Data Trustees:
A Whitelisting Approach for Trusted Data Sharing

Michael Steinert

Introduction

Brief Introduction

- Name is Michael Steinert
- Living in Germany near Dortmund
- Academic Qualifications: B.Sc. and M.Sc. in Computer Science
- Background in Software Development
- Research Focus: Data Spaces and Data Trustees
- Research Associate at Fraunhofer Institute for Software and Systems Engineering (Fraunhofer ISST)

Using software development to translate research questions into practical solutions

Agenda

Agenda

- What is a Data Space?
- What is a Data Trustee?
- Challenges and Research Question
- Research Methodology
- Research Findings and Implications
- Conclusion

What is a data space?

Definition: Data Space

A **Data Space** is a collaborative environment where one or more participants (typically organizations, services, or machines) share and interact with data. These environments are designed to maximize the value of data through interaction and exchange, rather than mere storage.

Functions of Data Spaces:

- **Metadata Sharing**: Metadata is shared while the data itself remains at its source.
- **Peer-to-Peer Transfer**: Direct transfer of data between participants, ensuring security and efficiency.
- **Participant Compliance**: Participants adhere to specific rules and regulations, fostering trust and reliability.
- **Infrastructure Agnostic**: Data spaces work across different environments such as on-premises, edge, or cloud systems.

What is a data trustee?

Definition: Data Trustee

A **Data Trustee** is a trusted intermediary that manages and protects sensitive data between data providers and users. Acting as an independent and neutral entity, it ensures that data is protected and managed in accordance with legal requirements and agreements.

Funktion of Data Trustees:

- **Transparency and Sovereignty** (covered by Data Spaces): Data owners retain full control over their data and can track how their data is being used.
- **Access and Permissions Management** (covered by Data Spaces): Data owners manage who can access the data, ensuring only authorized parties can interact with it.
- **Pseudonymization and Anonymization**: Techniques are applied to protect personal data by making it anonymous or pseudonymous, reducing privacy risks.
- **Monitoring and Logging** (covered by Data Spaces): Data usage is overseen, and logs are maintained to ensure compliance with agreements and regulations, enhancing accountability.

Problems and Research Question

Challenges

Core Problem: Lack of trust between data space participants, especially with sensitive data (e.g., personal or proprietary information).

Impact:

- Hesitancy in data sharing
- Limits on collaboration and innovation
- Hinders data-intensive applications (e.g., machine learning, predictive analytics)

Key Challenges:

- Establishing secure and trustworthy mechanisms for sharing sensitive data (covered by Data Spaces and Data Trustees)
- Meeting regulatory requirements and competitive concerns (covered by Data Spaces and Data Trustees)

Research Question

Role of Data Trustees:

- Critical for solving trust issues
- Manual selection of trustees is inefficient, prone to bias, and non-scalable

Research Question:

How can we automate the identification and selection of data trustees in a data space environment?

Research Methodology

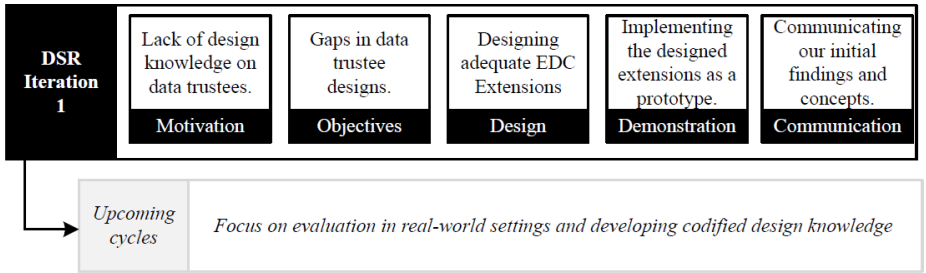
Research Methodology: Tackling Trust Challenges

Approach:

- Design Science Research (DSR) method
- Effective for problems with no pre-existing solutions (e.g., data trustees)
- Iterative process: Design, Implement, Evaluate (upcoming cycle)

Objective:

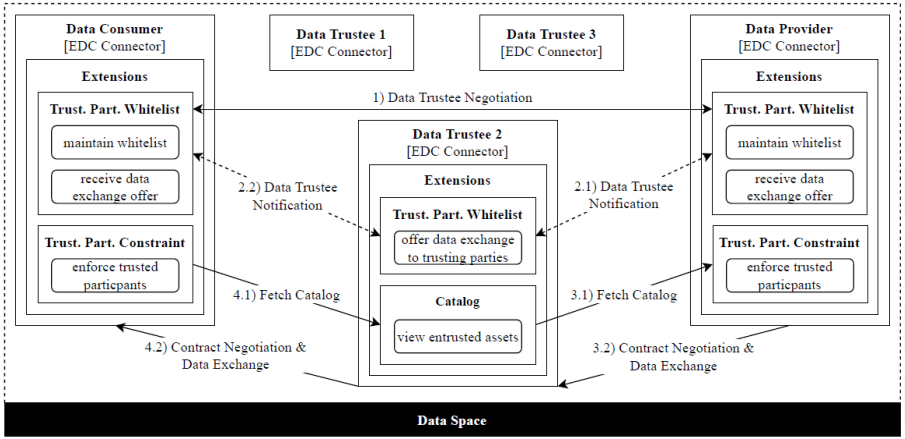
Extend Eclipse Dataspace Components (EDC) to create an automated, scalable mechanism for selecting data trustees



Research Methodology: Tackling Trust Challenges

Key Architectural Components (Artifacts):

1. Trusted Participant Constraint:
    - Automatically restricts data access to authorized parties
    - Uses verifiable credentials to ensure trust
2. Trusted Participant Whitelist:
    - Manages list of trusted entities
    - Facilitates data exchange negotiations
    - Ensures secure data-sharing



Research Findings and Implications

Research Findings

Increased Trust & Transparency

- Automated trustee selection builds confidence that only trusted entities are involved in data transactions.

Efficiency

- Automation removes manual, time-consuming, and subjective trustee selection.
- Provides a scalable and efficient solution.

Security & Compliance

- Ensures data is accessed only by authorized, trustworthy participants.
- Helps meet regulatory standards (e.g., GDPR).

Implications

- First step toward decentralized, scalable, and trustworthy data trustees in data spaces.
- Eclipse Dataspace Components (EDC) extensions provide a foundation for the community to build on.

HTTP request	Description
GET /health	Checks the health of the whitelist and returns its status.
POST /add	Adds a trusted participant to the whitelist and returns the outcome.
GET /list	Retrieves a list of trusted participants.
DELETE /remove	Removes a trusted participant from the whitelist and returns the outcome.
POST /negotiate/{counterPartyUrl}	Initiates a negotiation with another whitelist to determine common trusted participants.
POST /receive-negotiation	Handles incoming negotiation requests, matches trusted participants, and returns the negotiation outcome.
POST /notify	Receives notifications related to data trustee selection.

Conclusion

Conclusion

Key Takeaway: Addressed the challenge of trust in data spaces by automating the selection of data trustees through a whitelisting mechanism.

Impact: Promotes secure, compliant, and scalable data-sharing ecosystems.

Next Steps:

- Apply findings to real-world use cases
- Refine architectural design based on practical implementations

Looking Ahead: Excited to continue research and welcome community feedback to enhance trust and functionality in data spaces.

Thank you