to sell any data element, and professional services based on CTI data such as quality control, correlation, and translation. By supporting commercial activities, the industry's vast resources and expertise to procure the CTI data could be leveraged, thus, increasing the quality of data and keeping the costs low due to competition [11].

3.3 Use Case

In this section, we will describe our chosen use case that shows the requirements we identified in the last section. We will use this scenario in the following chapters to validate and justify the proposed solution.

3.3.1 Collaborative Cybersecurity in a Critical Infrastructure Sector

Our use case revolves around supply chain organizations in a critical infrastructure industry sector, e.g., Energy, that are subject to the NIS2 directive and collaboratively

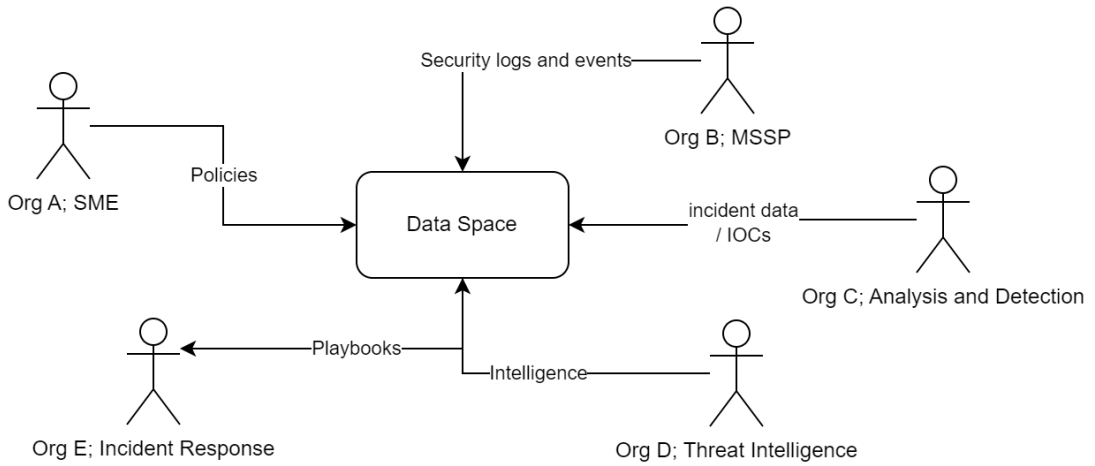

Figure 3.2: Use Case Actors; A summary of actors and the data types they exchange in our platform.

improve their security posture.

Due to the criticality of the cybersecurity in the energy sector, organizations in the energy supply chain try to improve their security posture. By using similar technologies, these organizations share the same vulnerabilities. Therefore, they attract the same attackers. These attackers use a specific set of tactics, techniques, and procedures (TTP) to attack their victims over and over again. As a result, the victims can prepare themselves for these threats by knowing the TTPs that were used against other victims.

Therefore, energy sector organizations can benefit from sharing threat information with each other. That is where they can use our proposed platform. Therefore, the participants using our platform are the organizations active in the energy supply chain. Here, we assume that participants are the security team of the aforementioned organizations, or a managed security service if the organization does not have its own security team. That is because the security team is the entity responsible for handling the CTI.

3.3.2 Actors

To walk through the scenario, we first describe different actors using the platform. A summary of actors and the data they provider/consumer is depicted in Figure 3.2.

Org A; An SME in Energy Sector

Org A is a small and Midsize Enterprise (SME) organization which operates in a Critical Infrastructure (CI) sector such as Energy. It is subject to the NIS2 directive and therefore is obliged to notify the national CERT of which it is a constituent about any breaches occurring to it. Therefore, it requires adequate monitoring infrastructure to

detect breaches quickly and have enough evidence to send to the CERT. It does not have technical capabilities for this on its own and thus outsources this task.

Org B; Managed Security Service Provider (MSSP)

Org B is a Managed Security Service Provider (MSSP), that deploys and manages services such as firewall, antivirus, and intrusion detection and prevention system (IDPS). These systems generate information and logs about their running environment, which is useful in detection of cyber incidents. To get insights from these data, Org B shares this data with a trusted community to leverage their analysis capabilities. However, Org A, as the owner of the data, wants to control the sharing and the usage of the data.

SOC A; Incident Detection as a Service

SOC A is a cybersecurity vendor specialized in real-time monitoring and management of logs and events from diverse sources. They employ a security information and event management (SIEM) system and human security analysts active 24/7. They filter logs and security events and detect potential security incidents. It is analogous to SOC Tier 1. They want to sell the detected incidents and IOCs with their trusted customers for profit.

SOC B; Threat Intelligence as a Service

SOC B focuses on threat intelligence. It proactively collects information about threat actors and their TTPs. It analyses the escalated incidents by correlating it with the potential actors and vulnerabilities. It provides strategic intelligence reports and Course of Actions (COA) to the interested parties. It is a member of some information sharing communities, and shares the information with those communities. It uses Traffic Light Protocol (TLP) to restrict circulation of sensitive information.

SOC C; Incident Response Team

SOC C does the task of Incident Response as part of a larger organization. It collects the playbooks that are shared within the community and performs them after incidents or before incidents as preventive measure. It might use a SOAR system internally to automate its tasks.

National CERT; Governance CERT

It is a national CERT as part of a governmental agency with the aim of improving the security posture of its constituency. Upon important incidents, it requires its constituencies to notify it.

3.3.3 Scenarios

Sharing Incident Data to National CERT

In this scenario, Org A can adhere to NIS2 directive by sharing incident related data after an incident happening to it. It will get notified by SOC A (Analysis and Detection) of an incident happened to it. It will share the information collected by Org B (MSSP) regarding its assets with the National CERT.

Collaboration Between Security Vendors

In this scenario, different vendors can collaboratively generate high level cyber threat intelligence and security services based on low level security events and sell it to their customers. SOC A will collect events from Org B (MSSP) and produce incident data. SOC B will further analyze this data to reach intelligence. SOC C will provide incident response capabilities. All participants want to control the usage of the data they provide and earn money. The actors are in subject of rules and regulations in different jurisdictions, and they have different trust level in each other.

Chapter 4

Conceptual Approach

The goal of this chapter is to derive some precise requirements for the design of a system that can facilitate the sharing of Cyber Threat Intelligence (CTI) in a decentralized manner and implement the requirements and scenarios we identified in the last chapter.

4.1 Methodology for Design

Taking inspiration from DSR methodology, we followed an iterative design methodology. We call it the construction phase, where we will build an artifact based on the high-level requirements and use case produced in the inception phase. To do so, we will follow a three stage approach (Figure 4.1):

- Design: We read IDS specifications, white papers, deliverables, and related literature and apply them to the requirements we identified.
- Implement: We review existing open-source implementations of the components of IDS, compare them, and implement our required features based on them.
- Evaluate: We measure whether the design and prototype implementation fulfilled our requirements, and we repeat the process.

4.2 Functional View

In the last section, we identified some high level requirements for a CTI sharing framework. In this section, we will decompose those requirements into some functional components. These components are either the ones defined in IDS terminology, or we describe how it will be mapped into IDS components. We will discuss why each component is necessary and how it will address the use cases we have identified. We will organize them according to the high-level requirements we identified in the last section, however, some components are useful in multiple requirements.

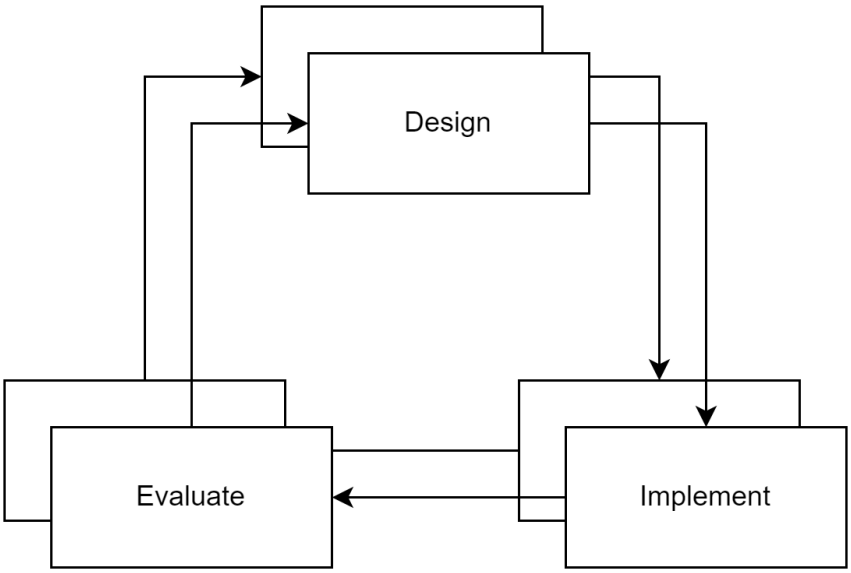

Figure 4.1: Construction Phase

4.2.1 Interoperability

In the following, we list the functional requirements that are related to interoperability.

- **Standard for Data Exchange:** In order to enable interoperability between different participants in a decentralized data space, a standard for data exchange is necessary. To implement that standard, an IT component that performs the actual exchange of bits, called the Connector, is necessary. In a data space, different participants could operate their own connector to perform data exchange without relying on a central entity.
- **Metadata Brokering:** To facilitate the exposure of the data and services offered by participants to a wider range of potential consumers, a brokering mechanism could be helpful. It should provide repositories for the metadata about the data and services offered by the providers, which allows the consumers to search for and discover them. It acts as a phone book by aggregating metadata to allow exposure of data and services within the data space.

4.2.2 Flexibility

We believe the following functional components will increase the flexibility of a CTI sharing platform.

- **Vocabularies:** A shared vocabulary hub is useful to enable a common understanding of the meaning of the provided data and to ease the collaboration between

participants. The vocabulary hub should allow for the development of new data models for the data that is exchanged in the data space.

- **Data Adapters:** Due to the importance of timeliness in the context o CTI, automation is necessary. One aspect to automate is the insertion of data to the data space on the data provider side. Therefore, the connector should be able to retrieve data from enterprise data sources. In the context of CTI, logs from tools like firewalls, intrusion prevention systems (IPS), and in the use case of energy sector, industrial control systems (ICS) such as SCADA (supervisory control and data acquisition) could provide useful data to share within the data space.

  Likewise, automation is crucial in the consumer side. The consumer might like to be notified automatically for each new update. They might want to automatically feed the received data into their internal systems, such as automatically updating firewall rules, or into their SIEM systems. However, integration to enterprise systems in the consumer side is more challenging because automatic enforcement of usage policies might require proper technical infrastructure to be compatible with the individual systems used.

- **Extensibility Through Data Apps:** The connectors could be extended by installing some additional software component, called Data Apps, which allow additional data processing capabilities. This will allow for flexibility and reduce the need to extract the data from the Connector.

4.2.3 Security and Trust

Based on the requirements identified in the last section, we think the following functions are crucial:

- **Digital Identities:** Each participant and their connectors should have a unique identifier assigned to them. A proper identity infrastructure is the foundation of any trust and security mechanism. Prior to each exchange, the involved parties might need to check the identity and attributes of the other party. Participants should be able to define the minimum trust level of the other participant in the exchange. When revealing the identity is not desired, e.g., anonymous transaction, it might be necessary to prove some attributes, e.g., verifiable credentials to the other parties. Therefore, an infrastructure for expressing, generating, revoking identities and the attributes possessed by them are necessary. These identities and attributes should be available for both participants and the IT components used by them.

  These identities are in the form of certificates that describe the subject's identity and attributes. It will be granted only after a so-called certification process which evaluates the operation environment and the characteristics of the IT component.

- **Dynamic Trust and Reputation Modelling and Monitoring:** A one-time certification of the participants is not sufficient, as their behavior or operational environment might change during time. To ensure the trustworthiness of the data

space participants, a continuous monitoring of all participants of the ecosystem is necessary. This should allow for certificate revocation and changes in the trust levels.

4.2.4 Data Privacy and Sovereignty

The following functional component will increase data privacy and sovereignty in a data space.

- **Technical Usage Contracts:** Organizations often rely on policies to control their handling of data. Appropriate specification and enforcement of data handling policies is required for success of any organization. Controlling sharing of information such as CTI data by policies is more important due to the sensitivity of CTI data. Therefore, organizations need to ensure that the data they share with other organizations is also handled properly by them. Traditionally, these contracts are written in natural language. An example of a contract with the end user is the "Review and Accept the Terms and Conditions" button often should press before using services on the internet. However, there are some issues with traditional contracts [4]: First, from a usability point of view, it is difficult to understand the language for expressing the terms and conditions and specifying the preferences. Second, the violation from the contract terms might not be easily detectable, which results in a lack of trust. Third, the content of the contracts might evolve over time, so a life cycle management is necessary and thus extra effort. Automation in definition and enforcement of the contracts could alleviate these issues, therefore machine-readable contract specification and tools to automatically enforce them has emerged [4]. There should be ways to automate or semi-automate negotiation of these contracts. The policy specification language should support three different type of clauses: authorization, prohibition and obligation [46], where each clause could be limited to some conditions.
  - **Conditions:** The policies should allow specifying the conditions that should hold before any data usage instance, such as allowed usage time, purpose, location, and the number of accesses.
  - **Obligations:** It should also support specifying some obligations for the usage of data. For example, notifying the clearing house for each usage instance for the case of metered usage. Another example is to perform the deletion of the data after a retention period.
- **Sanitization Component:** To enable easier sharing of data, the existence of tools to automatically sanitize the data before sharing is beneficial. They could help the provider comply to privacy regulations by removing the PII data or preventing sharing of information marked as classified. It should also be possible to further customize the sanitization process to support the organization's individual data handling policies. In IDS, this is possible via IDS Apps.

4.2.5 Commercial Activities

These functional components foster commercial activities in a data space.

- **Billing and Clearing:** In order to foster a commercial ecosystem, adequate mechanisms for billing and clearing are required. The data owner should be able to define the price for their data or service and billing model.
- **Data Processing Services:** It should be possible within the data space to offer services to process data owned by another party. In this case, the service provider is both a data consumer and a data provider. An example in the context of CTI is an entity providing correlation as a service by collecting events and alerts from their customers and providing the analyzed and deduplicated data.
- **App Store:** To incentivize the IDS software developers to create IDS Apps, an App Store that publishes the apps and provide monetary incentives to the developers is useful. App store will distribute the data apps to be downloaded and installed by the data space connector of the app user.

4.3 Business View

In this section, we overview the required business roles in the system and their interactions. We categorize them into three types, namely, Core, Intermediary, and Supporting.

4.3.1 Core Participants

Core participants perform the transactions, in other words, each transaction in IDS is between two core participants.

Data Provider

Data Provider provides the data asset. It has access to this data and shares this access with the data consumer. It might be possible that the data is not owned by the data provider. In such case, it processes the data on behalf of the owner. The data owner has the right to specify the usage policies, and the provider should ensure those policies are followed during the transaction.

Data Consumer

Data Consumer will receive the data from the provider. It should adhere to the usage policies. The Data Consumer and the Data User might be different, in such case, the Data User is the legal entity that is allowed to use the data, and the Data Consumer is the entity that is authorized by the Data User to use the data.

Service Provider

Service Provider is an actor in IDS, which is used in the scenario when an intermediary processes the data. This actor, will collect data from some data provider and create a

new data asset (data service) by processing that data. This data is then consumer by some data consumer. Service provider is therefore both data consumer and data provider.

4.3.2 Intermediary Participants

These actors facilitate the transactions between the core participants.

Metadata Broker

Metadata Broker will manage and provide metadata brokering as a service.

Vocabulary Intermediary

Vocabulary Intermediary is the participant that is managing the Vocabulary Hub. It will technically manage and offer vocabularies (i.e., ontologies, reference data models, or metadata elements).

App Store Provider

App Store business role will manage an App Store and verify and certify the Apps.

Clearing House

To facilitate the billing process, a trusted third party for exchanges is necessary. Clearing House provides clearing and settlement services for all financial transactions in the data space. It will monitor activities during a data exchange by collecting information from either data provider or consumer and use this information for conflict resolution between participants.

Identity Authority

Identity Authority manages the identity information of the participants. It consists of four subcomponents:

- Certification Authority: issuing and managing digitally verifiable certificates for the participants.
- Dynamic Attribute Provisioning Service (DAPS): providing dynamic attributes of the participants in the form of verifiable tokens.
- Dynamic Trust Monitoring (DTM): for continuous monitoring of the security and behavior of the network
- Participant Information Service (ParIS): Provides business-relevant attributes such as registered address and tax identity number of participants.

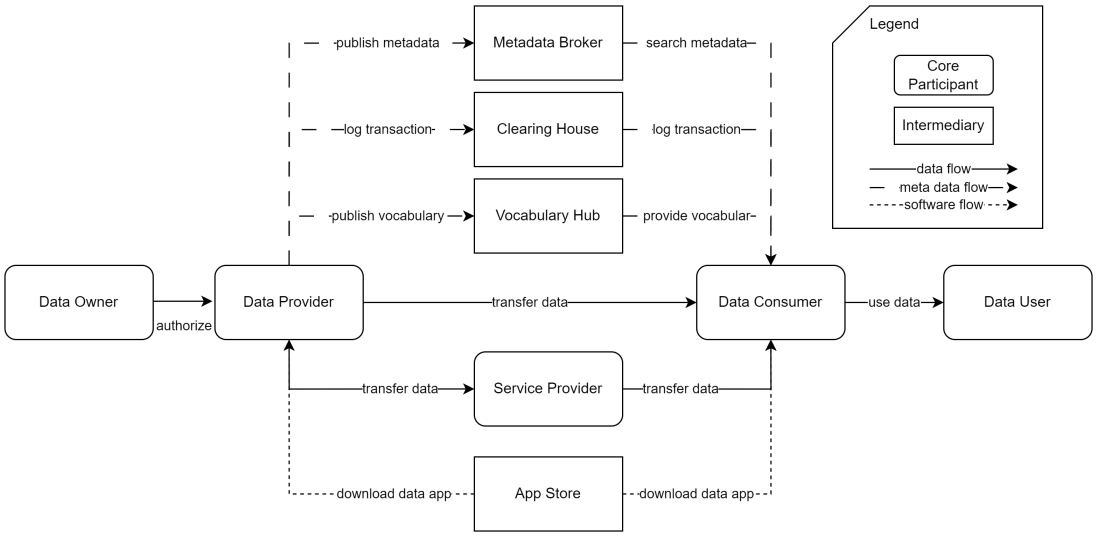

Figure 4.2: Roles and their interactions; Adopted from IDS RAM [41].

4.3.3 Supporting Roles

These actors are necessary to set up and initialization of the data space, but they do not involve in the everyday operations and transactions in the data space.

Software Developer

They develop the software stack needed for the operation of the data space. They create the data apps to be published in the app store. Furthermore, the Connector is also developed by the software developer and after proper certification it is used by the core participants.

Certification Body

It will perform evaluation of software components and participant's operational environment and issue a certificate for them.

4.3.4 Interaction Between Roles

A summary of all the roles and the interactions between them is shown in Figure 4.2.

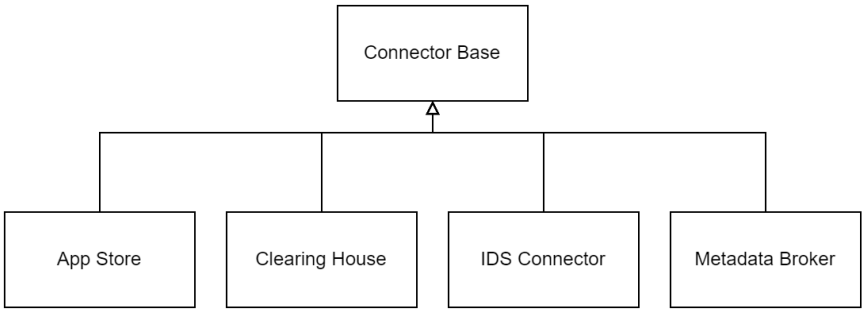

Figure 4.3: Different types of Connectors.

4.4 System View

4.4.1 Connector

The IDS network is a collection of many parts called connectors. A connector allows the data exchange via its Data Endpoints. A Participant may operate multiple IDS Connectors (e.g., to meet load balancing or data partitioning requirements). Other IDS components that require secure data exchange with IDS ecosystem, such as App Store, Clearing House, Metadata Broker, should conform to the Connector specification (Figure 4.3).

Components

The functionality of the connector could be decomposed to several functional components (Figure 4.4). These components are described below:

- Data Exchange: it provides and consumes APIs (i.e., application programming interface) to exchange data with other IDS participants (providers or consumers).
- IDS Protocol: it understands the IDS protocols and information model, and is able to perform IDS processes.
- (Optional) Remote Attestation: it answers the queries about the integrity of the Connector.
- Application Container Management: Connector is capable of running different isolated Apps concurrently. More information on IDS Apps is given in next section (4.4.2).
- Data App Management: supports the installation and uninstallation of IDS Apps.
- Policy Engine: It summarizes all components used for enforcing the IDS Usage Control Policies (more information on section 4.4.3).
- Contract Management: it performs the contract negotiation as specified in IDS protocol, and managing the Contract Agreements.

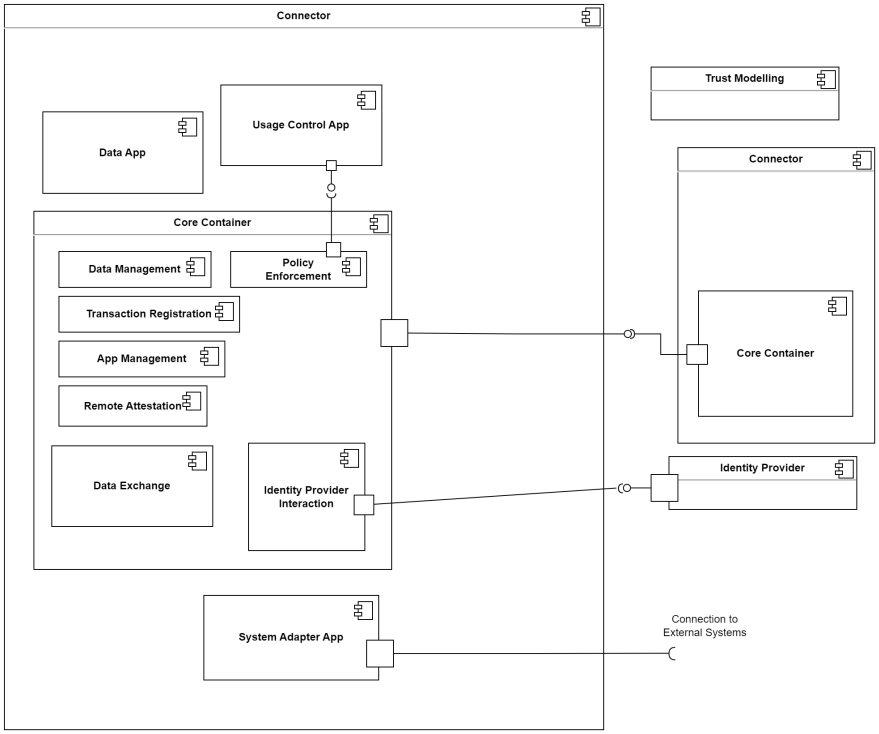

Figure 4.4: Connector Functional View

- Data Management: Stores and manages the data assets. In case of external data sinks, it keeps the link to the data and performs read and write via the data sink.

Provider Connector

Although a Connector can be deployed as both provider and consumer, the components active when performing each role is different. A Connector performing the provider role in our use case might be used for sharing monitoring data from a cybersecurity sensor to the data space. In such case, the usage control component is not needed, rather a system adapter app to collect the data from the sensor and a sanitization app to sanitize the data is needed. This architecture is depicted in Figure 4.5.

Consumer Connector

An example Connector playing the consumer role might be a connector used to display the received information in a controlled environment to a trusted security analyst. For this purpose, a certified data app for visualizing the received information is necessary. Needless to say, a usage control component is necessary to verify whether the data app is authorized to access the information. This architecture is depicted in Figure 4.6.

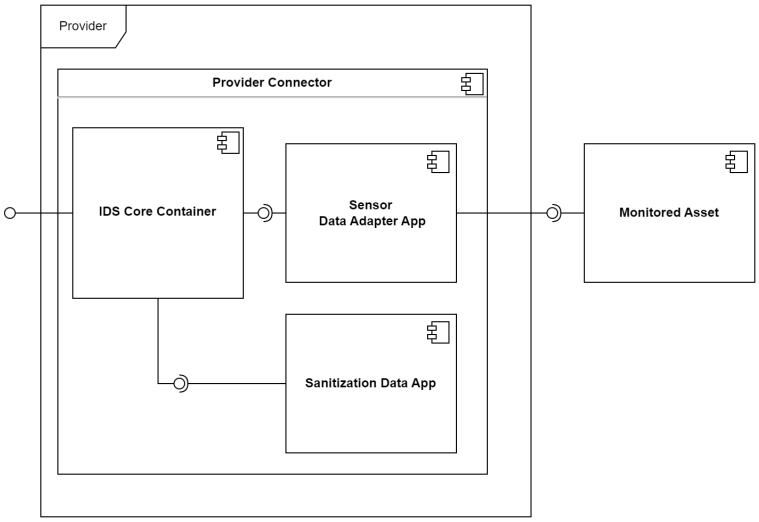

Figure 4.5: Provider Connector; Architecture of a connector providing data from a monitored asset.

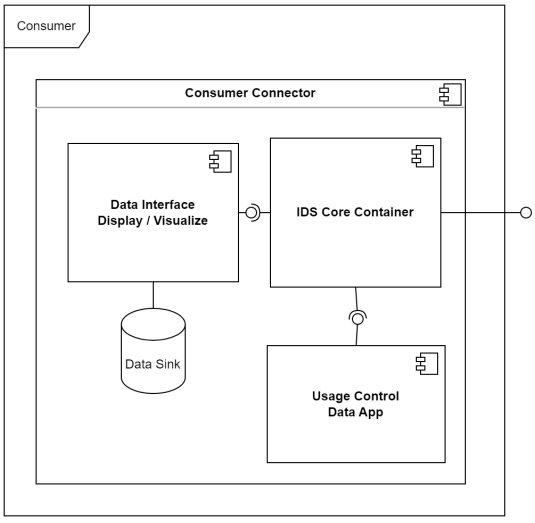

Figure 4.6: Consumer Connector; Architecture of a connector used to visualize the information in a usage controlled way.

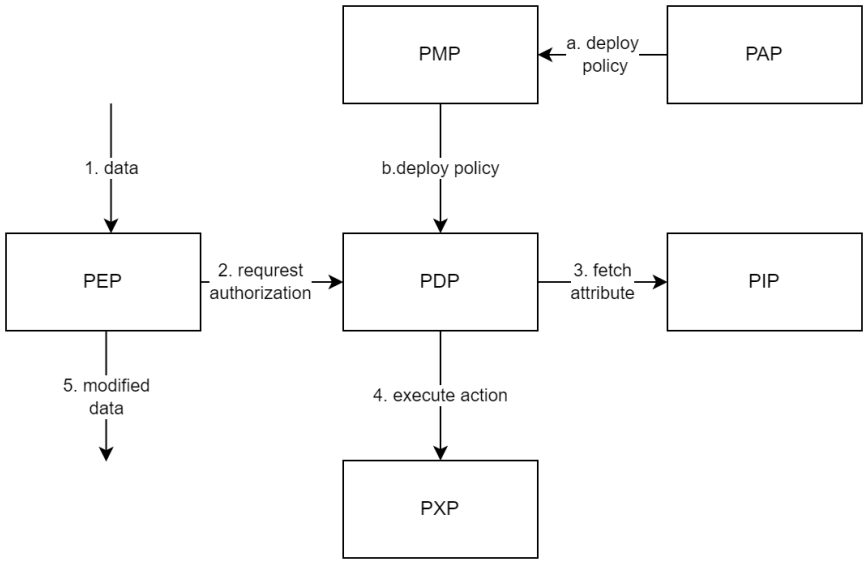

Figure 4.7: Policy Engine Components' Interaction Overview; Adopted from [16].

4.4.2 IDS Apps

IDS Apps allow the preprocessing of the data (e.g., filtering, anonymization, or translation), connecting to external backend systems, or control the Connector itself.

4.4.3 Policy Engine

The Policy Engine will cover administration and enforcement of policies. It follows a similar data flow as that of XACML (see 2.3.1) including PAP, PDP, PEP, PIP, obligation service, which is called PXP in IDS. A summary of the components and their interaction is depicted in Figure 4.7.

Policy Administration Point (PAP)

Policy Administration Point (PAP) or simply Policy Editor allows specification and deployment of policies. It should provide some template policies and a user interface to ease the specification by the user. Another feature of this component is the translation of the declarative, Specification-Level policies (SLP)s to Implementation-Level Policies. It is useful when we have different policy enforcement tools that have their own machine language.

Policy Management Point (PMP)

When data travels across organization boundaries, the policies associated to it must also be exchanged so that the receiver party get prepared for enforcing them. An approach

is sticky policies where the policy is embedded in the transferred data in the form of encryption and the data is decrypted only when the policy conditions met [47]. In IDS, a negotiation process exists before data exchange where these policies are exchanged. Policy Management Point (PMP) will collect and store these policies and track their life-cycle.

Policy Information Point (PIP)

A Policy Information Point (PIP) supplies the necessary information for decision-making. Furthermore, it can be used to obtain contextual information regarding the intercepted system action, such as data flow details or the geolocation of the requesting device.

Policy Execution Point (PXP)

PXP executes extra tasks in accordance with policy rules, like sending an email when data is accessed or logging the event in a designated system.

4.4.4 Identity Provider

To provide access control related features, an entity providing and managing identities is required. This entity is useful for performing both authentication (i.e., verifying an identity) and authorization (i.e., granting access to assets).

Certificate Authorities (CAs)

Certificate Authority issues X.509 certificates for participants upon request.

Dynamic Attribute Provisioning Service (DAPS)

DAPS issues short-lived tokens with up-to-date information about connectors. It is a trusted component by connectors providing current information about connectors, which eliminates the need for frequent certificate issuance and revocation. Also, it allows to selectively disclose attributes, which is not possible via normal X.509 certificate.

4.5 Process View

After defining actors and components of the data space, we will describe the processes existing in the system and the communications between the components and actors. We will present the data flows and activity diagrams.

4.5.1 Onboarding and Certification

An approach to reach trust is certification. In IDS, there are two types of certification, namely, operational environment certification and core component certification, both

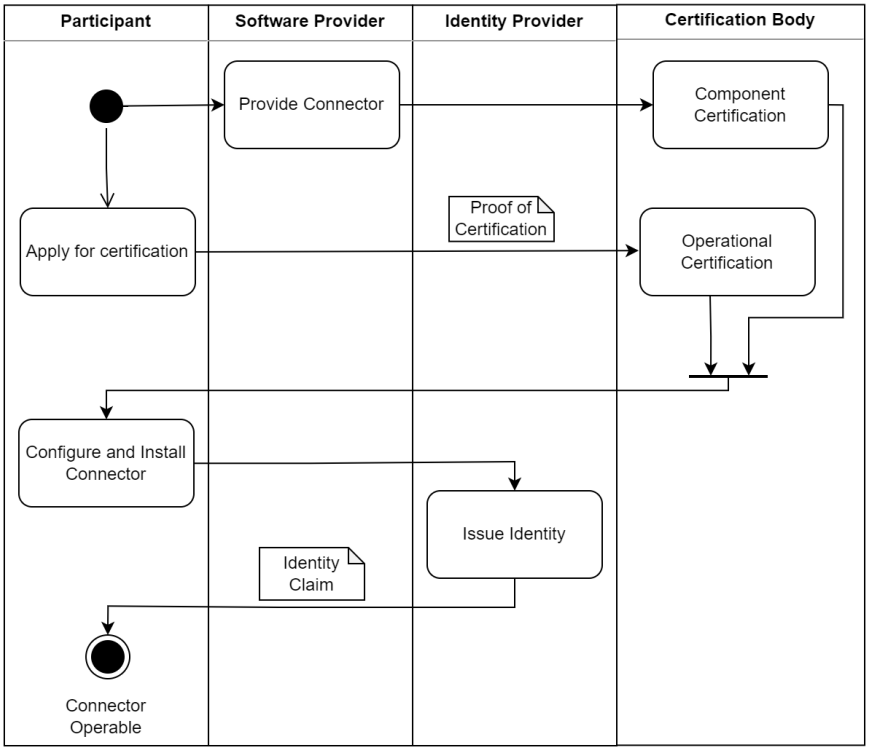

Figure 4.8: An activity diagram of the onboarding process of an IDS participant.

performed by the certification body. We will describe each in the following subsections, while proving a summary of the onboarding process in Figure 4.8.

Operational Environment Certification

Operational environment certification certifies the participants. The goal is to ensure that the participant is able to protect the data they access via the data space. Therefore, the process should evaluate the applicant's internal asset management capabilities, identity management, and physical security [61]. In order to reduce burden, the participant certification is designed to be compatible with existing certificates such as ISO/IEC 27001 [1]

Core Component Certification

The second certification type will certify the data space core components. This will evaluate both security and interoperability of the exchange. There are several assurance levels, namely, Base, Trust, Trust+. In the base level, the certification is done through a

[1] https://www.iso.org/standard/27001

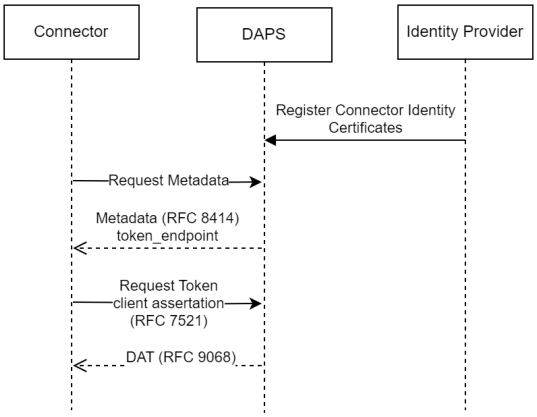

Figure 4.9: Summary of DAPS Interactions; It describes the sequence of messages between the DAPS, the Connector that obtains a DAT, and the Identity Provider which provides the Connector certificate.

checklist approach using a self-assessment. Higher levels involve concept review, testing, and source code audit. There are 3 trust levels, each having their own requirements: level 1—Data space interoperability, level 2—Feature complete for data usage control, and level 3—Additional protection from internal attacks [61].

Connector Identities

Connectors should have identity certificates to technically verify the trustworthiness of the communications. This is issued by the Certificate Authority (CA) in the form of an X.509 [2] certificate.

DAPS Interaction

Dynamic Attribute Provisioning Service (DAPS) as part of the IDS Identity Provider is used to provide signed dynamic attributes of the connectors. DAPS implements OAuth 2.0 Authorization Framework [3] and acts as an Authorization Server. It issues the Dynamic Attribute Token (DAT) which is a JSON Web Tokens (JWT) [4] containing verifiable information about the connector such as URI and the results of the certification such as security profile, and extended guarantee (e.g., supporting usage control enforcement). The receiving Connector will in turn utilize this DAT for communication with other IDS components. A summary of the communications with DAPS is depicted in Figure 4.9.

[2] https://www.itu.int/rec/T-REC-X.509
[3] https://datatracker.ietf.org/doc/html/rfc6749
[4] https://datatracker.ietf.org/doc/html/rfc7519

4.5.2 Publishing Data Offers

After successful on-boarding, the first step to perform data exchange is to advertise the data or service available by the data provider. To do so, the provider should prepare a catalog of resources that it offers along with attached policies, i.e., contract offer. This catalog is in turn published as part of the Connector Self Description via either an endpoint exposed by the connector itself or sent to the metadata broker.

Catalog Information Model

The catalog format is specified by the IDS Information Model, which builds upon and extends existing vocabularies, most significantly DCAT [5] and ODRL [6]. It comprises several data classes listed below and depicted in Figure 4.10:

- Resource Catalog: It is an extension of dcat:Catalog which aggregates the resources this catalog offers or requests.
- Resource: An extension of dcat:Dataset signifying digital content. It can specify the price, the endpoint to fetch the resource, a sample, and the owner of the resource. It also specifies the usage conditions in the contract object.
- Contract: Is based on odrl:Policy object and specifies a set of rules to govern the usage of the resource. Each rule can refer to an ids:Artifact, which can, but not have to, be associated with the resource in question.
- Distribution: Each IDS resource has some IDS representations, i.e., serializations, such as natural language, media-type or format. These are a subclass of the class Distribution in dcat.
- Artifact: It is the instantiation of the Representation that is used in the actual exchange.

Metadata Broker Interaction

As mentioned, the self-descriptions of the connectors, comprising the data offerings and requests, could be published via Metadata Broker. The sequence of the interactions are depicted in Figure 4.11.

4.5.3 Contract Negotiation

The provider specifies a contract offer as part of its self-description. In order to conclude the contract, both parties should sign. Therefore, a negotiation with the consumer is necessary. IDS Information Model defines the required message types communicated during a negotiation process, which is outlined in Figure 4.12. Three types of Contract Class is used, which differ in their semantics: Offer, Request, Agreement. There are different sequences for the process. First, the provider initiates a process by sending a

[5] https://www.w3.org/TR/vocab-dcat-2/
[6] https://www.w3.org/TR/odrl-model/

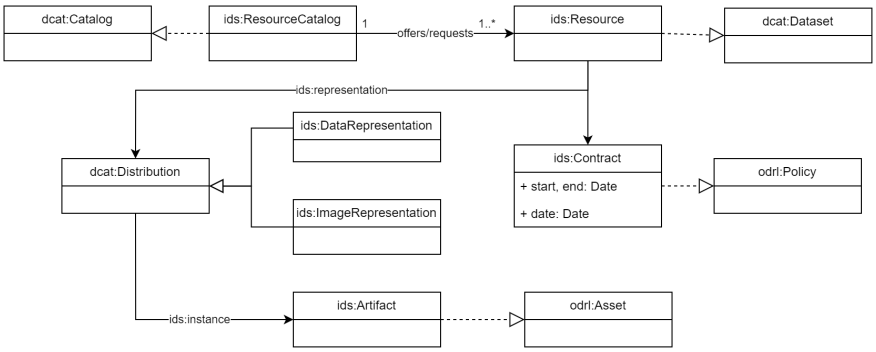

Figure 4.10: A simplified overview of the objects that comprise a resource catalog and their relation to ODRL and DCAT objects.

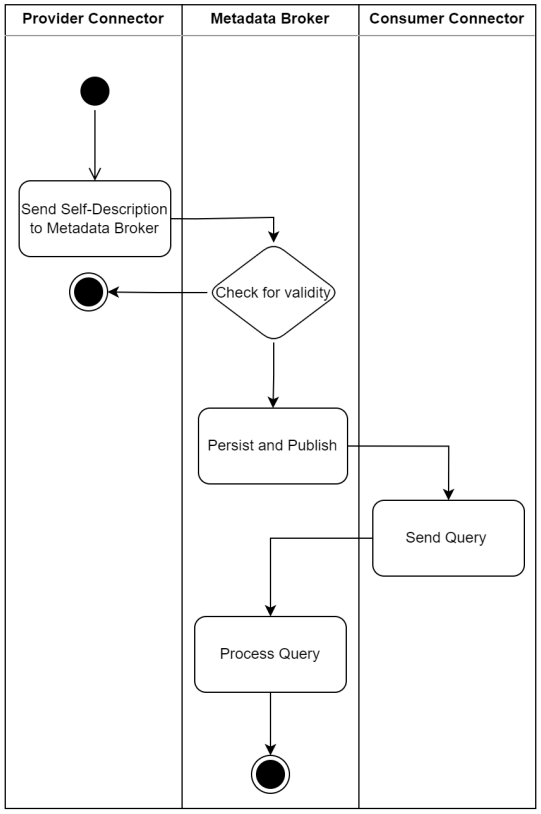

Figure 4.11: The process of publishing self-descriptions by the provider and querying them by the consumer, mediated by the metadata broker.

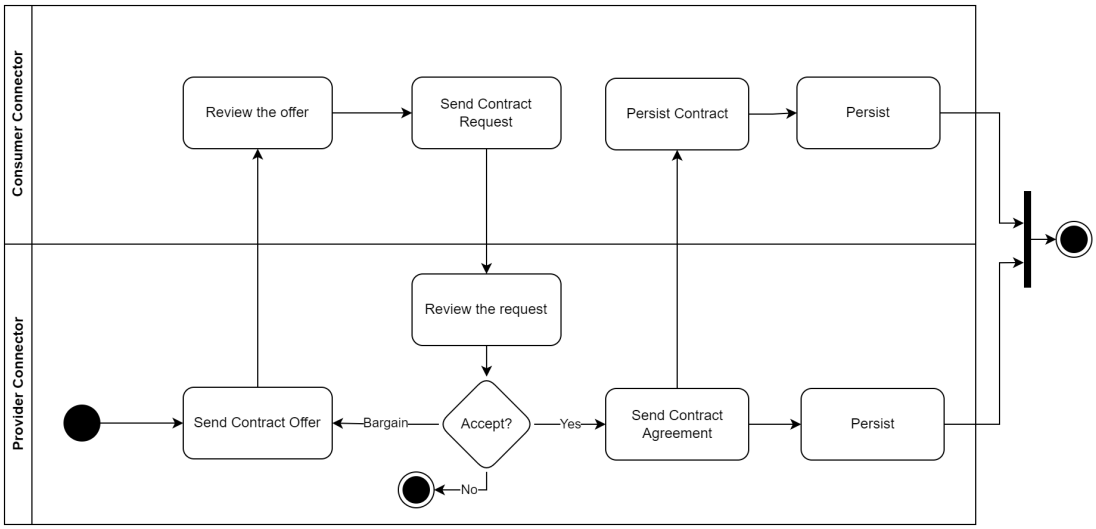

Figure 4.12: The contract negotiation process.

signed contract offer and the consumer in turn signs it and consequently and binding contract agreement signed by both parties is created. Second, the consumer takes the initiative by sending a signed contract request, which after being signed by the provider will result in a contract agreement. Third, a hybrid approach allows for subsequent offers and requests, mimicking a proper negotiation. In all cases, after reaching an agreement, the contract agreement could be sent to the Clearing House for a third party signature to allow the supervision of the contract observance by both parties by the Clearing House.

4.5.4 Data Exchange

The actual data exchange can happen after the on-boarding and contract negotiation. In this process, the consumer first fetches the metadata of the resource it wishes to acquire in order to check the different representations and find out the artifact's metadata. Then, it will send another request to fetch the selected artifact. This process is depicted in Figure 4.13.

Communication Protocols

IDS mentions three application level communication protocols for connector to connector communications in IDS-G [7]: Message passing based on HTTP/Multipart, IDS Communi-

[7] https://github.com/International-Data-Spaces-Association/IDS-G/blob/main/Communication/README.md

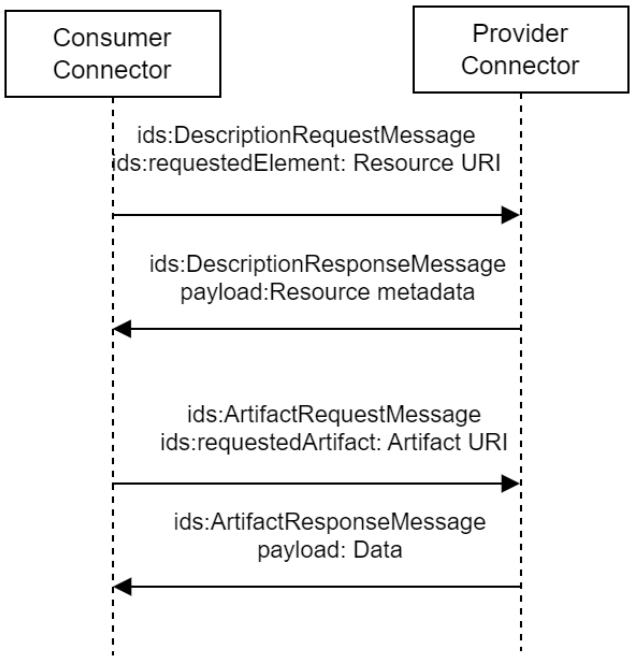

Figure 4.13: The data exchange flow. The data artifact is sent from the provider to the consumer.

cation Protocol Version 2 (IDSCP2) [8], and ids-rest [9] based on HTTP and REST. TRUE Connector also implements IDS Message passing on Web Socket Secure (WSS) standardized in RFC 6455 [10] which supports real-time communication between connectors with less overhead than normal HTTP. All protocols require the use of encrypted communications (SSL/TLS) and the provision followed by validation of the Dynamic Attribute Token (DAT) in the beginning.

4.5.5 Policy Engine

Here we describe, in more detail, how the policy enforcement works and how the policies look like. First, we will overview our policy language and the policy classes we need for our use case, and second, we look at the policy enforcement.

[8] https://github.com/International-Data-Spaces-Association/IDS-G/tree/main/Communication/protocols/idscp2

[9] https://github.com/International-Data-Spaces-Association/IDS-G/tree/main/Communication/protocols/ids-rest

[10] https://datatracker.ietf.org/doc/html/rfc6455

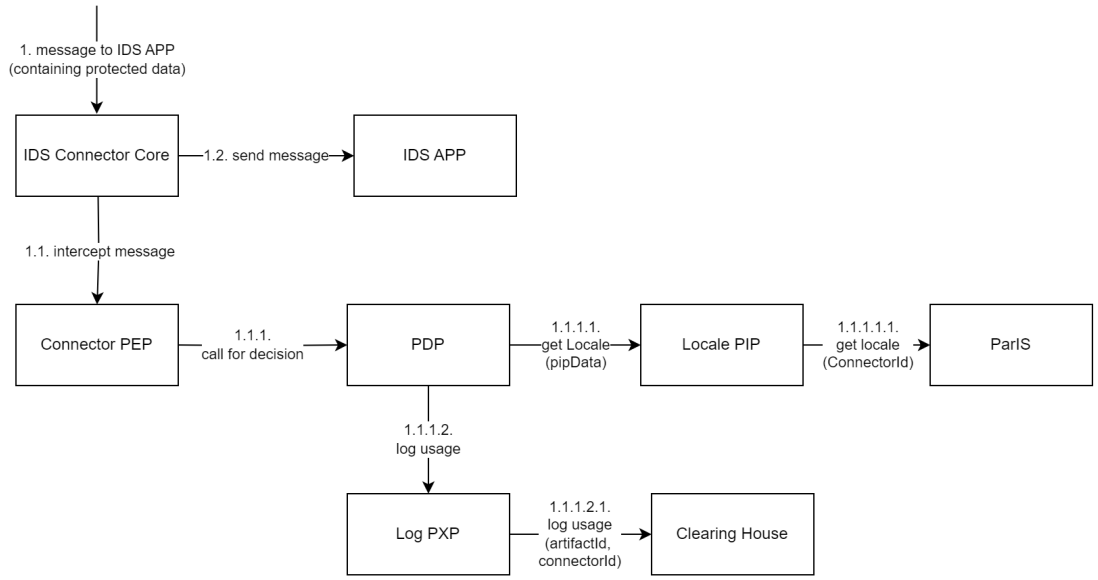

Figure 4.14: Example of usage control enforcement process; The policy states that the participants must be based in EU and the usage logged to the Clearing House [41].

Policy Specification

IDS defines a policy language which is based on ODRL and some added vocabulary. Rudolph [53] discusses ways to systematically define policies and how to make the process more user-friendly. IDS-G provides useful examples of how to describe different policy classes [11]. We will use these examples to specify our policies.

Policy Enforcement

To understand the process of policy enforcement, we look at how policy enforcement components interact to enforce the following policy: Data can only be used when the connector is in the EU (Locale) and the usage is sent to the Clearing House after data has been used (Log) (Example taken from IDS RAM). The provider ensures that the consumer has usage control capabilities during contract negotiation and accepted to enforce the policies. Next, the provider sends the requested artifact to the data app at the consumer side. The PEP will intercept the message, enforce the Locale constraint by requesting the ParIS, and then send the usage log to Clearing House, and finally continuing the data flow from PEP to the consumer IDS APP. This flow is depicted in Figure 4.14.

[11] https://github.com/International-Data-Spaces-Association/IDS-G/tree/main/UsageControl/Contract

Chapter 5

Realization

5.1 Tool Selection

In this section, we will take a programmer's perspective, focusing on the software components used to implement the logical components and processes described in the previous chapters.

We start the development by reviewing existing frameworks for the most critical aspects of our project, namely, IDS Connector and Usage Control Engine. We analyze and compare open source implementations to select one based on our criteria and consequently, develop those projects for our use cases.

5.1.1 IDS Connector

The IDS Association publishes a monthly report of the current state of all the data connectors used for exchange of data [12]. Dam et al. [10] investigated this report and published a survey in September 2023. They found that only 4 connectors have their source code available on a public repository: 1) IDS Dataspace Connector (DSC) by Sovity, Eclipse Dataspace Connector (EDC), the TRUsted Engineering (TRUE) Connector, and the Trusted Connector by Fraunhofer AISEC. In addition to that, I found two more: First, IDS Integration Toolbox by Open Logistics Foundation, which is a wrapper around the DSC. Second, TNO Security Gateway (TSG) initially developed by TNO which has implementations for many IDS components and is used in Smart Connected Supplier Network (SCSN) Dataspace [28].

Selection Criteria

We compared open-source IDS connectors by three criteria that were important to us:

1. Being actively maintained
2. Being well documented
3. Supporting usage control enforcement

Name	Created	Stars	Commits	Released	Hosted
DSC	07.10.2020	27[+101]	2600	10.22	Github
EDC	13.01.2021	202	1817	10.23	Github
TRUE	30.10.2020	19	122	08.23	Github
Trusted	05.09.2017	43	2221	02.23	Github
Toolbox	31.03.2022	3	172	04.23	Self-Hosted
TSG	12.05.2021	0	243	08.23	Gitlab

Table 5.1: Comparison of Open-source IDS Connectors

Selecting the Connector

The overview of different connectors is shown in Table 5.1. Regarding the first criterion, only EDC, TRUE, and TSG had a release after 05.23. TSG failed at the second criterion because of not having adequate documentation. So we compared EDC with TRUE Connector regarding the usage control features in section 5.1.3.

5.1.2 Policy Engine

There are different policy engines developed for IDS, which differ in the policy language supported, the enforcement method, support of provenance, being applicable in different systems, etc.

IDS RAM suggests the following policy engines [16]:

- **MYDATA** [1] is a policy-based usage control technology which is used for data sovereignty in a distributed environment. It is based on the IND2UCE (Integrated Distributed Data Usage Control Enforcement) [62] framework.
- **LUCON** [2] , first introduced by Schütte et al. [56], tries to achieve flawless policies with formal semantics and is embedded in the IDS Trusted Connector.
- **Degree (D°)** [3] Despite the former two that aim at providing usage control for existing applications, Degree's approach is to embed usage control during compile time of the applications.

The white paper "Usage Control in IDS" [16] compared the technologies in detail, and we attached their comparison table in the appendix (see 7).

Apart from the technologies suggested by IDS RAM, we found other technologies suitable for usage control enforcement in IDS:

- **Platoon** [4] is one that the TRUE Connector uses by default, however, it wasn't well documented nor supporting enforcement outside connector.
- **EDC** also implements its own Policy Engine and Policy Monitor, but we couldn't find adequate documentation on how to use it.

[1] https://www.dataspaces.fraunhofer.de/en/software/usage-control/mydata.html
[2] https://industrial-data-space.github.io/trusted-connector-documentation/docs/usage_control/
[3] https://www.dataspaces.fraunhofer.de/en/software/usage-control/d.html
[4] https://github.com/Engineering-Research-and-Development/true-connector-uc_data_app_platoon

Selection Criteria

- Maturity: It should have higher maturity expressed in Technology Readiness Level (TRL).
- Flexibility: It should allow policy enforcement in different scenarios, e.g., outside the Connector.
- Completeness: It should support all IDS defined policy classes.
- Documented: It should have adequate documentation.

Selecting the Policy Engine

MYDATA is best for our use case because first, it is more mature (i.e., Technology Readiness Level (TRL) of 7-8 compared to 5 (LUCON) and 4 (Degree)), second, it supports policy enforcement in the infrastructure (e.g., Database and external systems) and client system and services (e.g., operating system, external servers), and third, it supports all IDS defined policy classes [16].

5.1.3 Compatibility Analysis

In the section 5.1.3, we compared existing IDS Connector implementations and concluded that only EDC and TRUE Connector pass our first two criteria, namely, being actively maintained and having adequate documentation. In section 5.1.2, we concluded that MYDATA is the most suitable usage control technology for our use case. Therefore, we should find out which of these two IDS Connectors are compatible with MYDATA.

I started with trying EDC connector because of its more active GitHub status. I couldn't find any built-in support for MYDATA. I contacted the current maintainer of MYDATA, Robin Brandstädter [5], regarding integration with EDC, and his response was "It is not easy". I attached his explanation on this issue in the appendix 7. Therefore, we proceeded with TRUE Connector which had built-in support for MYDATA.

5.1.4 Minimum Viable Dataspace (MVDS)

To complete our Data Space setup, we need more components. The IDS association defines Minimum Viable Dataspace (MVDS) as the minimum set of components that provide the ability to do secure and sovereign data exchange. They specify the required components as follows: Two Connectors (a data provider and a consumer), an Identity Provider (Dynamic Attribute Provisioning Service, Certificate Authority).

IDSA also has published an open-source project, IDS Testbed, that contains instructions to install and orchestrate these set of minimum components. It references open-source implementations of these components, see table 5.1.4 to find source code of these components.

[5] robin.brandstaedter@iese.fraunhofer.de

Component	Source Code	Version	Language
IDS Testbed	Testbed Git	1.0	Docker-Compose
Dataspace Connector (DSC)	Connector Git	8.0.2	Java
Metadata Broker	Broker Git	5.0.3	Java
DAPS	DAPS Git	1.6.0	Ruby
Certificate Authority	Testbed Git	—	Python

Table 5.2: List of Components Used by the IDS Testbed.

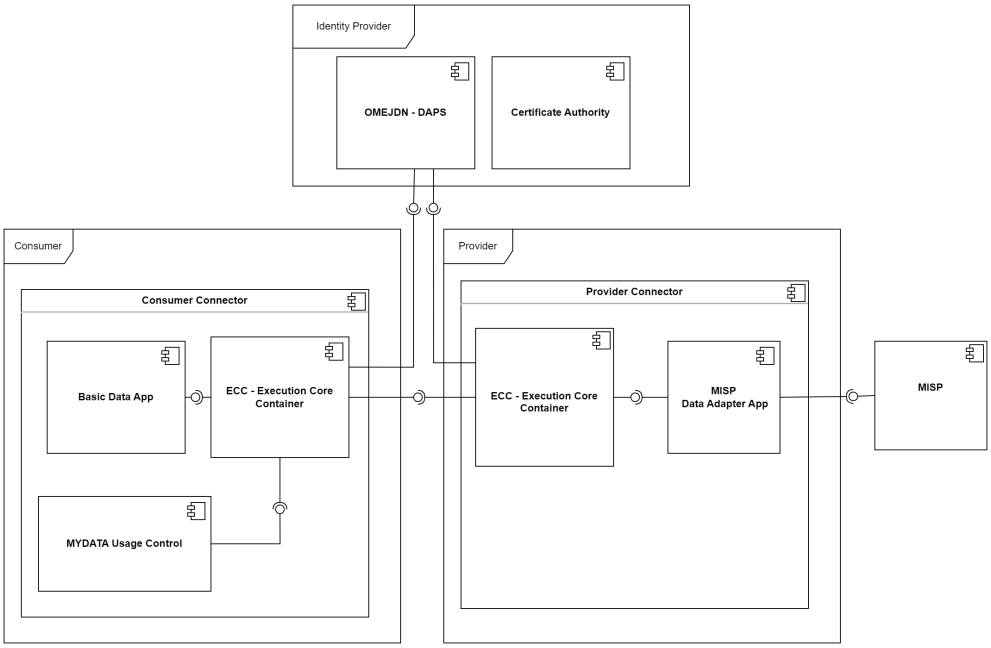

Figure 5.1: Deployed Components and Their Interactions.

5.2 Deployment

We created a fork of the selected open source tools to apply our configurations and deployed everything in a virtual machine. The overview of resulting components is depicted in Figure 5.1. We will discuss each component in this section.

TRUE Connector

IDS Connector Architecture is based on application container technology to allow isolation of IDS Apps from IDS Core while keeping the connector light-weight (Figure 5.2.

We installed TRUE Connector using the Docker-Compose specification provided by them. We configured the docker-compose to use MYDATA instead of the built-in usage

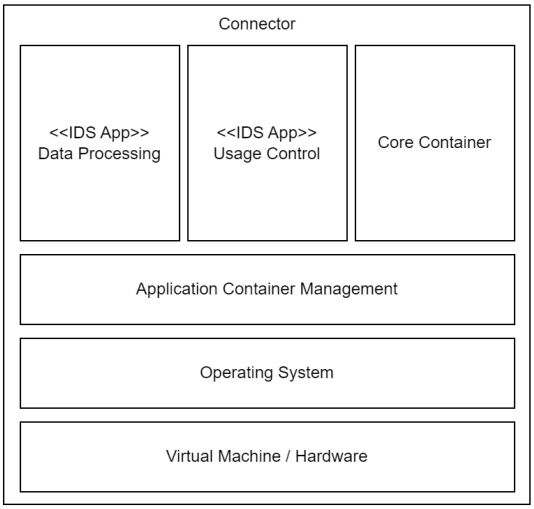

Figure 5.2: Connector Physical Architecture; IDS Apps are realized as Docker containers to keep them isolated while keeping the whole connector light-weight.

control. Useful documentation could be found here [6]. To start the docker containers, we ran the following command:

```
1    docker-compose up
```

It starts the following Docker containers, for the provider and the consumer:

- *ecc*: The core container executing IDS protocol.
- *uc–dataapp*: The usage control app running MYDATA.
- *be–dataapp*: A sample data app. In this demo, it stores and loads data from disk, which is a mounted docker volume.

IDS Testbed

We installed the *DAPS* service from the IDS Testbed using *docker*. Also, we used the certificates available in the Certificate Authority folder as our connector certificates. As the *DSC* connectors in the IDS testbed were not the focus of us, we only deployed the *DAPS* service with the following:

```
1    docker-compose up omejdn
```

[6] https://engineering-ing-inf-rd.gitbook.io/true-connector/advanced-doc/mydata_usage_control

Chapter 6

Evaluation

6.1 Investigation Procedure

Our investigations consist of four independent areas of study:

- **Application to Example Use Case Scenarios:** To show that our solution is applicable in some real-world scenarios, we will present some examples and explain how our solution could be utilized in each step of the scenario. In doing so, we will describe the data that is exchanged, the system components used and the roles that are involved in those scenarios.
- **Validation of the Policy Framework:** An adequate policy framework is crucial to success of a CTI sharing framework. To evaluate the usefulness of our policy framework in CTI sharing use cases, we measured the comprehensiveness of the policy classes by comparing it to a widely used policy framework for CTI sharing. We measured the proportion of the policy classes from the target policy framework that is covered by our policy framework.
- **Technical Tests and Metrics:** We will analyze some performance metrics of our solution using the implemented prototype as benchmark. Due to the importance of latency and resource consumption in our use case, we measured them in our experimental tests.
- **Analytical Verification of the Architecture:** To verify that the identified requirements are fulfilled by our solution, we will do an analysis of the requirements and evaluate the extent to which the requirement is satisfied. We would also compare it with widely used platforms for CTI sharing in practice.

6.2 Results

6.2.1 Application to Example Use Case Scenarios

We will proceed with the scenario we described in section 3.3.3 and refine it with more details to be able to simulate it with our prototype. It comprises three data exchanges

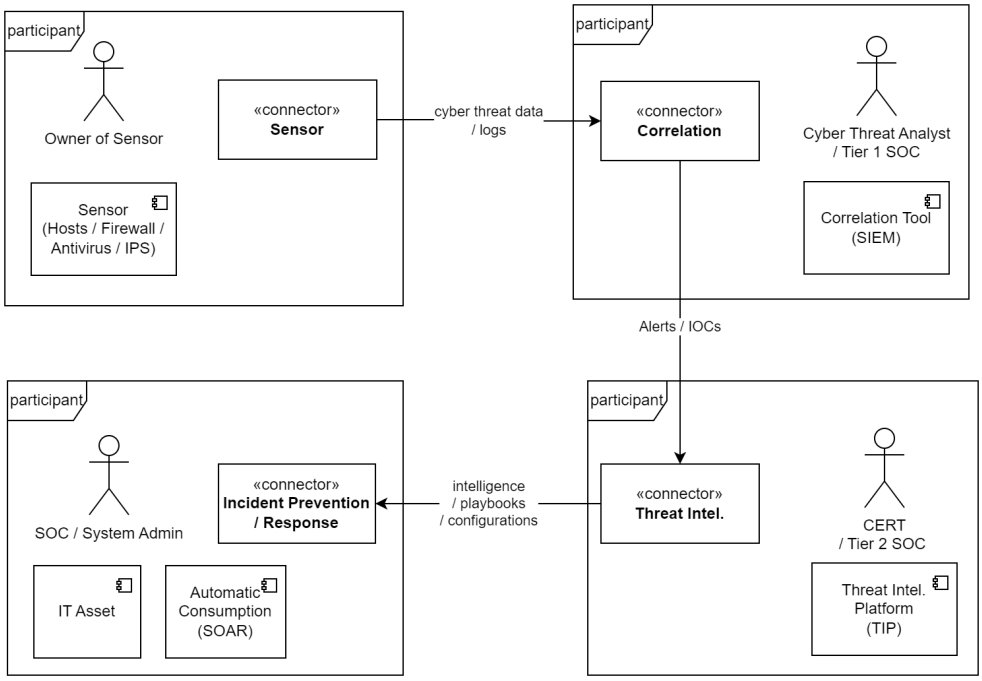

Figure 6.1: Validated Real-World Scenario; Four participants, each operating a connector and using respective backend systems exchanging different data types.

with different data and associated policy, which is illustrated in Figure 6.1.

Exchange 1: Organization Outsourcing Security Analysis

Org A is a small electricity provider that operates a smart grid, which is considered Critical Infrastructure and is therefore subject to incident reporting regulations. It does not host in-house SOC team and thus outsources the reporting task to an external SOC (SOC A). It continuously shares its security sensors data to SOC A and SOC A notifies as soon as it detects an anomaly.

A firewall could act as a sensor, and thus we downloaded a log dataset to test – see appendix section 7 for information about the dataset. The dataset contains personal data, so the exchange should be compliant with GDPR. It uses Sanitization App to remove and anonymize personal data. It is an IDS App that is developed by peers in the Energy sectors to process the same type of data.

To further ensure safety of the information, as the dataset contains internal network topology of Org A, it sets some policies for the proper handling of data. To set the policies, it uses the user interface provided by PAP component. The resulting policies are available in the appendices in the JSON-LD format (see section 7). The policy offers the data to participants located in EU with two permissions: First, use the logs by an

specific IDS App which does log management and correlation, a SIEM equivalent, and second, use of the logs by external apps only when aggregated and in Germany.

Exchange 2: SOC Sharing Incident Data with the Community

SOC A after analyzing data from Org A has detected an anomaly and indicating an incident. SOC A wants to share the IoCs found in this incident to help peers and get financial reward. It is a member of a CTI sharing community for energy sector, i.e., ISAC, which uses our framework for exchanges between its members. The community allows a provider-defined billing scheme to encourage the members to share. The community manages a metadata broker to provide a catalog of all data available to the community. When a member is approved to join the community, the community manager will contact the identity provider to register the participant. This membership will be saved into the DAPS as an attribute of the joining member. Consequently, to limit the access to messages to the members of the community, SOC A will check the DAT token of the consumer to see if it has the membership.

The Certification process validates the identity and trustfulness of the participants and therefore the participants could ensure that only the community member will get access to the data.

Exchange 3: National CERT Notifying a Constituent Organization

A national CERT wants to notify an organization about a new threat. Due to the high risk associated with the threat, the CERT wants to ensure proper handling of data by the Consumer. It knows that the Consumer uses a SOAR system to automatically respond to threats. The CERT will set a usage policy rule to only let the SOAR system on the consumer side to have access rights to the data. This way, not only the classified information is protected, but also the response is performed quickly. The CERT also provides the report for manual consumption by the SOC analyst. These reports are commonly labeled with TLP to specify how the data could be distributed. In section 6.2.2, we discussed how TLP labels could be expressed with our policy framework.

6.2.2 Validation of the Policy Framework

Selecting a Policy Framework as a Benchmark

We searched for a policy framework that is commonly used by CTI sharing communities to compare our solution with it. We decided for the Information Exchange Policy 2.0 Framework (IEP) [35] set by the Forum of Incident Response and Security Teams (FIRST) [1], established in 1990, which is a prominent community with 756 members in 111 different countries. IEP defines several different classes of statements that comprise a policy, which we summarized them in Table 6.1. We will evaluate the support of our

[1] https://www.first.org/

Policy Class	Meaning
ENCRYPT-IN-TRANSIT	Encrypt when retransmit.
CONTACT FOR INSTRUCTION	Must contact the provider for instructions.
INTERNALLY VISIBLE	Only actions that are visible in internal networks and systems.
EXTERNALLY VISIBLE INDIRECT	Only indirect, passive actions outside internal network.
EXTERNALLY VISIBLE DIRECT	Any actions based on the information is permitted.
NOTIFY-AFFECTED-PARTY	Permission to notify affected parties of a potential compromise or threat.
TLP:RED	Redistribution is not permitted.
TLP:AMBER	Redistribution permitted on a need-to-know basis within the recipient organization and its clients.
TLP:GREEN	Redistribution permitted within the community.
TLP:CLEAR	Redistribution permitted publicly.
PROVIDER-ATTRIBUTION	Consumer MAY/MUST/MUST NOT attribute the provider when redistributing.
UNMODIFIED-RESALE	Permission to resell the information received unmodified or in a semantically equivalent format.

Table 6.1: List of Policy Statements Supported by IEP [35]. This serves as a benchmark to evaluate our policy engine.

proposed solution for each policy class. If some requirement is not implemented, we will estimate of how difficult it is to extend the solution to fulfill it.

Fulfillment Levels

To assess the extent to which the aforementioned policy classes are supported, we investigated both policy specification and enforcement capabilities of our solution separately. First, we assessed whether the policy specification language can express a policy class by manually comparing it with the IDS vocabulary. We also looked into the ODRL vocabulary, which is used by the IDS Information model. The result was either one or some terms defined in the IDS vocabulary, or found out it is missing and an extension of the vocabulary is needed. Second, to assess the policy enforcement capabilities, given the policy specification exists, we subjectively estimated the difficulty of extending the policy engine to enforce them automatically. We used a four-level assessment, which is presented in Table 6.2.

Implementation Difficulty	Description
ZERO	The implemented prototype can enforce it, or enforcement is not needed.
LOW	Enforcement is possible with implementing missing policy engine components, i.e., PIPs/PXPs.
MEDIUM	Enforcement is possible with the extension of IDS specifications and existing components, such as Clearing House.
HIGH	Enforcement requires implementation or strict monitoring of complex domain specific workflows, e.g., forensic actions.

Table 6.2: Policy Enforcement Implementation Estimated Difficulty Levels and Their Descriptions.

Results

The validation results of the policy framework are presented in Table 6.3. We found that for 7 policy classes, the IDS information model has enough vocabulary to express the meaning of these classes completely. For 6 policy classes we need to define new vocabulary which is possible through the designated Vocabulary Hub. For example, CONTACT FOR INSTRUCTION is an obligation which is expressed as odrl:Duty, however, to express the action we need to define new vocabulary. TLP levels are also possible to be expressed with ODRL using constraints and the Distribute action, however for the AMBER level, we need to define the condition "NEED-TO-KNOW" in a new vocabulary.

As for the enforcement, ENCRYPT-IN-TRANSIT is already implemented by MY-DATA, and TLP:CLEAR and EXTERNALLY-VISIBLE DIRECT do not require any enforcement. NOTIFY-AFFECTED-PARTY can be easily implemented, we need a PIP that will assess if the recipient is affected by a vulnerability or incident and only then permits the data flow. TLP:RED is also easy to implement, the PDP should check the Distribute permission. TLP:GREEN can be implemented by the PIP checking if the recipient is part of the community. PROVIDER-ATTRIBUTION can be implemented by a PIP that checks if the distributed data has attribution to the original provider or a PXP that automatically removes the attribution. UNMODIFIED-RESALE needs more changes in the implementation. We should implement a resale functionality to automatically check for this field on resale. Also, for more safety, clearing house could be extended to check the authenticity of the commercial transactions.

IEP Policy Class	**IDS Information Model Object**	**Implementation Difficulty**
ENCRYPT-IN-TRANSIT	ids:DistributeEncryptedAgreement	ZERO
CONTACT FOR INSTRUCTION	odrl:Duty & Extended Vocabulary Needed	HIGH
INTERNALLY VISIBLE	Extended Vocabulary Needed	HIGH
EXTERNALLY VISIBLE INDIRECT	Extended Vocabulary Needed	HIGH
EXTERNALLY VISIBLE DIRECT	Extended Vocabulary Needed	ZERO
NOTIFY-AFFECTED-PARTY	ids:Permission and odrl:Distribute & Additional Vocabulary (Affected)	LOW
TLP:RED	ids:Prohibition & odrl:Distribute	LOW
TLP:AMBER	Extended Vocabulary Needed (Need-to-know)	HIGH
TLP:GREEN	odrl:Distribute & odrl:Recipient & odrl:Refinement & odrl:NextPolicy	LOW
TLP:CLEAR	odrl:Permission & odrl:Distribute	ZERO
PROVIDER-ATTRIBUTION	odrl:Distribute & odrl:Attribute	LOW
UNMODIFIED-RESALE	odrl:Commercialize & odrl:Distribute	MEDIUM

Table 6.3: Mapping of IEP to IDS Policy Engine; Column 2: The suggested IDS Information model object to express each IEP policy class. Column 3: An estimate of the implementation effort to enforce the policy automatically.

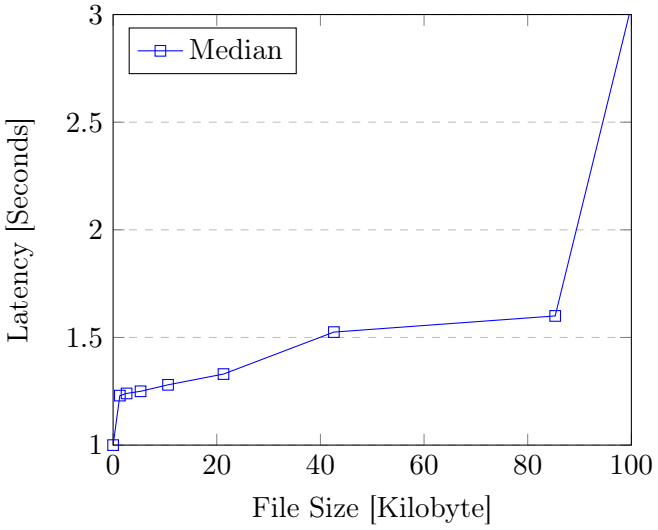

Figure 6.2: Latency Measurement; Based on an example data exchange in our platform.

6.2.3 Technical Tests and Metrics

Latency

Given the importance of timeliness in the usefulness of CTI information, we measured the latency of a data exchange using the prototype we developed. We fetched some OSINT on a publicly available MISP feed using our MISP client and saved it into a file. It summed up to 75 records with the size of $\sim 100.0KB$. We created 6 truncated versions of it with different sizes. We uploaded it to the Provider Connector and created a usage policy to allow the usage of it by the Consumer Connector. Finally, we measured the latency of transferring it from one Connector to another. We repeated the measurements 10 times for each file, and we calculated the mean across the values. We also tested a non-existing file and measured the response time. We ran these experiments on a virtual machine with $32Gb$ of RAM and Intel Xeon Processor with 8 Cores each running with $2.1 - 2.3Ghz$ frequency. The measurements are illustrated in a plot in Figure 6.2. The results indicate a range of 1–3 seconds, which includes the authentication and authorization as well as enforcing the usage policies. This latency is acceptable for most use cases. However, more complex deployments with more distribution and network delay or more complex usage policies might lead to different results.

Resource Usage

Another metric is resource usage that plays a significant role in the operation costs of the platform. We measured the memory usage of different components of the prototype and the results are shown in Table 6.4. It amounts to $1.6GB$ for each connector, which is affordable for most use cases.

Container Name	Memory Usage
uc-dataapp-consumer	1.1 GiB
be-dataapp-consumer	0.2 GiB
ecc-consumer	0.3 GiB
uc-dataapp-provider	1.1 GiB
be-dataapp-provider	0.2 GiB
ecc-provider	0.3 GiB
DAPS	0.06 GiB

Table 6.4: Memory Usage of Different Containers in the Prototype.

6.2.4 Analytical Verification of the Architecture

To verify that the identified requirements are fulfilled by our solution, we did an analysis and discussed how each requirement is satisfied by our solution. To better understand the novelty of our approach, we will perform this analysis for some existing solutions: namely, MISP and ThreatConnect. We present the results in Table 6.5 and discuss the rationale as follows:

Interoperablity

To achieve interoperability and avoid vendor lock-in, an approach is to use open standards, data models. IDS defines an open standard and some open source tools. The standard is delivered via its IDS RAM, which references its Information Model and its open protocols. It is based on existing open standards such as W3C ODRL and IETF OAuth. MISP is free and uses open data models and standards. CrowdStrike on the other hand is, similar to most other proprietary services, not open-source.

Flexiblity

Due to the numerous requirements of an overarching CTI sharing framework, a solution should be flexible and expandable. It should define adequate extension points to implement new requirements for data processing, data models, third-party systems, and data policies. IDS defines the concept of App Store to implement new data processing methods and adapting to different external systems. It allows management of new data models with the Vocabulary Hub. ThreatConnect is a flexible solution, however its expansion is bounded by the vendor resources. Open source solutions like MISP are highly expandable due to the large community behind them.

Trust

Trust is a central design goal of IDS. We divide aspects related to trust in three and discuss each separately.

- Component Certification: IDS ensures integrity of the software components by the Certification of the components and IDS Apps done by third party certification

Sharing Platform		Our Solution	MISP	ThreatConnect [a]
Approach		IDS Based	Open Source	Vendor-Driven
Sharing Model		Hybrid	Hybrid	Hub and Spoke
Implemented Requirements				
I	Open Standard	Yes	Yes	No
F	Different Data Models	High	High	Limited
	External Integration	High	High	Limited
T	Component Certification	3rd Party	Local Components	Self Certified
	Participant Certification	3rd Party	Possible	By Vendor
	Multi-Level Participant Trust Level	Yes	No	No
	Dynamic Trust	Yes	No	By Vendor
C	Data and Service Marketplace	Flexible	Limited	No
	Digital Rights Management	Yes	No	No
D	Distribution Control	In OS	In Platform	In Platform
	Usage Control	Yes	No	No
	Automatic Sanitation	Yes	Yes	Yes

[a] https://threatconnect.com/

Table 6.5: Comparative Analysis; Comparison between our solution and other CTI sharing solutions based on the implemented features; Features are categorized as **I**nteroperability, **F**lexibility, **T**rust, **C**ommercial, **D**ata Privacy and Sovereignty; We chose Crowdstrike and MISP as examples, and most of the results also apply to other alternatives of the same type.

body based on a transparent certification process. Proprietary tools are usually black-box and obfuscated, so only the vendor itself can certify it. Open-source tools, e.g., MISP, are verified by the community, however, the component used by the remote user cannot be verified.

- Participant Certification: IDS defines an Identity and trust management protocols that are based on state-of-the-art encryption methods. Furthermore, any participant will go through a certification that ensures proper data handling capabilities, which is done by a third party Certification Body. In MISP, it is possible to set some certification as requirement of joining a community, however, it is not embedded in the protocol. In ThreatConnect, the vendor manages the trustworthiness of the participants.
- Multi-Level Participant Trust Level: In IDS, multiple levels of trust levels are defined (i.e., Security Profile), and it is possible to filter a specific trust level during

sharing. However, in MISP and ThreatConnect, a participant is either trusted or not.

- Dynamic Trust: In IDS, DTM monitors participants continuously and the access could be revoked or reissued easily with DAPS. In MISP and ThreatConnect, this process is manual and requires reconfiguration of the instances.

Commercial Activities

In IDS, it is possible to sell data and services with flexible billing models based on usage enabled by the Clearing House. Also, Software Developers could sell their Applications in the App Store. It is possible to provide paid or subscription based services in MISP, however due to not having DRM it makes the intentional or accidental illegal sharing probable. ThreatConnect is aimed at providing their own premium content and services and is not a marketplace.

Data Privacy and Sovereignty

IDS achieves data sovereignty with strict enforcement of usage control even after data has travelled across organizations.

- Distribution Control: All platforms implement distribution control, i.e., TLP. However, only IDS can restrict distribution outside the platform with technical measures.
- Usage Control: IDS can enforce technically prohibitions, permissions, and obligations attached to the data, which is not the case for other platform.
- Automatic Sanitization: Removing trade secrets, personal information, classified information before sharing automatically is possible in IDS with IDS Apps. Other platforms also provide some sanitation functionalities.

6.3 Discussion

We performed 4 different evaluation tasks, each highlighting the potentials of Data Spaces in its own way. First, we applied our solution in a practical CTI sharing scenario and discussed how our platform will fulfill the requirements and saw a significant effectiveness of our solution. Second, we dove deep in the policy engine of our platform, and we saw a decent compatibility of our policy engine with a widely used policy framework in practice. Third, we found the adequacy of the performance metrics through our experimental analysis. And finally, we demonstrated the distinguishing merits of our solution compared to the platforms used in practice.

The results indicate that the IDS architecture model is a viable solution for CTI sharing, offering a robust framework that addresses key challenges such as data privacy, trust, and interoperability. It calls for further investment in data space-based approaches to implement different scenarios of CTI sharing in practice.

The suitability of IDS can also apply to other use cases beyond CTI sharing, provided that a business data exchange in a sovereign way is part of the use case.

Chapter 7

Summary and Outlook

Summary

We found significant results for all of our objectives:

- **Objective:** Find specific CTI sharing use cases and scenarios where current platforms cannot adequately address the challenges and find out why.

  Contributions: We identified 5 high-level requirements that current systems do not cover all 5 items completely. We described real-world use cases that will prove these requirements can happen together in practice and are not imaginary. The requirements are as follows: Interoperability, flexibility, trust, data sovereignty, and commercial.

- **Objective:** Find the degree to which dataspaces can alleviate the challenges in selected CTI sharing scenarios.

  Contributions: We designed an IDS based architecture model for CTI sharing for the identified scenarios and implemented a prototype based on it. We evaluated the architecture model and the prototype against the scenarios and requirements, and all evaluations suggested strong capabilities of our solution.

- **Objective:** Find the implementation considerations when setting up a dataspace for CTI sharing.

  Contributions: We performed a comparative analysis of the available implementation frameworks to implement an IDS based dataspace. We extended those implementations to build our prototype, and documented our findings. Our source codes are available on request.

Future Works

Our analysis showed that IDS is a promising approach in cybersecurity. Therefore, a pilot project with real actors involved would allow for better evaluation of the platform and some empirical results. Thereby, we could complete this work by adding a user based evaluation of more subjective aspects, such as usability and perceived trust. It also calls

for a more detailed requirement analysis of the data apps required in the consumer side to implement more use cases.

Appendices

Comparison of Usage Control Technologies

In the white paper "Usage Control in IDS" [16], the usage control enforcement technologies supporting IDS is listed and compared. A table summarizing their discussion follows (Table 1).

	MYDATA	**LUCON**	**Degree (D°)**
Purpose	Usage Control Enforcement	Control of data flows, enforcement of obligations dependent on data flows	Development of data processing applications with integrated usage control.
Documentation	https://www.mydata-control.de/	https://industrial-data-space.github.io/trusted-connector-documentation/docs/usage_control/	T3 – Deliverables (available on request)
License	Open-Source SDK: Apache 2	Apache 2	Apache 2
Programming Language	Java	Java	D° (Java as Host)
Management	On premise hosting and Java library.	IDE & LUCON Environment	IDE & D° runtime environment
Graphical User Interface	Web UI	Web UI (IDS Policy Editor)	No
TRL	7-8	5	4

Table 1: General Usage Control Technologies Comparison [16].

Integration of EDC and MYDATA

I contacted the current maintainer of MYDATA, Robin Brandstädter [1], regarding integration with EDC, and he believed it is not easy. Here is his explanation:

"On the question of integration with the EDC. It is not easy: We had a research project on this and had to give up, in principle we had an approach. Unfortunately, the EDC failed to accept the IDS and ODRL policies. I can try to explain the approach again. After several attempts to do it according to the EDC documentation (to use or define scopes) we came to the conclusion that it must be solved via the listeners offered by the EDC. Because otherwise you work with the edc policy engine, which has already decomposed the policies, but we have a completely different approach and need a complete policy to deploy in MYDATA. In other words, we need a different policy engine and this must be integrated via the listener. Now the policies still have to be transferred via the EDC. However, the EDC uses its own language, which means that it has to be translated from ODRL into the EDC language and back again. The EDC should actually be able to do this. .. I hope it can by now. The task is then to integrate the library that we have developed to translate ODRL into MYDATA policies and deploy them."

Policy Source Codes for Example Scenario

In the following, we attach the policies we defined for the example use case scenario. For ease of readability, we present it in separate listings. In listing 1, the structure of an IDS Contract is presented, which is the same for all three policies we defined. In listing 2, we present how to limit the valid consumers of the data. In listing 3, an IDS permission is presented which will grant the usage permission of the logs to the assignee by the SIEM app.

```
1  {
2      "@context": {
3          "ids": "https://w3id.org/idsa/core/",
4          "idsc": "https://w3id.org/idsa/code/",
5          "cti": "https://w3id.org/cti/dataspace/"
6      },
7      "@type": "ids:ContractOffer",
8      "@id": "https://w3id.org/idsa/autogen/contractOffer/
       sensor",
9      "ids:contractStart": {
10         "@value": "2024-03-29T20:43:42.331Z",
11         "@type": "http://www.w3.org/2001/XMLSchema#
       dateTimeStamp"
12     },
13     "ids:contractDate": {
```

[1] robin.brandstaedter@iese.fraunhofer.de

```
        "@value": "2024-03-29T20:43:42.939Z",
        "@type": "http://www.w3.org/2001/XMLSchema#
    dateTimeStamp"
    },
    "ids:provider": {
        "@id": "http://w3id.org/engrd/connector/sensor"
    },
    "ids:consumer": {
        ...
    },
    "ids:permission": [
        ...
    ]
}
```

Listing 1: The structure of an IDS Contract Offer; It is represented in JSON-LD, has references to IDS vocabulary, i.e., IDS Information Model, and specifies the contract clauses such as participants, date, usage rules.

```
"ids:consumer": {
    "ids:participantRefinement": [
        {
            "ids:leftOperand": "vcard:hasGeo",
            "ids:operator": "ids:IN",
            "ids:rightOperand": [
                {
                    "@value": "http://ontologi.es/place/EU",
                    "@type": "xsd:anyURI"
                }
            ],
            "ids:pipEndpoint": {
                "@type": "ids:PIP",
                "ids:interfaceDescription": {
                    "@value": "https://example.com/ids/pip/id
    /geo",
                    "@type": "xsd:anyURI"
                },
                "ids:endpointURI": {
                    "@value": "https://consumer.org/pip/ep/
    geo",
                    "@type": "xsd:anyURI"
                }
            }
        }
```

```
    ]
}
```

Listing 2: Specifying the Consumer; In IDS, it is possible to specify constraints on the allowed consumer using participantRefinement class. This constraint reuires the consumer to be located in EU. It uses a PIP to inquire the geolocation which is available at endpoint ep/geo with a defined interface accessible at pip/id/geo.

```
{
    "ids:description": [{
            "@value": "Permission to use by SIEM Data App",
            "@type": "http://www.w3.org/2001/XMLSchema#string
    "
        }],
    "ids:target": {"@id": "http://w3id.org/engrd/connector/
    artifact/firewall.log"},
    "ids:action": [{"@id": "idsc:USE"}],
    "ids:constraint": [{
            "@type": "ids:Constraint",
            "ids:leftOperand": { "@id": "idsc:APPLICATION"
            },
            "ids:operator": { "@id": "idsc:EQUALS"
            },
            "ids:rightOperand": {
                "@value": "http://example.com/ids/application
    /siem-app",
                "@type": "xsd:anyURI"
            },
            "ids:pipEndpoint": {
                "@type": "ids:PIP",
                "ids:interfaceDescription": {
                    "@value": "https://example.com/ids/pip/id
    /application",
                    "@type": "xsd:anyURI"
                },
                "ids:endpointURI": {
                    "@value": "https://consumer.org/pip/ep/
    application",
                    "@type": "xsd:anyURI"
}}}]},
```

Listing 3: The permission to use the firewall logs by the SIEM data app. The assignee, which will be added after contract negotiation, will be granted the permission to use data by the SIEM IDS App.

Firewall Logs

We downloaded some firewall log samples from the documentation of FortiGate, an advanced firewall application [2]. One of the samples is shown in Listing 4. The logs contain network traffic data, security events detected by the built-in intrusion detection and prevention system, and changes in the configurations. These logs contain device identifiers which can indirectly identify a person, such as International Mobile Equipment Identity (IMEI), medium access control (MAC) address, and so on. Therefore, handling of this data should be GDPR-compliant.

```
1 date=2019-05-10 time=11:37:47 logid="0000000013" type="
    traffic" subtype="forward" level="notice" vd="vdom1"
    eventtime=1557513467369913239 srcip=10.1.100.11 srcport=58
    012 srcintf="port12" srcintfrole="undefined" dstip=23.59.1
    54.35 dstport=80 dstintf="port11" dstintfrole="undefined"
    srcuuid="ae28f494-5735-51e9-f247-d1d2ce663f4b" dstuuid="
    ae28f494-5735-51e9-f247-d1d2ce663f4b" poluuid="ccb269e0
    -5735-51e9-a218-a397dd08b7eb" sessionid=105048 proto=6
    action="close" policyid=1 policytype="policy" service="
    HTTP" dstcountry="Canada" srccountry="Reserved" trandisp="
    snat" transip=172.16.200.2 transport=58012 appid=34050 app
    ="HTTP.BROWSER_Firefox" appcat="Web.Client" apprisk="
    elevated" applist="g-default" duration=116 sentbyte=1188
    rcvdbyte=1224 sentpkt=17 rcvdpkt=16 utmaction="allow"
    countapp=1 osname="Ubuntu" mastersrcmac="a2:e9:00:ec:40:01
    " srcmac="a2:e9:00:ec:40:01" srcserver=0 utmref=65500-742
```

Listing 4: Firewall Log Sample Record. It is a network traffic log containing source and destination addresses and MAC addresses which might be considered as sensitive information in some use cases.

[2] https://docs.fortinet.com/document/fortigate/7.4.0/administration-guide/986892/sample-logs-by-log-type

Bibliography

[1] *2023 Cyber Security Statistics Trends & Data*. PurpleSec. URL: https://purplesec.us/resources/cyber-security-statistics/ (visited on 06/10/2024).

[2] Sean Barnum. "STIX - Standardizing cyber threat intelligence information with the structured threat information expression". In: *Mitre Corporation* 11 (2012), pp. 1–22.

[3] M. Blaze, J. Feigenbaum, and J. Lacy. "Decentralized trust management". In: *Proceedings 1996 IEEE Symposium on Security and Privacy*. Proceedings 1996 IEEE Symposium on Security and Privacy. ISSN: 1081-6011. May 1996, pp. 164–173. DOI: 10.1109/SECPRI.1996.502679. URL: https://ieeexplore.ieee.org/document/502679/?arnumber=502679 (visited on 07/09/2024).

[4] Marco Casassa-Mont et al. "Towards safer information sharing in the cloud". In: *International Journal of Information Security* 14.4 (Aug. 1, 2015), pp. 319–334. ISSN: 1615-5270. DOI: 10.1007/s10207-014-0258-5. URL: https://doi.org/10.1007/s10207-014-0258-5 (visited on 01/15/2024).

[5] David W Chadwick et al. "A cloud-edge based data security architecture for sharing and analysing cyber threat information". In: *Future Generation Computer Systems* 102 (Jan. 1, 2020), pp. 710–722. ISSN: 0167-739X. DOI: 10.1016/j.future.2019.06.026. URL: https://www.sciencedirect.com/science/article/pii/S0167739X19300895 (visited on 01/04/2024).

[6] Julie Connolly, Mark Davidson, and Charles Schmidt. "The trusted automated exchange of indicator information (taxii)". In: *The MITRE Corporation* (2014), pp. 1–20.

[7] Luigi Coppolino et al. "Building Cyber-Resilient Smart Grids with Digital Twins and Data Spaces". In: *Applied Sciences* 13.24 (Jan. 2023). Number: 24 Publisher: Multidisciplinary Digital Publishing Institute, p. 13060. ISSN: 2076-3417. DOI: 10.3390/app132413060. URL: https://www.mdpi.com/2076-3417/13/24/13060 (visited on 03/16/2024).

[8] Edward Curry et al. "A Real-time Linked Dataspace for the Internet of Things: Enabling "Pay-As-You-Go" Data Management in Smart Environments". In: *Future Generation Computer Systems* 90 (Jan. 2019), pp. 405–422. ISSN: 0167-739X. DOI: 10.1016/j.future.2018.07.019. URL: https://www.sciencedirect.com/science/article/pii/S0167739X1732887X (visited on 11/26/2023).

[9] *Cyber Threat Intelligence Technical Committee*. URL: https://oasis-open.github.io/cti-documentation/ (visited on 06/15/2024).

[10] Tobias Dam et al. *A Survey of Dataspace Connector Implementations*. Sept. 20, 2023. arXiv: 2309.11282[cs]. URL: http://arxiv.org/abs/2309.11282 (visited on 10/01/2023).

[11] Luc Dandurand and Oscar Serrano Serrano. "Towards improved cyber security information sharing". In: *2013 5th International Conference on Cyber Conflict (CYCON 2013)*. 2013 5th International Conference on Cyber Conflict (CYCON 2013). ISSN: 2325-5374. June 2013, pp. 1–16. URL: https://ieeexplore.ieee.org/abstract/document/6568369 (visited on 06/26/2024).

[12] *Data Spaces Radar*. International Data Spaces. URL: https://internationaldataspaces.org/adopt/data-spaces-radar/ (visited on 06/20/2024).

[13] José M. De Fuentes et al. "PRACIS: Privacy-preserving and aggregatable cybersecurity information sharing". In: *Computers & Security* 69 (Aug. 2017), pp. 127–141. ISSN: 01674048. DOI: 10.1016/j.cose.2016.12.011. URL: https://linkinghub.elsevier.com/retrieve/pii/S0167404816301821 (visited on 12/27/2023).

[14] Ruilong Deng, Peng Zhuang, and Hao Liang. "False Data Injection Attacks Against State Estimation in Power Distribution Systems". In: *IEEE Transactions on Smart Grid* 10.3 (May 2019), pp. 2871–2881. ISSN: 1949-3053, 1949-3061. DOI: 10.1109/TSG.2018.2813280. URL: https://ieeexplore.ieee.org/document/8307441/ (visited on 10/29/2023).

[15] *Detect, SHARE, Protect - Solutions for Improving Threat Data Exchange among CERTs*. ENISA. 2013. URL: https://www.enisa.europa.eu/publications/detect-share-protect-solutions-for-improving-threat-data-exchange-among-certs (visited on 06/24/2024).

[16] Andreas Eitel et al. *Usage Control in the International Data Spaces*. en. Tech. rep. Version Number: 3.0. Zenodo, Mar. 2021. DOI: 10.5281/ZENODO.5675884. URL: https://zenodo.org/record/5675884 (visited on 10/03/2023).

[17] *eXtensible Access Control Markup Language (XACML) Version 3.0*. Jan. 22, 2013. URL: http://docs.oasis-open.org/xacml/3.0/xacml-3.0-core-spec-os-en.html (visited on 05/11/2024).

[18] Michael Franklin, Alon Halevy, and David Maier. "From databases to dataspaces: a new abstraction for information management". en. In: *ACM SIGMOD Record* 34.4 (Dec. 2005), pp. 27–33. ISSN: 0163-5808. DOI: 10.1145/1107499.1107502. URL: https://dl.acm.org/doi/10.1145/1107499.1107502 (visited on 06/05/2023).

[19] Di Freeze. *Global Cybersecurity Spending To Exceed $1.75 Trillion From 2021-2025*. Cybercrime Magazine. Section: Blogs. Sept. 10, 2021. URL: https://cybersecurityventures.com/cybersecurity-spending-2021-2025/ (visited on 06/10/2024).

[20] Julio Hernandez, Lucy McKenna, and Rob Brennan. "TIKD: A Trusted Integrated Knowledge Dataspace For Sensitive Healthcare Data Sharing". en. In: *2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC)*. Madrid, Spain: IEEE, July 2021, pp. 1855–1860. ISBN: 978-1-66542-463-9. DOI: 10.1109/COMPSAC51774.2021.00280. URL: https://ieeexplore.ieee.org/document/9529379/ (visited on 10/30/2023).

[21] J. L. Hernandez-Ardieta, J. Tapiador, and Guillermo Suarez-Tangil. "Information sharing models for cooperative cyber defence". In: *2013 5th International Conference on Cyber Conflict (CYCON 2013)* (June 4, 2013). URL: https://www.semanticscholar.org/paper/Information-sharing-models-for-cooperative-cyber-Hernandez-Ardieta-Tapiador/f20e54ae9b67dc072d680c96acd5016cb8b00c07 (visited on 06/25/2024).

[22] Alan R Hevner. "A Three Cycle View of Design Science Research". In: 19 (2007).

[23] Daire Homan, Ian Shiel, and Christina Thorpe. "A New Network Model for Cyber Threat Intelligence Sharing using Blockchain Technology". In: *2019 10th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*. 2019 10th IFIP International Conference on New Technologies, Mobility and Security (NTMS). ISSN: 2157-4960. June 2019, pp. 1–6. DOI: 10.1109/NTMS.2019.8763853. URL: https://ieeexplore.ieee.org/document/8763853?denied= (visited on 01/14/2024).

[24] *ISO/IEC 27010 Inter-sector comms*. URL: https://www.iso27001security.com/html/27010.html (visited on 06/26/2024).

[25] Vitor Jesus, Balraj Bains, and Victor Chang. "Sharing Is Caring: Hurdles and Prospects of Open, Crowd-Sourced Cyber Threat Intelligence". In: *IEEE Transactions on Engineering Management* (2023), pp. 1–20. ISSN: 0018-9391, 1558-0040. DOI: 10.1109/TEM.2023.3279274. URL: https://ieeexplore.ieee.org/document/10146036/ (visited on 07/14/2023).

[26] Christopher S. Johnson et al. *Guide to Cyber Threat Information Sharing*. NIST SP 800-150. National Institute of Standards and Technology, Oct. 2016, NIST SP 800–150. DOI: 10.6028/NIST.SP.800-150. URL: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-150.pdf (visited on 12/01/2023).

[27] Panos Kampanakis and Mio Suzuki. *Incident Object Description Exchange Format Usage Guidance*. Request for Comments RFC 8274. Num Pages: 33. Internet Engineering Task Force, Nov. 2017. DOI: 10.17487/RFC8274. URL: https://datatracker.ietf.org/doc/rfc8274 (visited on 06/15/2024).

[28] Andreas Kembuegler. *The Smart Connected Supplier Network by TNO*. International Data Spaces. Apr. 23, 2020. URL: https://internationaldataspaces.org/the-smart-connected-supplier-network-by-tno/ (visited on 04/08/2024).

[29] Masoudeh Keshavarzi and Hamid Reza Ghaffary. "I2CE3: A dedicated and separated attack chain for ransomware offenses as the most infamous cyber extortion". In: *Computer Science Review* 36 (May 1, 2020), p. 100233. ISSN: 1574-0137. DOI: 10.1016/j.cosrev.2020.100233. URL: https://www.sciencedirect.com/science/article/pii/S1574013719300838 (visited on 10/29/2023).

[30] Ahmed El-Kosairy, Nashwa Abdelbaki, and Heba Aslan. "A survey on cyber threat intelligence sharing based on Blockchain". In: *Advances in Computational Intelligence* 3.3 (May 23, 2023), p. 10. ISSN: 2730-7808. DOI: 10.1007/s43674-023-00057-z. URL: https://doi.org/10.1007/s43674-023-00057-z (visited on 12/27/2023).

[31] Philippe Kruchten. "Architectural Blueprints—The "4+1" View Model of Software Architecture". In: ().

[32] Aliaksandr Lazouski, Fabio Martinelli, and Paolo Mori. "Usage control in computer security: A survey". In: *Computer Science Review* 4.2 (May 1, 2010), pp. 81–99. ISSN: 1574-0137. DOI: 10.1016/j.cosrev.2010.02.002. URL: https://www.sciencedirect.com/science/article/pii/S1574013710000146 (visited on 01/20/2024).

[33] Martin Lee. *Cyber threat intelligence*. Oxford, UK ; Hoboken, NJ, USA: Wiley, 2023. 1 p. ISBN: 978-1-119-86176-8 978-1-119-86175-1.

[34] Rafał Leszczyna. *Cybersecurity in the electricity sector: managing critical infrastructure*. Cham: Springer, 2019. 213 pp. ISBN: 978-3-030-19538-0 978-3-030-19537-3.

[35] Terry MacDonald, Paul McKitrick, and Merike Kaeo. *IEP 2.0 Framework Definition*. FIRST — Forum of Incident Response and Security Teams. Nov. 6, 2019. URL: https://www.first.org/iep/iep_framework_2_0 (visited on 07/27/2024).

[36] *MAEC - Malware Attribute Enumeration and Characterization | MAEC Project Documentation*. URL: https://maecproject.github.io/ (visited on 06/15/2024).

[37] Fabio Martinelli et al. "A Formal Support for Collaborative Data Sharing". In: *Multidisciplinary Research and Practice for Information Systems*. Ed. by Gerald Quirchmayr et al. Berlin, Heidelberg: Springer, 2012, pp. 547–561. ISBN: 978-3-642-32498-7. DOI: 10.1007/978-3-642-32498-7_42.

[38] *melicertes/csp*. original-date: 2019-05-14T13:59:38Z. Apr. 30, 2024. URL: https://github.com/melicertes/csp (visited on 06/19/2024).

[39] National Institute of Standards and Technology. *NIST - Framework for Improving Critical Infrastructure Cybersecurity, Version 1.1*. NIST CSWP 04162018. Gaithersburg, MD: National Institute of Standards and Technology, Apr. 16, 2018, NIST CSWP 04162018. DOI: 10.6028/NIST.CSWP.04162018. URL: http://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.04162018.pdf (visited on 05/10/2023).

[40] *ODRL Information Model 2.2*. URL: https://www.w3.org/TR/odrl-model/ (visited on 07/09/2024).

[41] B. Otto et al. *IDS Reference Architecture Model*. en. Tech. rep. Version Number: Version 3.0. Zenodo, Apr. 2019. DOI: 10.5281/ZENODO.5105529. URL: https://zenodo.org/record/5105529 (visited on 09/24/2023).

[42] Boris Otto. "The Evolution of Data Spaces". In: *Designing Data Spaces : The Ecosystem Approach to Competitive Advantage*. Ed. by Boris Otto, Michael ten Hompel, and Stefan Wrobel. Cham: Springer International Publishing, 2022, pp. 3–15. ISBN: 978-3-030-93975-5. DOI: 10.1007/978-3-030-93975-5_1. URL: https://doi.org/10.1007/978-3-030-93975-5_1.

[43] Prof. Dr. Boris Otto. *GAIA-X and IDS*. en. Tech. rep. Version Number: 1.0. Zenodo, Jan. 2021. DOI: 10.5281/ZENODO.5675897. URL: https://zenodo.org/record/5675897 (visited on 05/10/2023).

[44] Maryam Pahlevan, Artemis Voulkidis, and Terpsichori-Helen Velivassaki. "Secure exchange of cyber threat intelligence using TAXII and distributed ledger technologies - application for electrical power and energy system". In: *Proceedings of the 16th International Conference on Availability, Reliability and Security*. ARES '21. New York, NY, USA: Association for Computing Machinery, Aug. 2021, pp. 1–8. ISBN: 978-1-4503-9051-4. DOI: 10.1145/3465481.3470476. URL: https://doi.org/10.1145/3465481.3470476 (visited on 10/15/2023).

[45] Alan Paice and Sean McKeown. "Practical Cyber Threat Intelligence in the UK Energy Sector". en. In: (Mar. 2023). Publisher: Springer. DOI: 10.1007/978-981-19-6414-5_1. URL: https://napier-repository.worktribe.com/output/2880543/practical-cyber-threat-intelligence-in-the-uk-energy-sector (visited on 10/30/2023).

[46] Jaehong Park and Ravi Sandhu. "Towards usage control models: beyond traditional access control". In: *Proceedings of the seventh ACM symposium on Access control models and technologies*. SACMAT '02. New York, NY, USA: Association for Computing Machinery, June 3, 2002, pp. 57–64. ISBN: 978-1-58113-496-4. DOI: 10.1145/507711.507722. URL: https://dl.acm.org/doi/10.1145/507711.507722 (visited on 01/04/2024).

[47] Siani Pearson and Marco Casassa-Mont. "Sticky Policies: An Approach for Managing Privacy across Multiple Parties". In: *Computer* 44.9 (Sept. 2011). Conference Name: Computer, pp. 60–68. ISSN: 1558-0814. DOI: 10.1109/MC.2011.225. URL: https://ieeexplore.ieee.org/abstract/document/5959137?casa_token=lNtkpc_ySW0AAAAA:fnGYz_AqpWfCpIZ6fCfoIhUvzIXcCrKsV6Z1Hv2BUeoxgbI47Bi-4wkOMXkE476xy28AjDM8 (visited on 05/12/2024).

[48] Heinrich Pettenpohl, Markus Spiekermann, and Jan Ruben Both. "International Data Spaces in a Nutshell". In: *Designing Data Spaces : The Ecosystem Approach to Competitive Advantage*. Ed. by Boris Otto, Michael ten Hompel, and Stefan Wrobel. Cham: Springer International Publishing, 2022, pp. 29–40. ISBN: 978-3-030-93975-5. DOI: 10.1007/978-3-030-93975-5_3. URL: https://doi.org/10.1007/978-3-030-93975-5_3.

[49] *PHOENIX*. Electrical Power System's Shield against complex incidents and extensive cyber and privacy attacks. URL: https://phoenix-h2020.eu/ (visited on 06/19/2024).

[50] *Proactive detection of network security incidents, report.* ENISA. URL: https://www.enisa.europa.eu/publications/proactive-detection-report (visited on 12/26/2023).

[51] Reiberg, Abel and Niebel, and Crispin. *What is a Data Space?* 2022.

[52] Matteo Repetto. "Adaptive monitoring, detection, and response for agile digital service chains". In: *Computers & Security* 132 (Sept. 1, 2023), p. 103343. ISSN: 0167-4048. DOI: 10.1016/j.cose.2023.103343. URL: https://www.sciencedirect.com/science/article/pii/S0167404823002535 (visited on 10/15/2023).

[53] Manuel Rudolph. "User-friendly and Tailored Policy Administration Points". In: Jan. 1, 2015.

[54] Clemens Sauerwein et al. "Threat Intelligence Sharing Platforms: An Exploratory Study of Software Vendors and Research Perspectives". In: *Wirtschaftsinformatik 2017 Proceedings* (Jan. 23, 2017). URL: https://aisel.aisnet.org/wi2017/track08/paper/3.

[55] Daniel Schlette, Marco Caselli, and Günther Pernul. "A Comparative Study on Cyber Threat Intelligence: The Security Incident Response Perspective". In: *IEEE Communications Surveys & Tutorials* 23.4 (2021). Conference Name: IEEE Communications Surveys & Tutorials, pp. 2525–2556. ISSN: 1553-877X. DOI: 10.1109/COMST.2021.3117338. URL: https://ieeexplore.ieee.org/abstract/document/9557787 (visited on 06/25/2024).

[56] Julian Schuette and Gerd Stefan Brost. "LUCON: Data Flow Control for Message-Based IoT Systems". In: *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE).* 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE). ISSN: 2324-9013. Aug. 2018, pp. 289–299. DOI: 10.1109/TrustCom/BigDataSE.2018.00052. URL: https://ieeexplore.ieee.org/document/8455920 (visited on 07/13/2024).

[57] Oscar Serrano, Luc Dandurand, and Sarah Brown. "On the Design of a Cyber Security Data Sharing System". In: *Proceedings of the 2014 ACM Workshop on Information Sharing & Collaborative Security.* WISCS '14. New York, NY, USA: Association for Computing Machinery, Nov. 3, 2014, pp. 61–69. ISBN: 978-1-4503-3151-7. DOI: 10.1145/2663876.2663882. URL: https://doi.org/10.1145/2663876.2663882 (visited on 07/27/2024).

[58] Dimitrios Skias et al. "Pan-European Cybersecurity Incidents Information Sharing Platform to support NIS Directive". In: *Proceedings of the 16th International Conference on Availability, Reliability and Security.* ARES '21. New York, NY, USA: Association for Computing Machinery, Aug. 17, 2021, pp. 1–7. ISBN: 978-1-4503-9051-4. DOI: 10.1145/3465481.3470477. URL: https://dl.acm.org/doi/10.1145/3465481.3470477 (visited on 01/21/2024).

[59] Florian Skopik, Giuseppe Settanni, and Roman Fiedler. "A problem shared is a problem halved: A survey on the dimensions of collective cyber defense through security information sharing". In: *Computers & Security* 60 (July 2016), pp. 154–176. ISSN: 01674048. DOI: 10.1016/j.cose.2016.04.003. URL: https://linkinghub.elsevier.com/retrieve/pii/S0167404816300347 (visited on 12/04/2023).

[60] Florian Skopik, Giuseppe Settanni, and Roman Fiedler. "A problem shared is a problem halved: A survey on the dimensions of collective cyber defense through security information sharing". In: *Computers & Security* 60 (July 2016), pp. 154–176. ISSN: 01674048. DOI: 10.1016/j.cose.2016.04.003. URL: https://linkinghub.elsevier.com/retrieve/pii/S0167404816300347 (visited on 12/04/2023).

[61] Sebastian Steinbuss et al. *IDS Certification explained*. Zenodo, Nov. 1, 2019. DOI: 10.5281/zenodo.5675945. URL: https://zenodo.org/records/5675945 (visited on 07/18/2024).

[62] Martin Steinebach et al. "Datenschutz und Datenanalyse: Herausforderungen und Lösungsansätze". In: *Datenschutz und Datensicherheit - DuD* 40 (July 1, 2016), pp. 440–445. DOI: 10.1007/s11623-016-0633-7.

[63] *Strengthening EU-wide cybersecurity and resilience – provisional agreement by the Council and the European Parliament*. URL: https://www.consilium.europa.eu/en/press/press-releases/2022/05/13/renforcer-la-cybersecurite-et-la-resilience-a-l-echelle-de-l-ue-accord-provisoire-du-conseil-et-du-parlement-europeen/ (visited on 06/19/2024).

[64] Hubert Tardieu. "Role of Gaia-X in the European Data Space Ecosystem". en. In: *Designing Data Spaces*. Ed. by Boris Otto, Michael Ten Hompel, and Stefan Wrobel. Cham: Springer International Publishing, 2022, pp. 41–59. ISBN: 978-3-030-93974-8 978-3-030-93975-5. DOI: 10.1007/978-3-030-93975-5_4. URL: https://link.springer.com/10.1007/978-3-030-93975-5_4 (visited on 07/13/2023).

[65] *The VERIS Framework*. URL: https://verisframework.org/ (visited on 06/15/2024).

[66] Thomas D. Wagner et al. "Cyber threat intelligence sharing: Survey and research directions". In: *Computers & Security* 87 (Nov. 1, 2019), p. 101589. ISSN: 0167-4048. DOI: 10.1016/j.cose.2019.101589. URL: https://www.sciencedirect.com/science/article/pii/S016740481830467X (visited on 10/20/2023).

[67] Tania Wallis and Rafał Leszczyna. "EE-ISAC—Practical Cybersecurity Solution for the Energy Sector". en. In: *Energies* 15.6 (Mar. 2022), p. 2170. ISSN: 1996-1073. DOI: 10.3390/en15062170. URL: https://www.mdpi.com/1996-1073/15/6/2170 (visited on 05/10/2023).

[68] Qi Wang et al. "Coordinated Defense of Distributed Denial of Service Attacks against the Multi-Area Load Frequency Control Services". In: *Energies* 12.13 (Jan. 2019). Number: 13 Publisher: Multidisciplinary Digital Publishing Institute, p. 2493. ISSN: 1996-1073. DOI: 10.3390/en12132493. URL: https://www.mdpi.com/1996-1073/12/13/2493 (visited on 10/29/2023).

[69] *What is a Threat Intelligence Platform I Resources I Anomali.* URL: https://www.anomali.com/resources/what-is-a-tip (visited on 06/16/2024).

[70] Adam Zibak and Andrew Simpson. "Cyber Threat Information Sharing: Perceived Benefits and Barriers". In: *Proceedings of the 14th International Conference on Availability, Reliability and Security*. ARES '19. New York, NY, USA: Association for Computing Machinery, Aug. 2019, pp. 1–9. ISBN: 978-1-4503-7164-3. DOI: 10.1145/3339252.3340528. URL: https://dl.acm.org/doi/10.1145/3339252.3340528 (visited on 12/04/2023).

License