Master Thesis
Towards a Dataspace for Cyber Threat Intelligence

Navid Rahimi Danesh

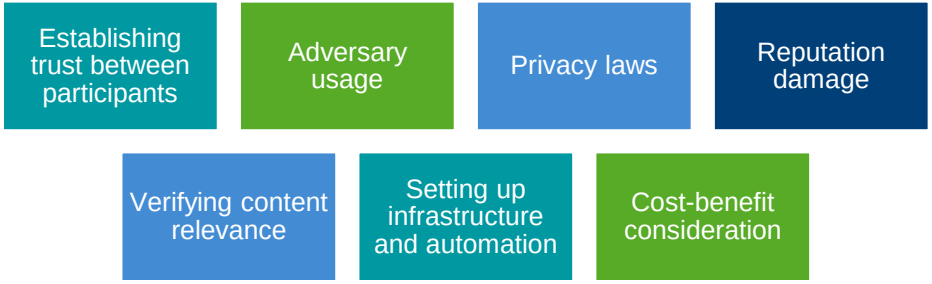
RWTH Aachen, Informatik 5
Lehrstuhl Prof. Decker

Supervisor: Prof. Dr. Stefan Decker, Prof. Dr. Andreas Ulbig
Advisors: Mehdi Akbari Gurabi, Ömer Sen

Introduction - Motivation

Information Sharing in Cyber Security: Motivations and Challenges

- Cyber attacks are evolving
- Cyber Threat Intelligence
  - Proactive: Risk management
  - Reactive: Mitigation and containment
- Collaboration is helpful
  - Reduce duplicate work
  - Faster response
  - Compliance (e.g., NIS2)
- It is open for research due to its challenges [26]

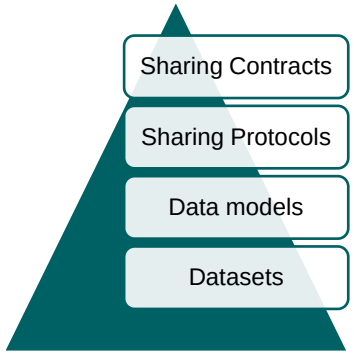


Introduction - Motivation

The Rise of Data Space Technologies

- Data is becoming more valuable
  - Digitalization / AI
- Organizations are sharing more data
  - Data value chains / Ecosystems / Economy
- EU Data strategy
- Notable initiatives
- Dataspaces
  - Help organizations share data
  - Components
  - Are being implemented

Catena-X
Your Automotive Network



Data Space components

iSHARE
gaia-x
BDV BIG DATA VALUE ASSOCIATION
FIWARE
INTERNATIONAL DATA SPACES ASSOCIATION

Introduction - Thesis Goal

Investigating the Suitability of Dataspaces for CTI Sharing Use Case

- Identify gaps in the current CTI sharing platforms
- Address the gaps of current platforms with a dataspace-based solution
- Find implementation considerations when setting up a dataspace for CTI sharing

Methodology

- Identify gaps in current systems
- Potentials of Dataspaces

- Architecture Design
- Data Modelling
- Business Flows

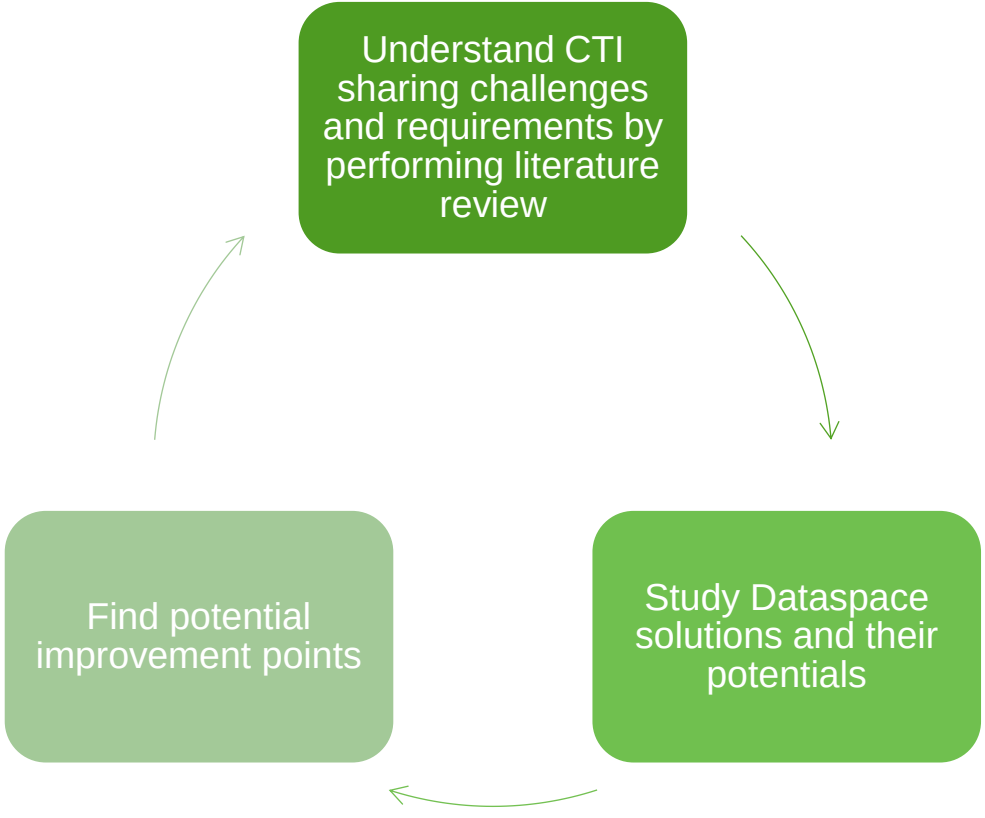
- Tool Selection
- Deployment

- Real-world Scenarios
- Comparative Analysis
- Performance

Use Case and Requirements

Methodology

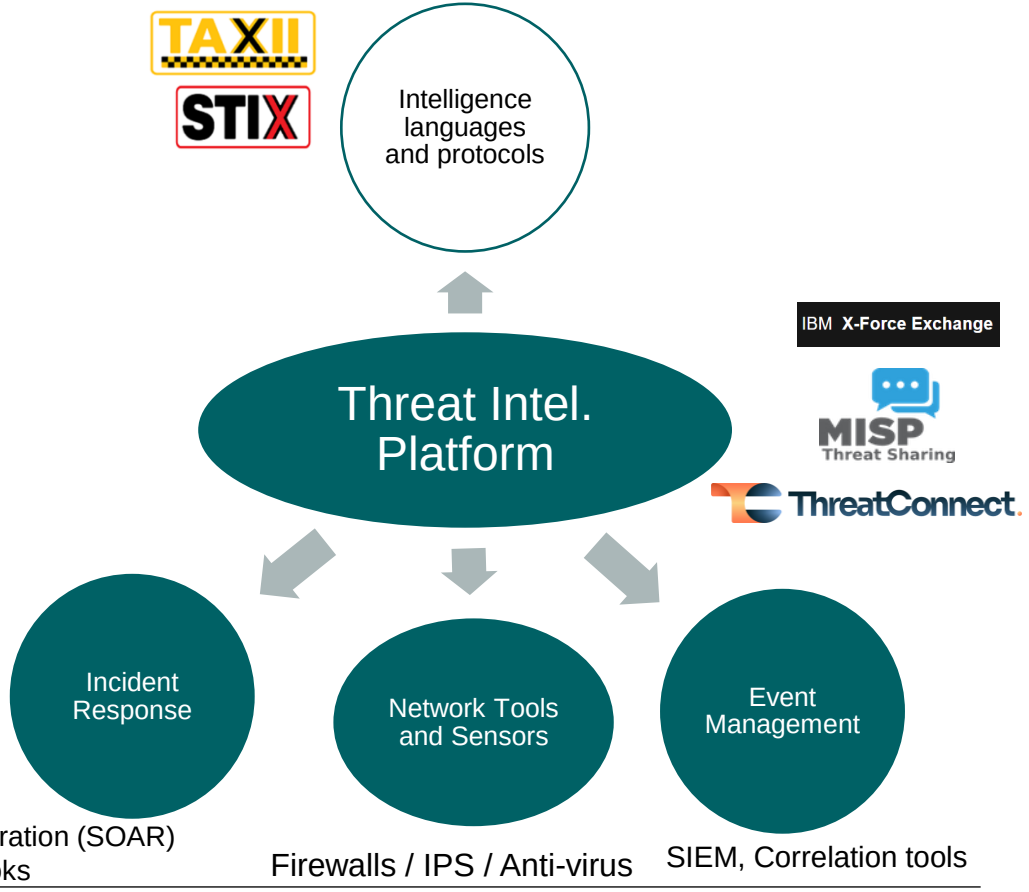
- Goals
  - Identify gaps in current systems
  - Potential benefits of Dataspaces
- Method



Use Case and Requirements

Requirement: Automation and Flexibility

- Role of Automation in CTI
  - Growing complexity and amount of data
  - Human delay can be costly
- Threat Intelligence Platforms (TIPs)
  - Implement automatic collection, integration and sharing of CTI
- We have several TIPs
- CTI Languages and Protocols
  - Allow compatibility between different systems
- TIPs should integrate with many external systems



Use Case and Requirements

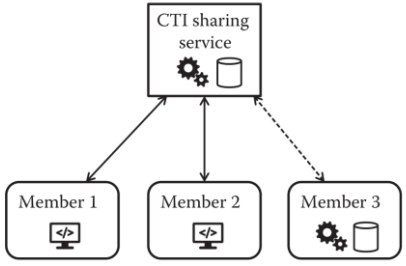
Requirement: Privacy and Sovereignty

- CTI contains sensitive information
  - Data regarding clients: risk of violating GDPR
  - Company secrets: reputation damage
  - Classified information: reports from government
- Approaches
  - Data Sanitization
    - Removing Attribution (Anonymization)
    - PETs
  - Sharing Policies
    - automating the legal aspects of information sharing
    - TLP: Traffic Light Protocol
    - Existing standard policies: IEP / DSA / ISO/IEC

Use Case and Requirements

Problems with existing TIPs

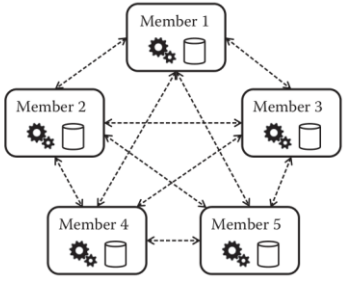
Centralized



- ✅ Verified Content and Participants
- ✅ Automatic sanitization
- ❌ Vendor lock-in
- ❌ One entity controls all exchanges

IBM X-Force Exchange · ThreatConnect · VMRay

Peer to Peer



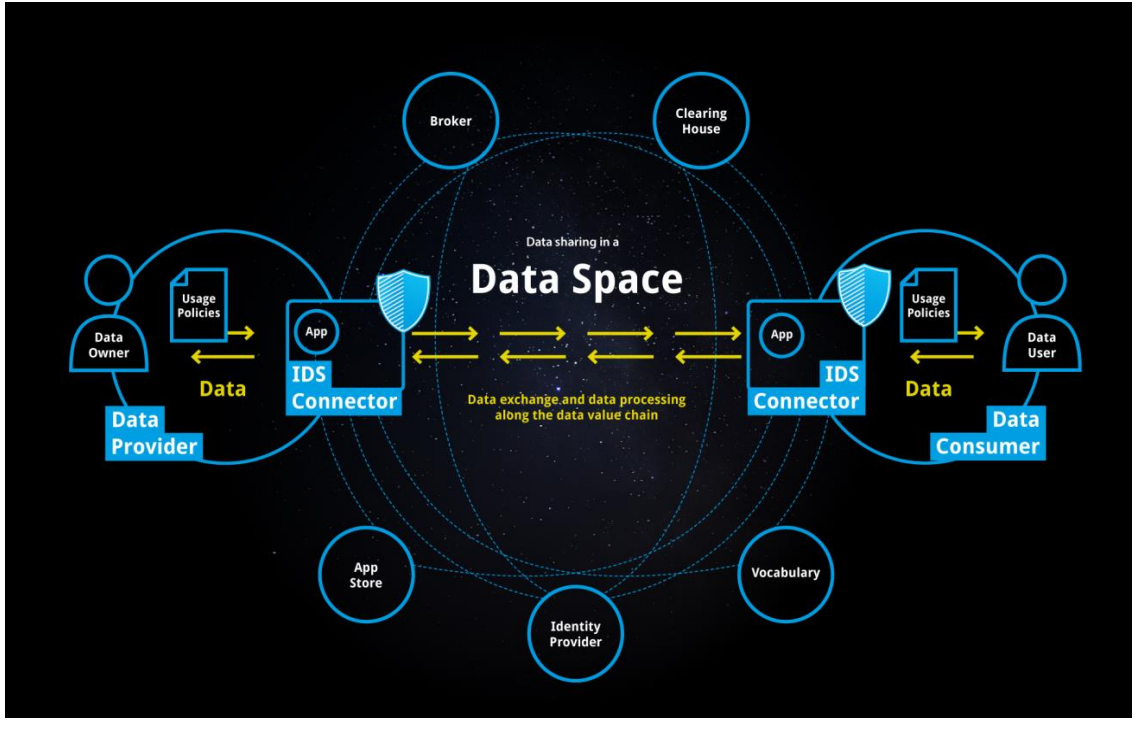
- ❌ Establishment and management of trust
- ❌ Liability risks
- ✅ Open and Interoperable
- 🟡 More privacy and data sovereignty

MISP Threat Sharing · OPENCTI

Use Case and Requirements

International Data Spaces (IDS)

- History
  - 2015: Fraunhofer
  - 2016: non-profit IDSA
- It facilitates
  - Secure and standard data exchange
  - In a trusted business ecosystem
  - Guaranteeing data sovereignty for data owner
- Features
  - Standard Data Exchange Component
  - Usage Policies (Specification and Enforcement)
  - Certification of Participants
  - Certification of Components
  - Extension via Data Apps
  - Clearing and Billing

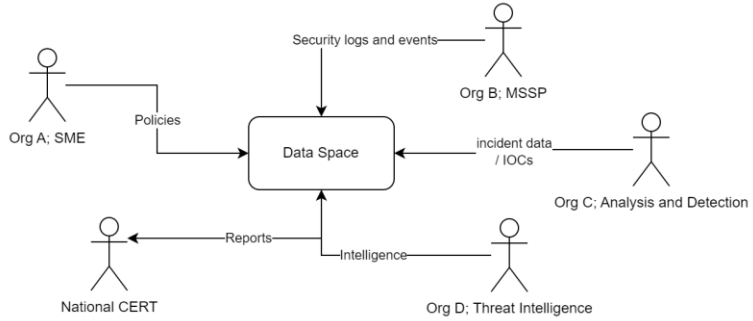


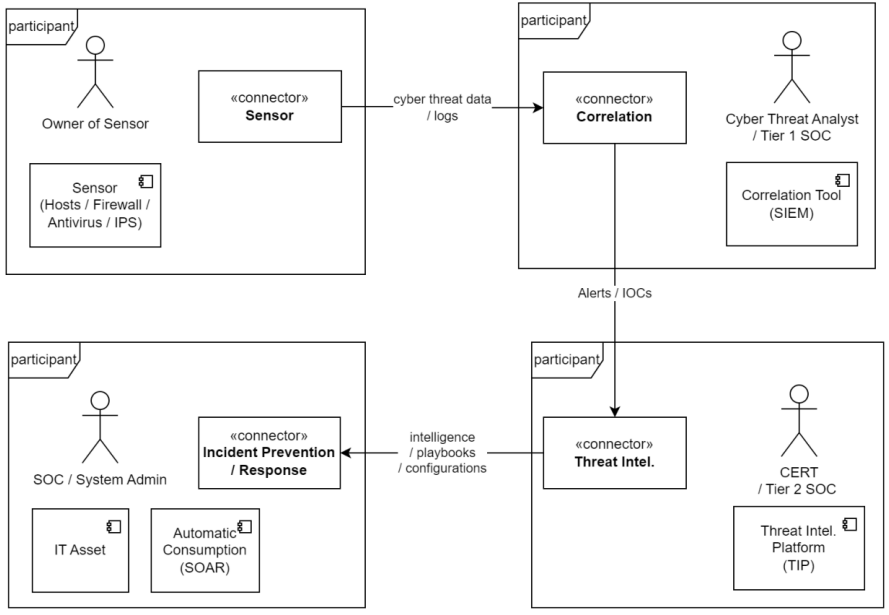
https://internationaldataspaces.org/why/data-spaces/

Use Case and Requirements

Use Case: Critical Infrastructure (e.g., Energy Sector)

- Scenarios:
  1. An SME outsourcing security analysis
     - Ensure purpose / data retention policy
  2. SOC selling incident data to a members of an international community
     - DRM / TLP
  3. National CERT Notifying a Constituent Organization
     - Protect classified information (Distribution Control)
     - Automatic IR
- Actors

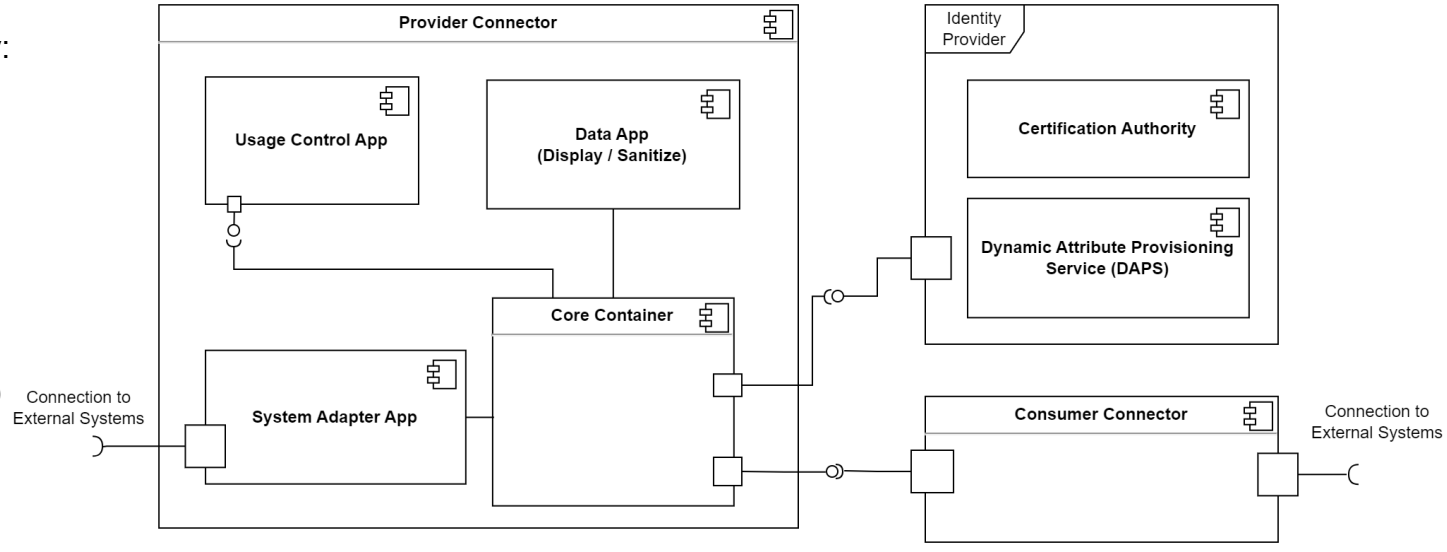




Conceptual Approach

Overview

- Processes
  - Certification Authority:
    - On-boarding and Certification
  - DAPS: Connection Establishment
  - Data Offers
  - Contract Negotiation
    - Core Container
  - Policy Enforcement
    - Usage Control App



Conceptual Approach

Information Model

- IDS policies could ensure
  - Restrict consumer
  - Restrict application
  - Restrict location of use
  - Restrict purpose of use
  - Restrict time interval
  - Restrict number of usage
  - Log usage information
  - Delete data after some time
- CTI vocabularies
  - STIX
  - MISP Taxonomies and Galaxies
  - VERIS
  - IODEF

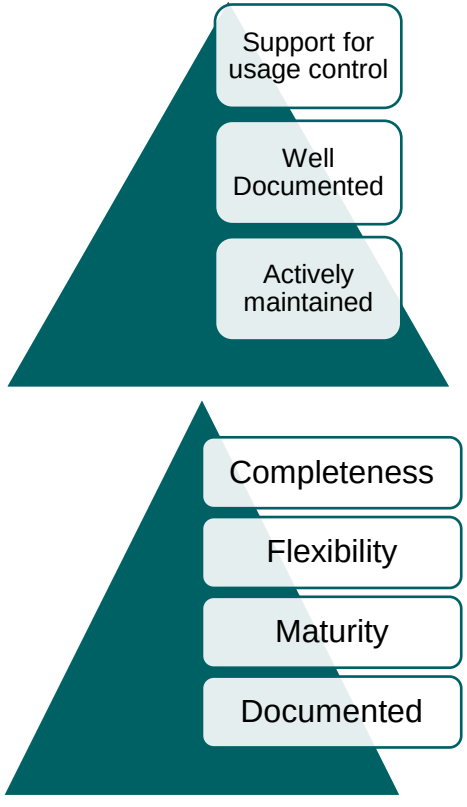
[1] https://www.w3.org/TR/vocab-dcat-3/
[2] https://www.w3.org/TR/odrl-model/

```
{
    "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"
}}}]},
```

Implementation

Technology Selection

- Base Connector
  - Eclipse Dataspace Connector (EDC)
  - ➢ TRUsted Engineering Connector (TRUE)
  - Trusted Connector by Fraunhofer AISEC
  - IDS Dataspace Connector (DSC)
- Policy Engine
  - ➢ MYDATA
  - LUCON
  - Degree (D)
- External System
  - MISP (misp-docker)
- IDS Testbed
  - DAPS: Omejdn
  - Certificate Authority: cfssl
- Source: https://github.com/Navidda/master-thesis



Evaluation

Comparative Analysis

- Interoperability
- Flexibility
- Trust and security
- Commercial
- Data sovereignty and privacy

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/

Evaluation

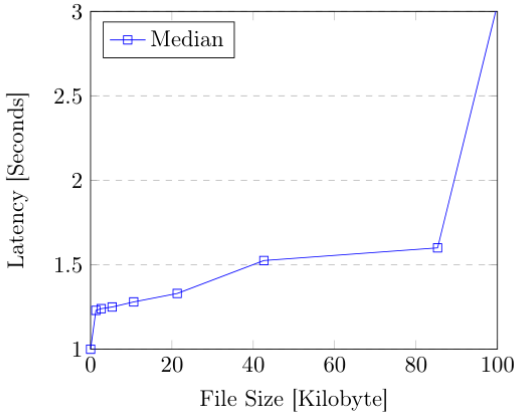
Comparative Analysis of the Policy Framework

- We selected a policy language for CTI data
  - Information Exchange Policy (IEP)
- Which is widely used
  - used by Forum of Incident Response and Security Teams (FIRST)
    - Est. 1990 / 756 members in 111 countries
- We compared its clause classes with our policy framework
  - Does our policy language express it?
    - 7 completely
    - 2 partially
    - 4 not expressed
  - The difficulty of automatically enforcing it with our platform
    - 3 Zero: Our prototype already implements it
    - 4 Low: We need to implement some missing apps
    - 1 Medium: We need to extend the IDS specification
    - 4 High: implementation of complex (currently manual) workflows are required

Evaluation

Prototype Performance Overhead

- Two metrics
- Running machine spec
  - virtual machine
  - Memory: 32Gb
  - Processor: Intel Xeon 8 * (2.1−2.3) Ghz
- Latency: Transferring a sample CTI file from one connector to another
  - < 3s for 100Kb of data
- Memory usage: idle state
  - 1.6GB Per connector



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

Discussion and Conclusion

- Results implications
  - Multiple evaluation results show significant potentials
  - Call for more implementation and investment
- Summary
  - Requirement analysis
  - Design and evaluate
  - Find implementation considerations
- Future works
    - Expert questionnaire about subjective aspects
    - Implement more apps (esp. consumer side)
    - Pilot project and empirical results

References

Selected references for presentation (full list in the thesis)

- [26] Christopher S. Johnson et al. Guide to Cyber Threat Information Sharing. NIST SP 800-150.
- [18] Michael Franklin, Alon Halevy, and David Maier. “From databases to dataspaces: a new abstraction for information management”. en.
- [13] José M. De Fuentes et al. “PRACIS: Privacy-preserving and aggregatable cybersecurity information sharing”.
- [23] Daire Homan, Ian Shiel, and Christina Thorpe. “A New Network Model for Cyber Threat Intelligence Sharing using Blockchain Technology”.
- [58] Dimitrios Skias et al. “Pan-European Cybersecurity Incidents Information Sharing Platform to support NIS Directive”.
- [5] David W Chadwick et al. “A cloud-edge based data security architecture for sharing and analysing cyber threat information”.

Thank you for your attention

Feel free to ask your questions.

Appendices

IEP Policy Classes

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.

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.

Appendices

Monitoring and Detection Feeds and Tools

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

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

Appendices

Incident Response Formats and Tools

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

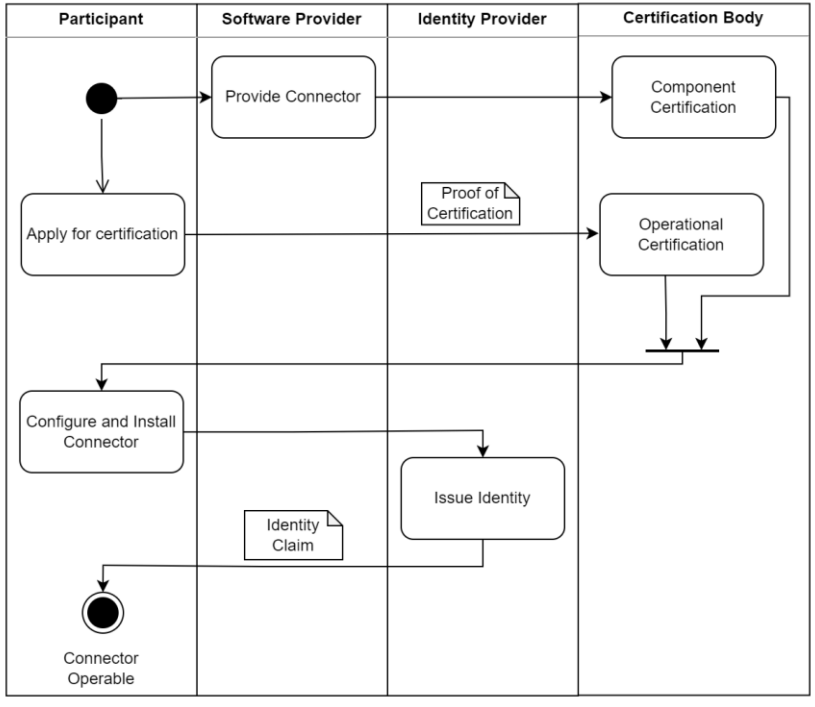
Appendices

Comparison of Open-source IDS Connectors

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

Appendices

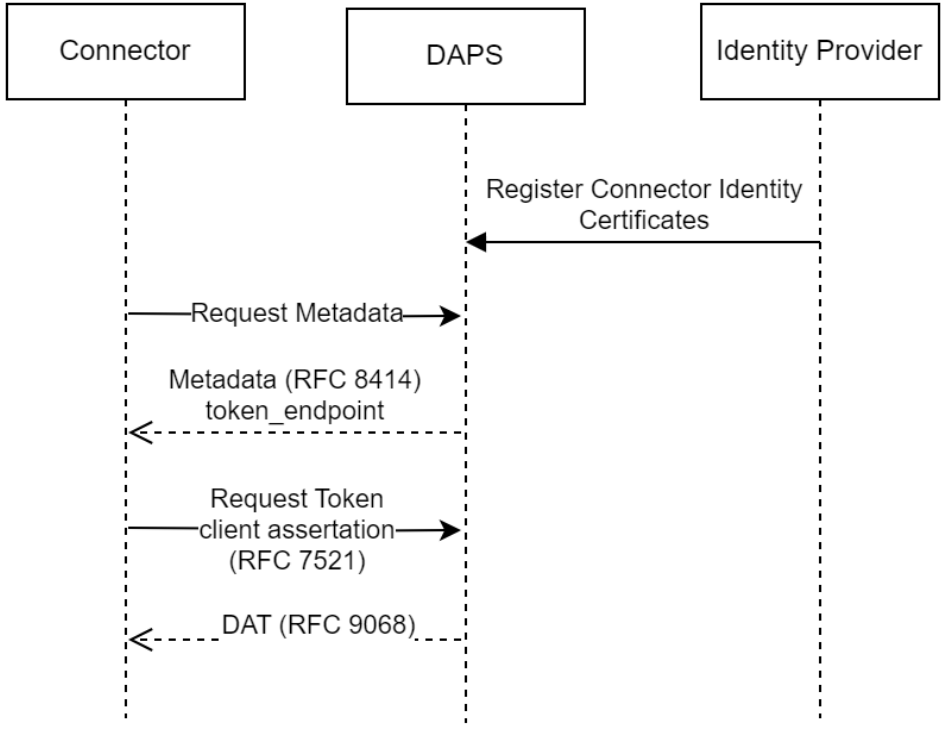
On-boarding and Certification



Appendices

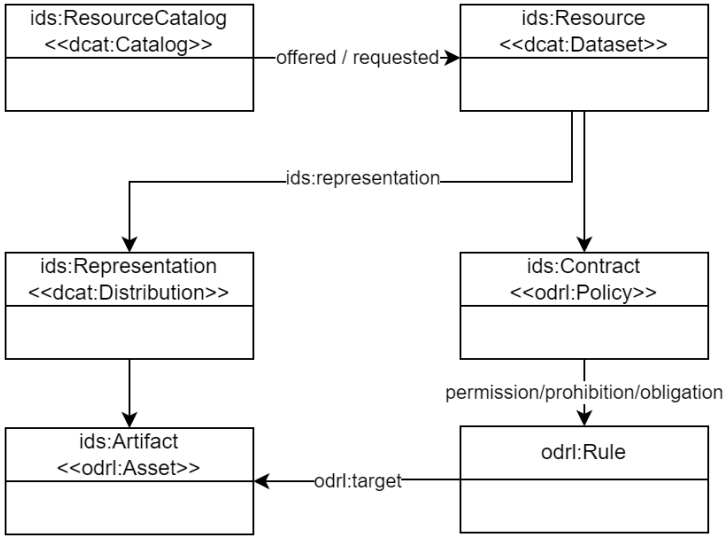
Connection Establishment

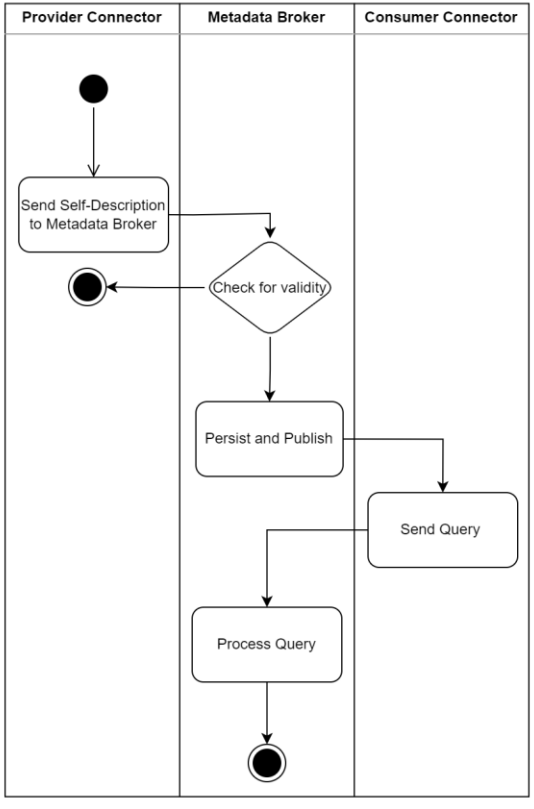
- DAPS Interaction (Fetch DAT)



Appendices

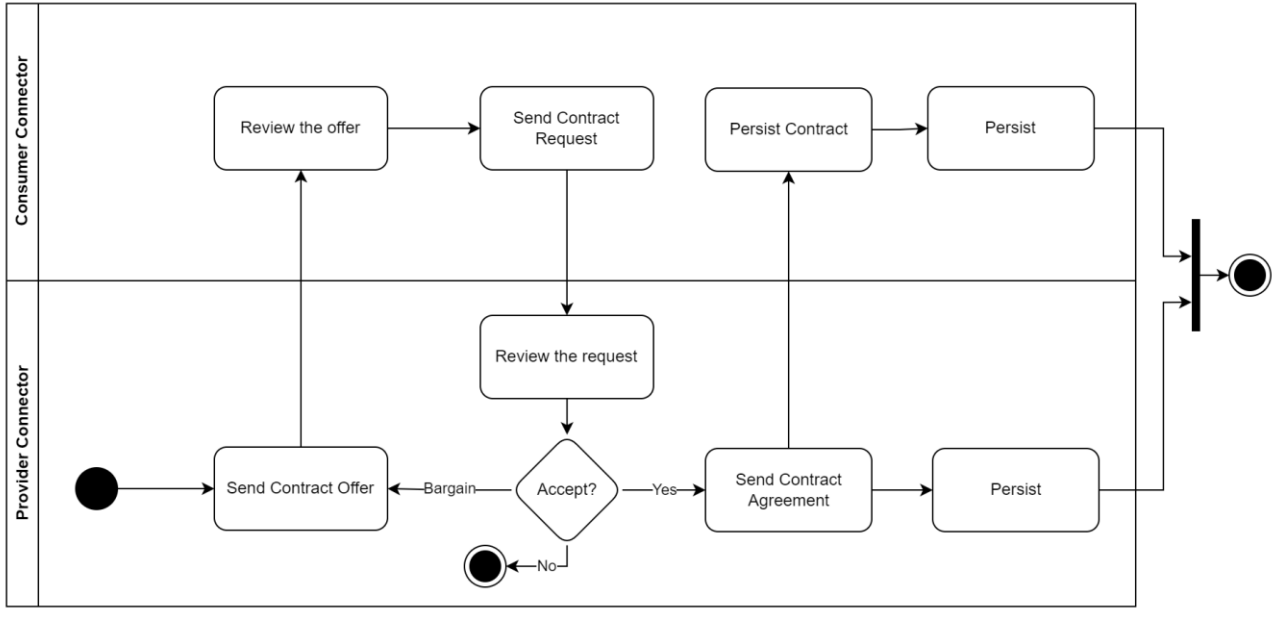
Publishing Data Offers





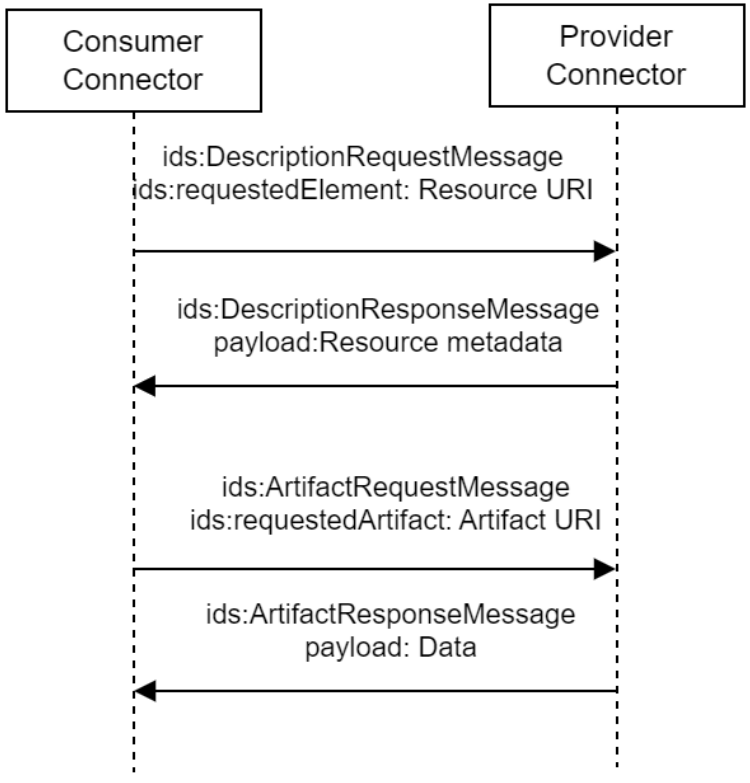
Appendices

Contract Negotiation



Appendices

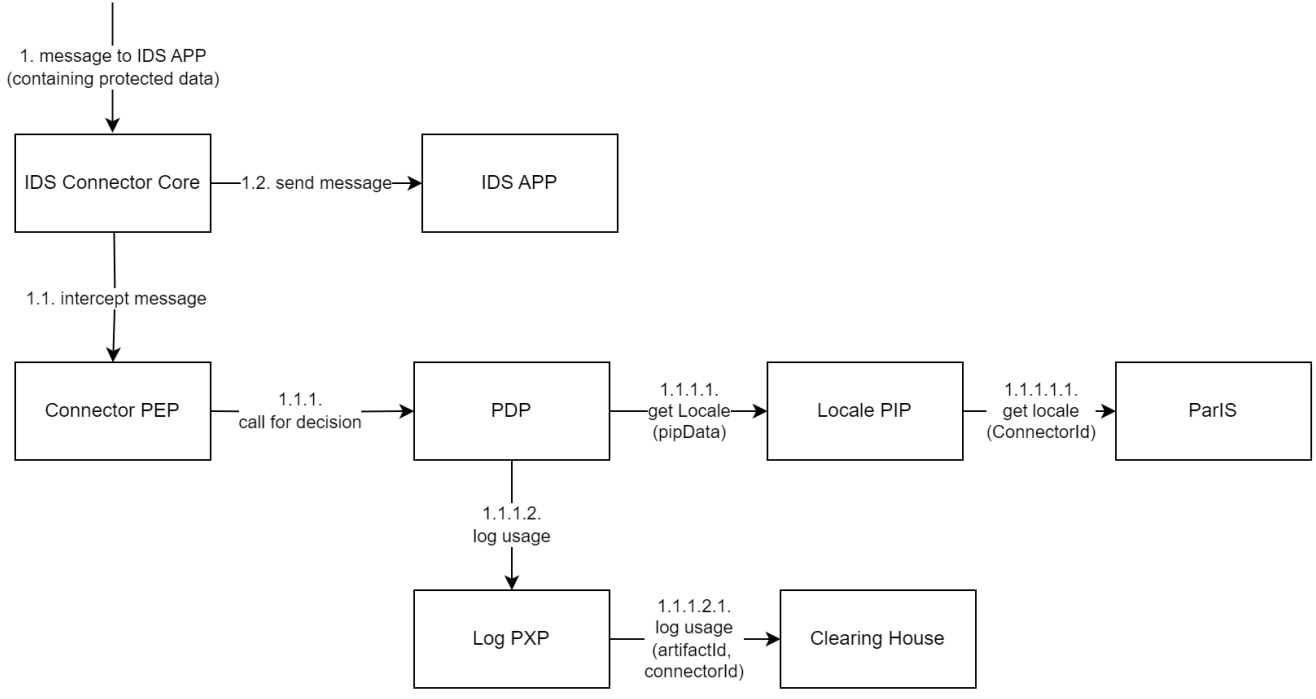
Data Exchange



Appendices

Example of usage control enforcement process

- “Participants must be based in EU and the usage logged to the Clearing House”



Appendices

Research Works Addressing CTI Sharing Challenges

- PRACIS (De Fuentes et al. [13])
- CTI Sharing on Blockchain (Daire Homan et al. [23])
- Incidents Information Sharing Platform (I2SP) (Dimitrios Skias et al [58])
- C3ISP (Chadwick et al. [5])

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.

Appendices

Comparative Analysis of the Policy Framework (Results)

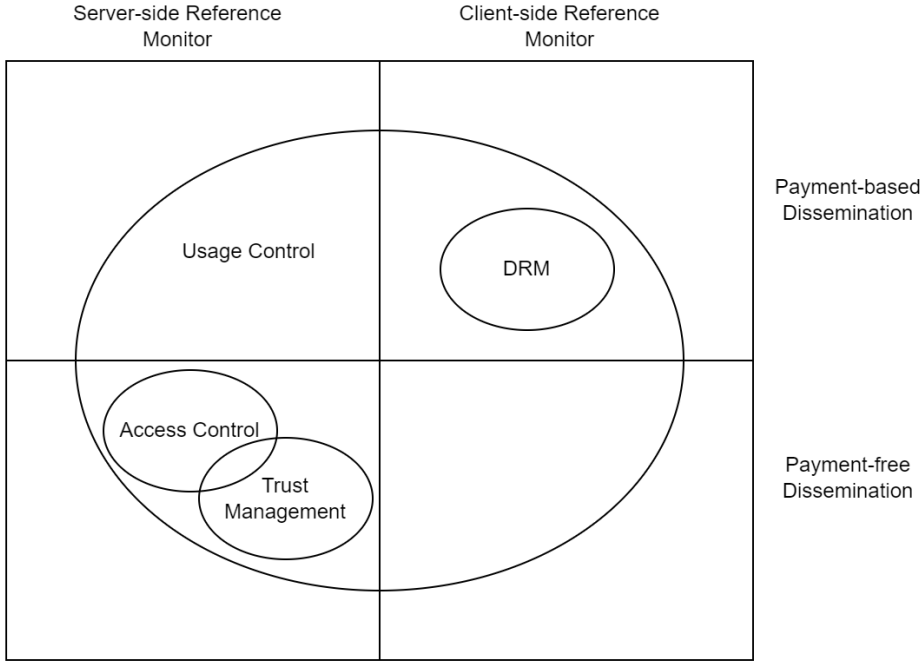
- Expressivity
  - 7 complete
  - 2 partial
  - 4 not expressed
- Enforcement Capability (lower means better)
  - 3 Zero
  - 4 Low
  - 1 Medium
  - 4 High

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

Background

Concepts Related to Usage Control

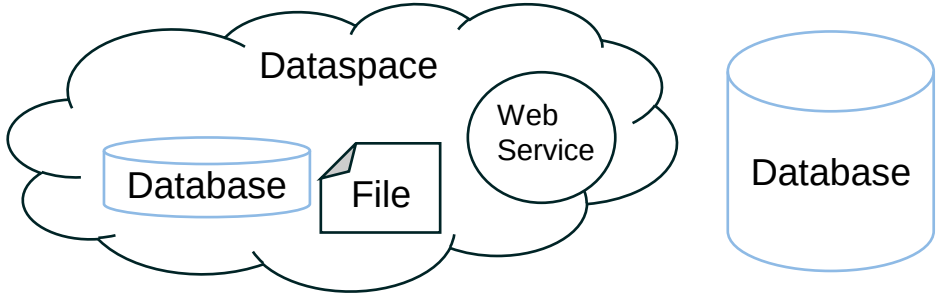
- Access Control: Authorize access
  - e.g., RBAC / XACML
- Trust Management: Authenticate strangers
  - e.g., X.509, PGP
- Digital Rights Management: Prevent illegal distribution
  - e.g., Microsoft PlayReady, Google Widevine
- Usage Control
  - Make decision on each action on data
  - Provision + Obligation

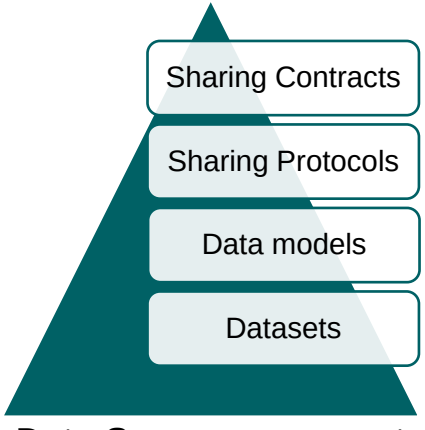


Background

Data Space Concepts

- DataSpace (Franklin 2005) [18]
  - Context: Data management and integration
  - Heterogeneous format, location, or model
- Data Ecosystems
  - Free flow of data
  - Data value chains
- DataSpace in the context of Data Ecosystems
  - Goal: Data sharing between organizations
  - Components





Data Space components

Use Case and Requirements

Intro to Cyber Threat Intelligence

- Contains information about:
- Concerns different actors:
  - Security analysts and SOCs
  - Security researchers
  - Executive management
  - IT department
- Is helpful in:
  - Cyber risk management
  - Incident response
  - Automating related tasks

Conceptual Approach

Roles and Functionalities

- Data Provider
- Data Consumer

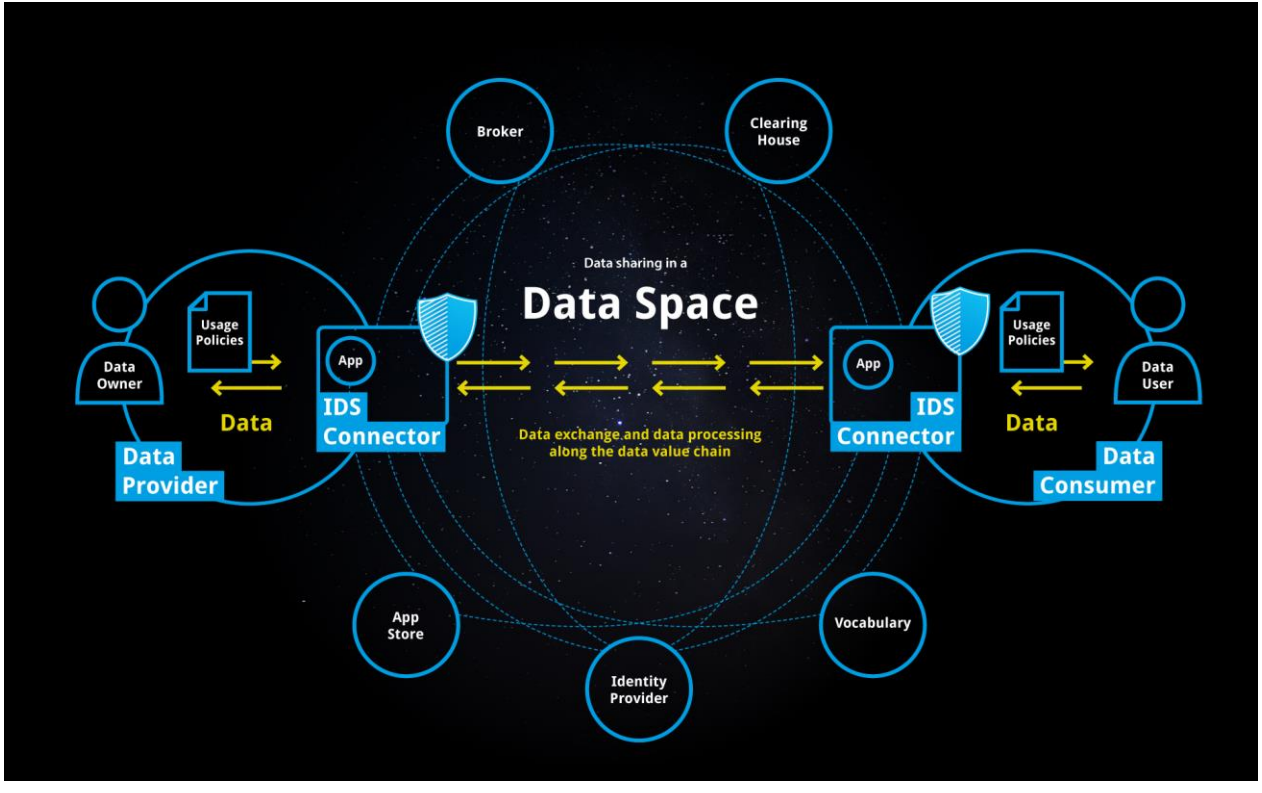
- Metadata Broker
- Clearing House
- App Store
- Identity Provider
- Vocabulary Hub

- Software Developer
- Certification Body

Conceptual Approach

Processes

- On-boarding and Certification
- Connection Establishment
- Publishing Data Offers
- Contract Negotiation
- Data Exchange
- Policy Enforcement

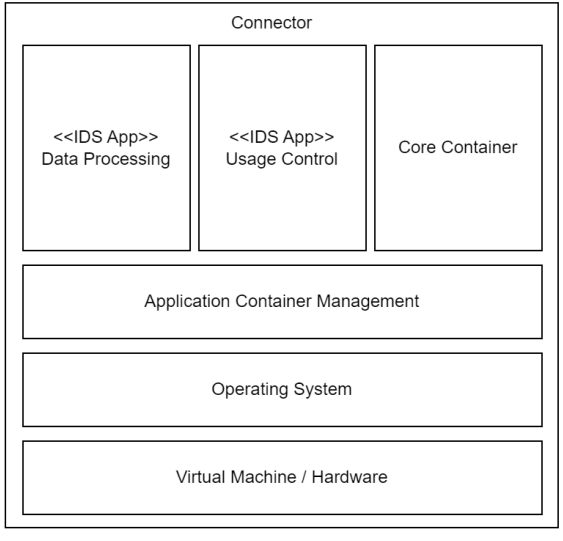


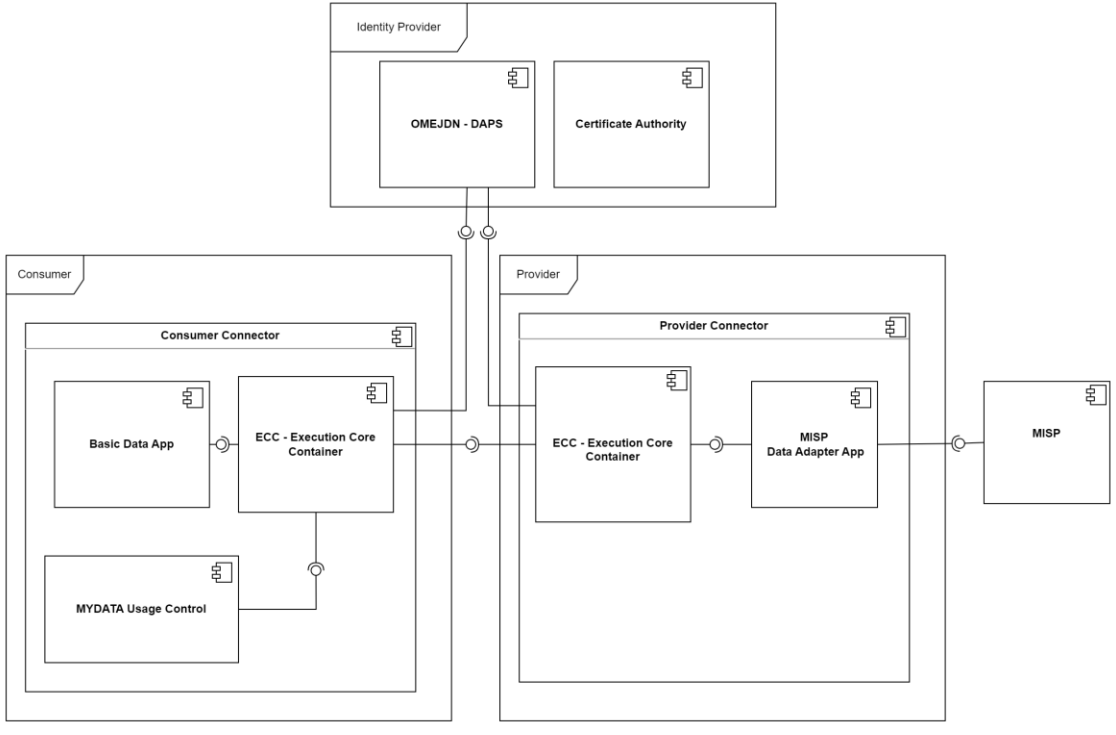
https://internationaldataspaces.org/why/data-spaces/

Implementation

Deployed Components

- Using container technology
  - docker-compose





Use Case and Requirements

High Level Requirements

1. Interoperability and decentralization
   1. Open-standards
2. Flexibility and automation
   1. Integration with existing systems
   2. Support multiple data models
3. Trust and security
   1. Participant and components certification
   2. Reputation and trust monitoring
4. Privacy and sovereignty
   1. Not share sensitive information (sanitization)
   2. Control data handling of the shared information (sharing policies)
5. Commercial activities
   1. Digital rights management (DRM)
   2. Support revenue models (marketplace)