

Challenges & Lessons learnt

Challenges & Lessons learnt from TEADAL & TRUSTEE: Creating Trustworthy European Data Space Federation

Sebastian Werner & Alberto
Berreteaga
27 October, 11:30 – 12:30



IN COLLABORATION WITH



INCLIVA | VLC
Biomedical Research Institute



UNDER THE AUSPICES OF



Agenda

- Speaker Introduction
- Trust Challenges in Data Spaces
- a TEADAL Perspective
- a TRUSTEE Perspective
- Open Panel Discussion



Trust Challenges in Data Spaces

- Adapting legal and regulatory frameworks to many heterogenies data space implementations
- Automatic selection of appropriate means to ensure compliant data access from a growing portfolio of solutions.
- Enabling transparent yet confidential auditing and reviews of data exchanges

TRUSTWORTHY, ENERGY-AWARE FEDERATED DATA LAKES ALONG THE COMPUTING CONTINUUM

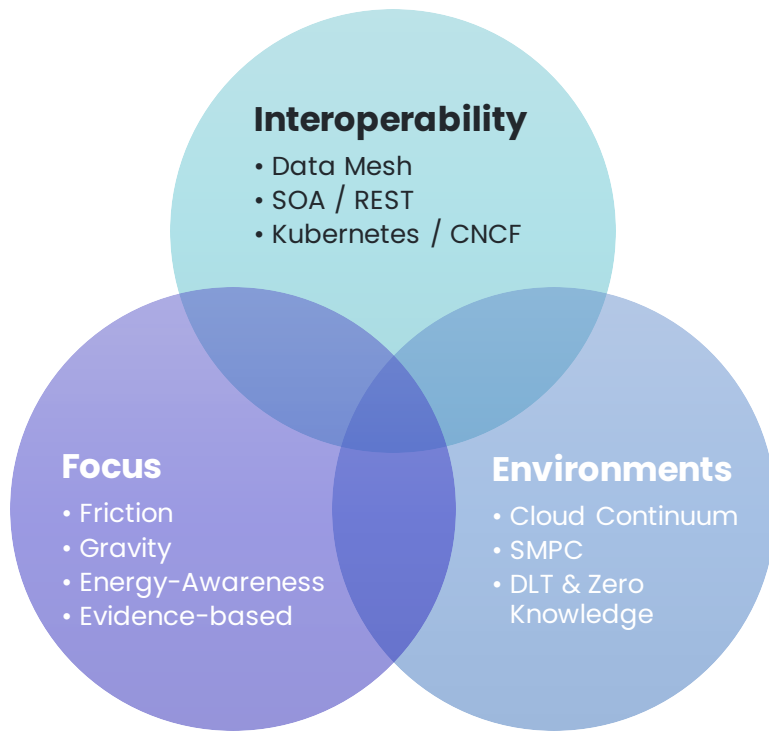
Dr. Sebastian Werner,
TU Berlin



Teadal's Objectives

- 1 To establish confidence in handling of data across the continuum and deliver efficiency for building and using stretched data lakes solutions.
- 2 To enable the construction of trustworthy data lakes and mediatorless federation of trustworthy data lakes.
- 3 To reduce the environmental impact of data analytics by carefully managing how data are stored, reused, moved, and processed in a federation of stretched data lakes.
- 4 To simplify the specification and enforcement of privacy/confidentiality requirements, constraints and policies for federated stretched data lakes to be compliant with regulations, norms, and organizations' policies.
- 5 To contribute and influence research, data-centric European initiatives, open-source communities, and industry with methods, and tools to improve data sharing.

Teadal in the Data Space Ecosystem



Rich pool of Pilot-Cases



Health

Marina Salud, Spain



Industrial

ERT, Portugal



Agriculture

Terraview, Switzerland



Finance

ING, The Netherlands



Mobility

AMTS, Italy & UITP, Belgium



Energy/Green Deal

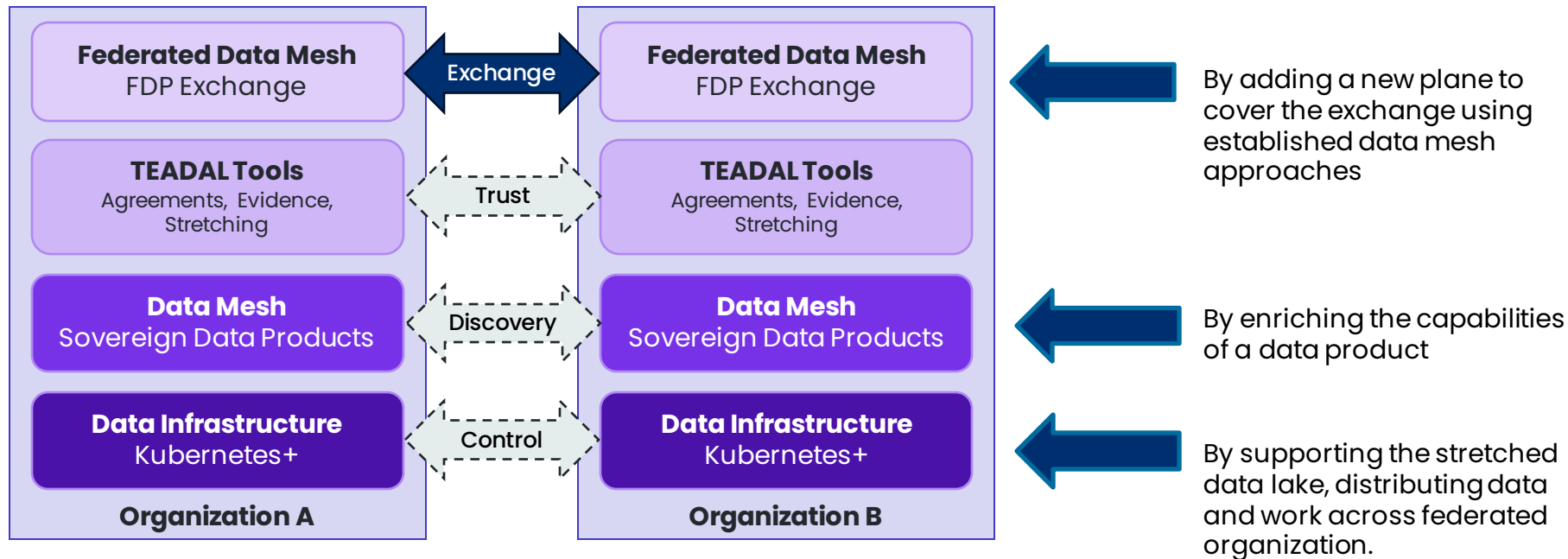
Box2M, Romania



Public Administration

Regione Toscana, Italy

Teadal's Data Exchange Model



Federated Data Products

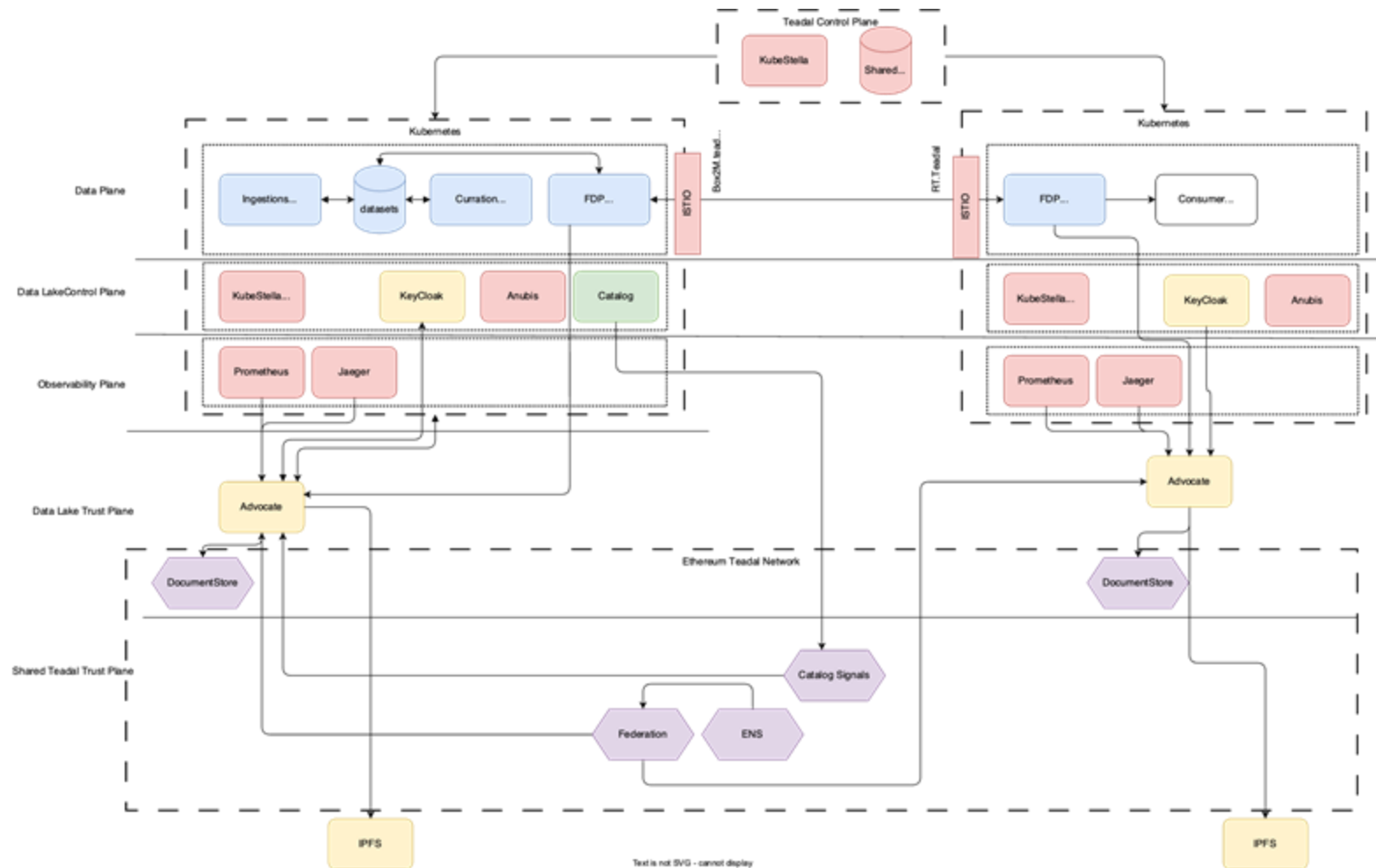
- Self-describing data service, accessible through REST APIs
- Deployable in CNCF environments, e.g., Kubernetes, K3s
- Governance through federated data catalog, incl. policies, capabilities
- Consumption only based on enforced agreements
- Data owner can audit, control and validate data consumption

Core Challenges

1. How to reduce wasted data movement, executions while ensuring data exchange compliance?
2. **How to establish trustworthiness of data consumers?**
3. How can we exploit the cloud continuum to reduce energy consumption without hindering data exchange?
4. ...

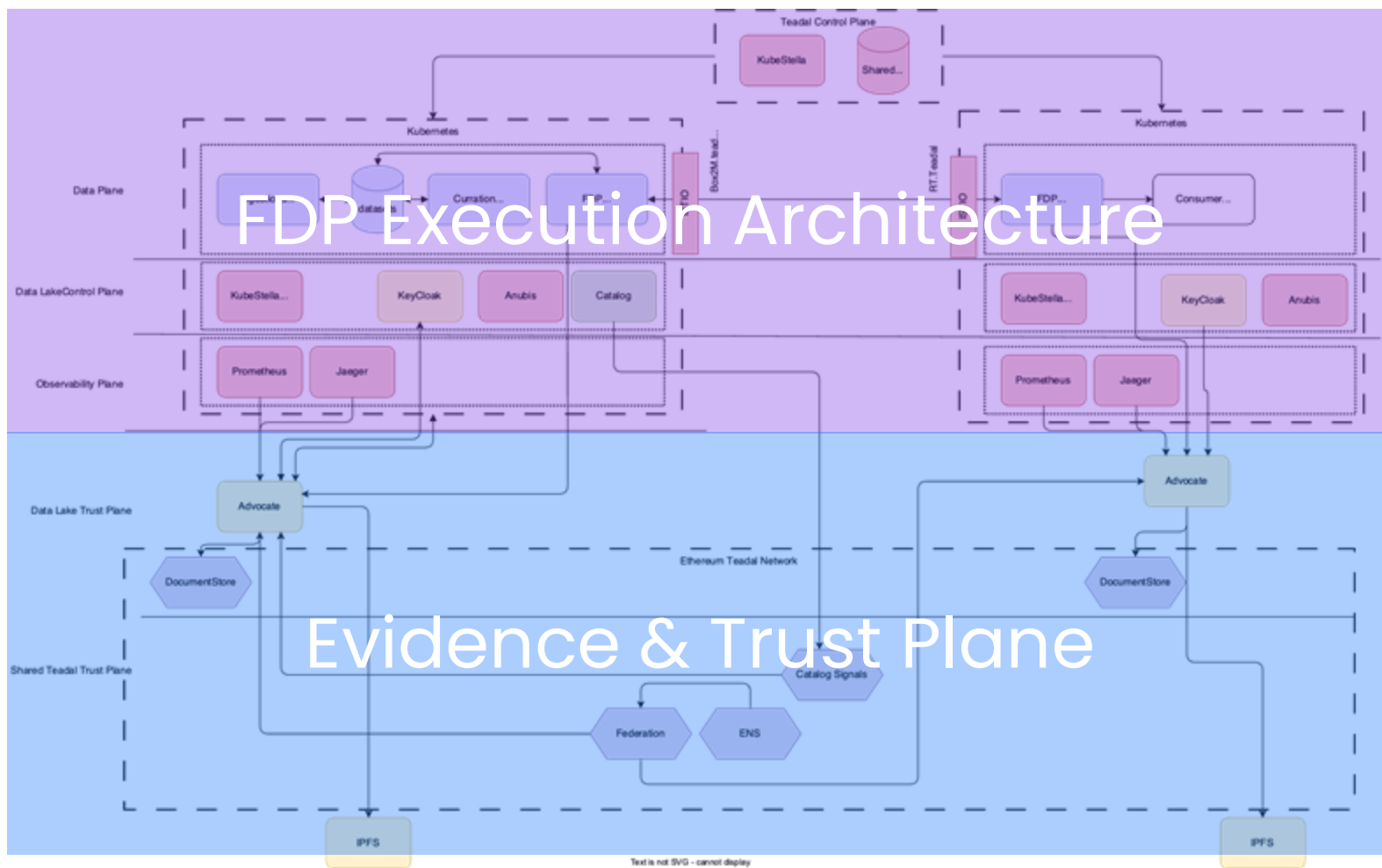
Teadal's Evidence-based System Design

- **Verifiable data provenance**
- Agreement compliance
- Realtime data flow observability on a user level
- Flexible policy enforcement for deployment, consumption and execution level



Text is not SVG - cannot display

FDP Execution Architecture



Evidence & Trust Plane

Evidence Sources for Audit/Validation/Verification

Lifecycle Events

- discovery, agreement, consumption, ...

Catalog:

BPMN based reporting to SmartContracts

Access / Enforcement Events

- user access and enforcement events

OPA:

immutable storing of decision record and policy code

Platform Events

- platform level events through tracing and instrumentation on all components in the data exchange

Open Telemetry:

Convert observations into immutable and verifiable records

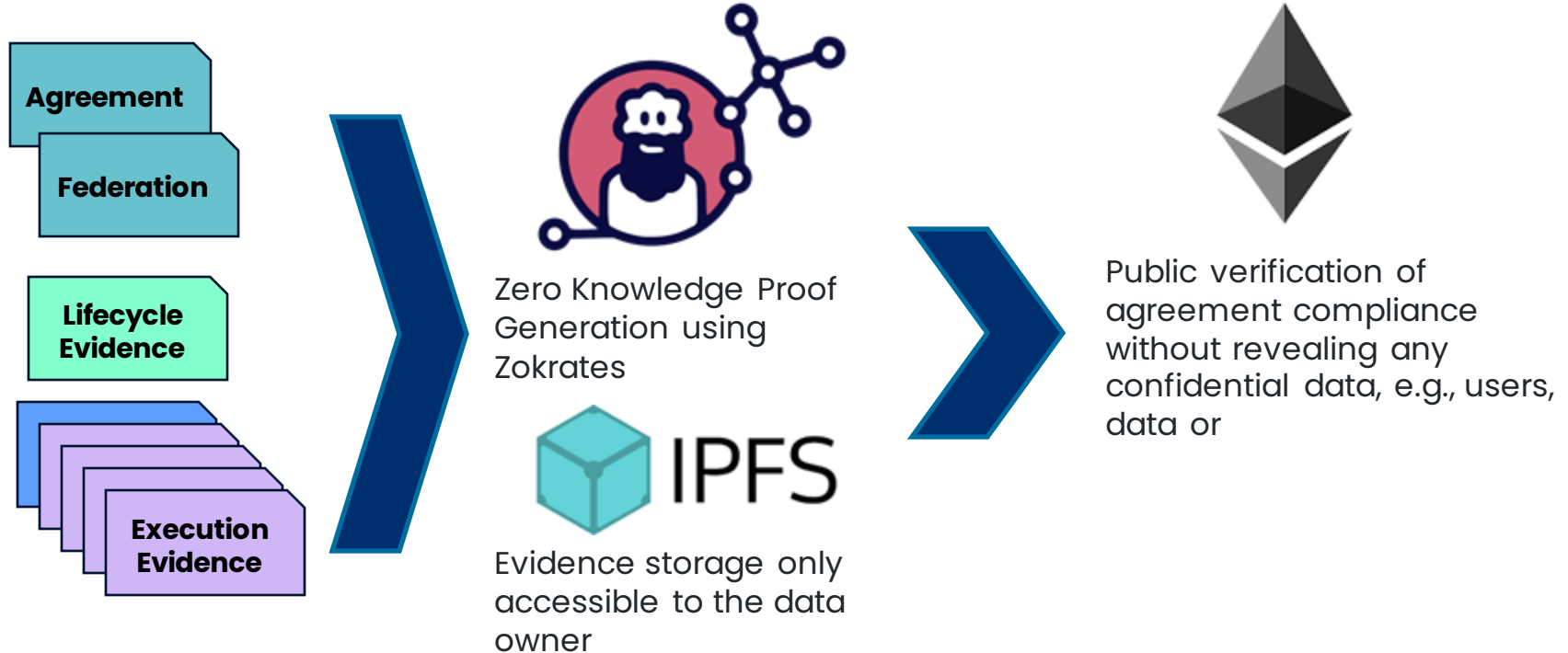
FDP Events

- user interactions and processing steps

Advocate:

Explicit verification of what data is provided under which agreement for which user

Confidential Evidence Verification





Follow our progress



<https://www.teadal.eu/newsletter/>



www.linkedin.com/company/teadal/



[@TEADAL_eu](https://twitter.com/TEADAL_eu)



<https://www.tu.berlin/ise/teadal>



werner@tu-berlin.de

Meet the TEADAL Consortium

ubiwhere



CYBERNETICA

Cefriel





**Funded by the
European Union**



TRUSTEE

Funded by the European Union.

Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

Project Facts

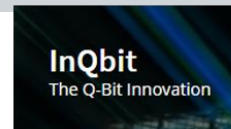


- **Duration:** 3,5 years | 42 months | Started 1st July 2022
- **Consortium:** 22 partners
- **Demonstrations:** in 6 pilots
- **Project's Budget:** € 8 706 263,75
- **Funding Scheme:** Digital, Industry and Space
- **Work Programme:** Horizon Europe
- **GA No:** 101070214
- **Coordinator:** Foundation for Research & Technology – Hellas

Consortium



TRUSTEE



ATHENA Research & Innovation
Information Technologies



ZORTENET



etoile




UNIVERSITÀ
CATTOLICA
del Sacro Cuore



About TRUSTEE

In accordance with the principles of responsible/trustworthy AI by employing a co-development approach as the foundation methodology TRUSTEE proposes a secure-by-design Federated Platform in accordance with EU data strategy (COM (2020) 66) and the main EU reference architectures (GAIA-X, EOSC, EGI) in the sector, capable of ensuring interoperability, enabling cross-border scenarios, and scaling a variety of AI-based applications through the use of open APIs, with the goal of making the EU the world's most secure and trustworthy data hub.



To achieve this, we employ a novel homomorphic approach that ensures user-friendly, safe, trustworthy, compliant, fair, transparent, accountable, and long-term data collection, storage, processing, querying, and delivery.

About TRUSTEE



TRUSTEE will **analyse the current threat landscape, the security and trust requirements** for sharing and processing heterogeneous cross border data, while evaluating the **security and privacy requirements** of **data sharing** from six critical sectors to improve their openness and fairness in line with national and European legislations.

Our pilots:



Education



Space



Data



Automotive

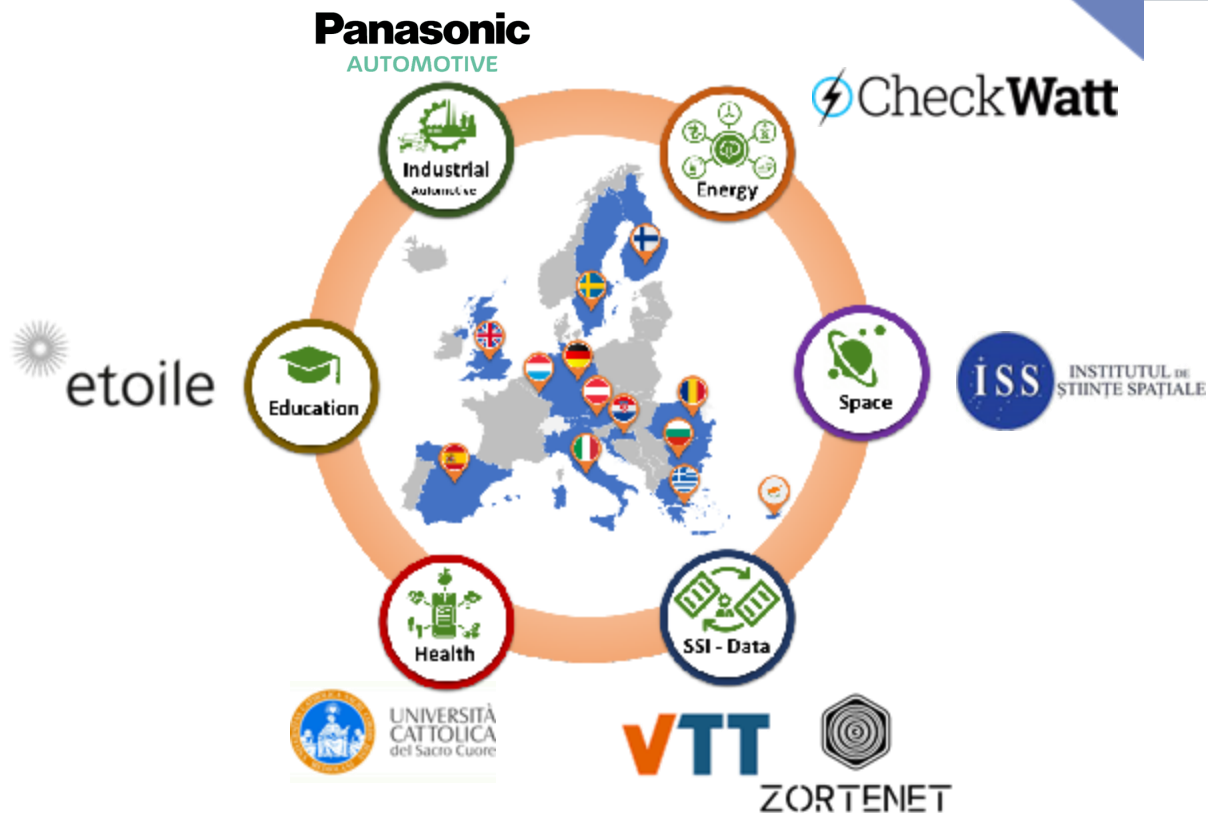


Energy



Health

TRUSTEE Use Cases



Objectives



TRUSTEE

1. Design a Secure and Trust framework and Reference Architecture to ensure end-to-end trust and privacy for stakeholders across European data

2. Design and develop a distributed Homographic-capable self-sovereign framework for accessing, sharing and manipulating data, compliant to national and EU legislation

3. Novel Big Data management and analytics infrastructure to facilitate use and re-use of data in data spaces across interdisciplinary science fields and business sectors

Objectives

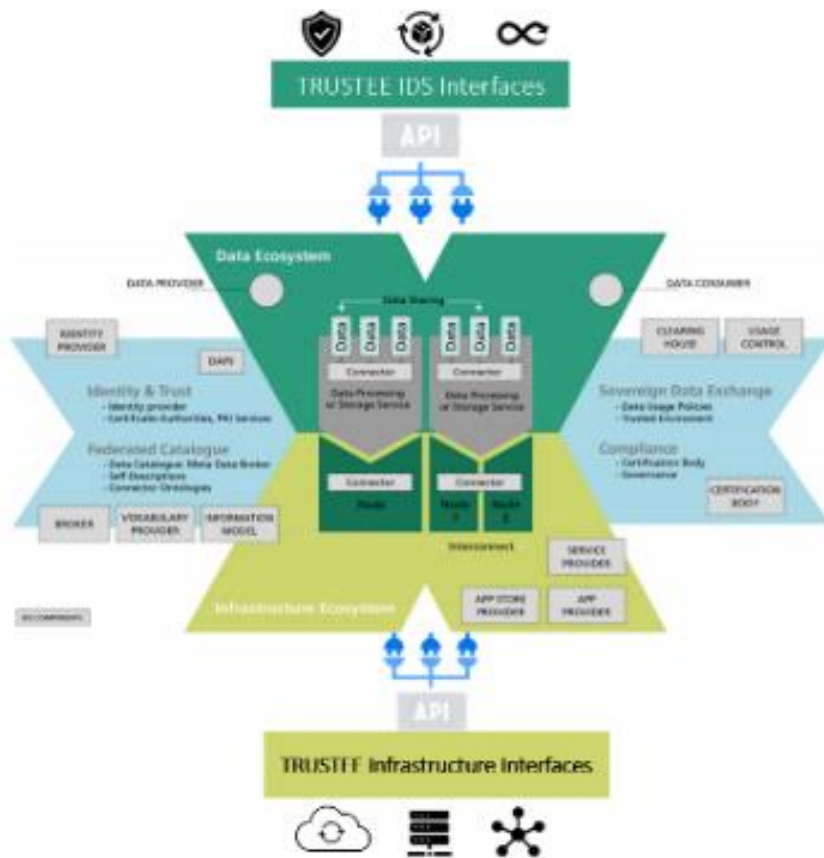


TRUSTEE

4. Analyze the ethical, legal, privacy and IPR issues for collection, storage, processing, querying, analytics and delivery of data enabling the European single market for data

5. Design and develop TRUSTEE explainable and trustworthy AI for efficient and robust use and re-use of data and metadata across interdisciplinary domains

6. Provide OPEN and FAIR databases facilitating sharing and manipulation of data in compliance with prevailing and emerging legislation (e.g. GDPR)



Data Pilot – TRUSTED MULTI-DISCIPLINARY DATA EXCHANGE PILOT

Objectives



TRUSTEE

TRUSTEE will leverage the secure IDS interfaces and channels to establish the necessary pathway for multi-stakeholder data exchange and data sovereignty across its platform implementing a set of IDS interfaces with the GAIA-X federation layer to allow secure data inference across the entire platform.

This use case will enhance the security and privacy aspect of TRUSTEE by creating a trustworthy environment for stakeholders to collaborate over heterogeneous business agreements with each other.



Data Type



The IDS Information Model is a modular meta-model (ontology) describing the capabilities of IDS infrastructure components, such as the Connector or the Data Endpoints. Extending the Open Digital Rights Language (ODRL), a W3C standard, the Information Model's Usage Control module provides machine- readable specifications of usage control policies.

To implement and enforce usage policies at a specification level within individual target environments, it is necessary to map organizational and technical measures to the individual target environments.

While organizational measures are out of scope here, technical measures involve a variety of additional information sources (PIPs) and tight integration with the host environment (PEPs).

Here, the Information Model enhances ODRL constructs via predefined extension “hooks” to support mapping onto lower-level, implementation-oriented policy languages (e.g., INDUCE XML).

KPIs

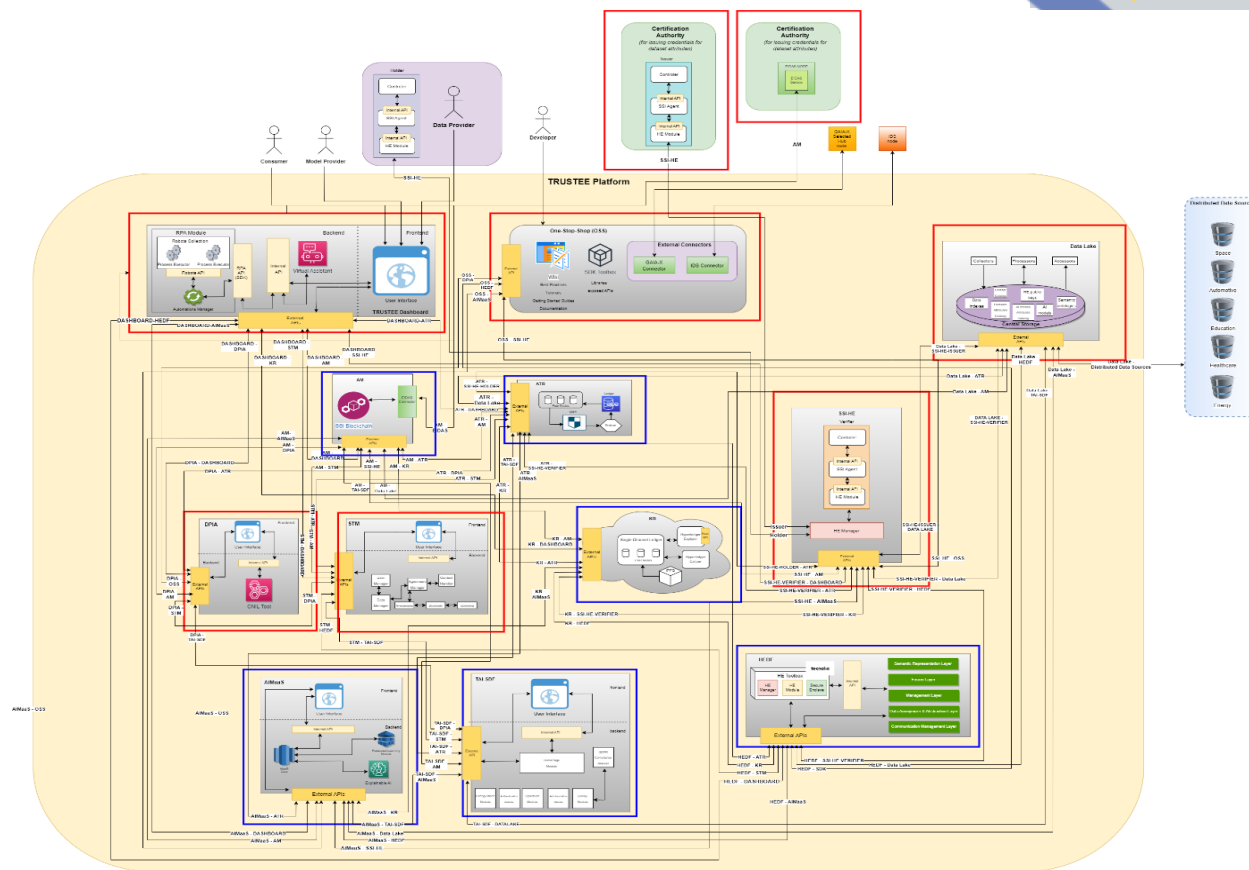


- Validate trust, reputation, and cooperation mechanisms for at least three (3) data providers in the TRUSTEE ecosystem, including at least three (3) confidence-focused parameters included in SSI e.g., successful delivery of computation, total time of computation and reliability of resource (measured as percentage of uptime).
- Achieve at least 70% People satisfaction in the following processed (user-friendly, safe, trustworthy, compliant, fair, transparent, accountable)
- Ensure interoperability and reasonable re-use of common reference models of Gaia-X and EOSC.
- Facilitate at least 500 Person to use and validate TRUSTEE platform.
- Ensure data subjects/rightsholders and other stakeholders 100%

TRUSTEE Detailed architecture

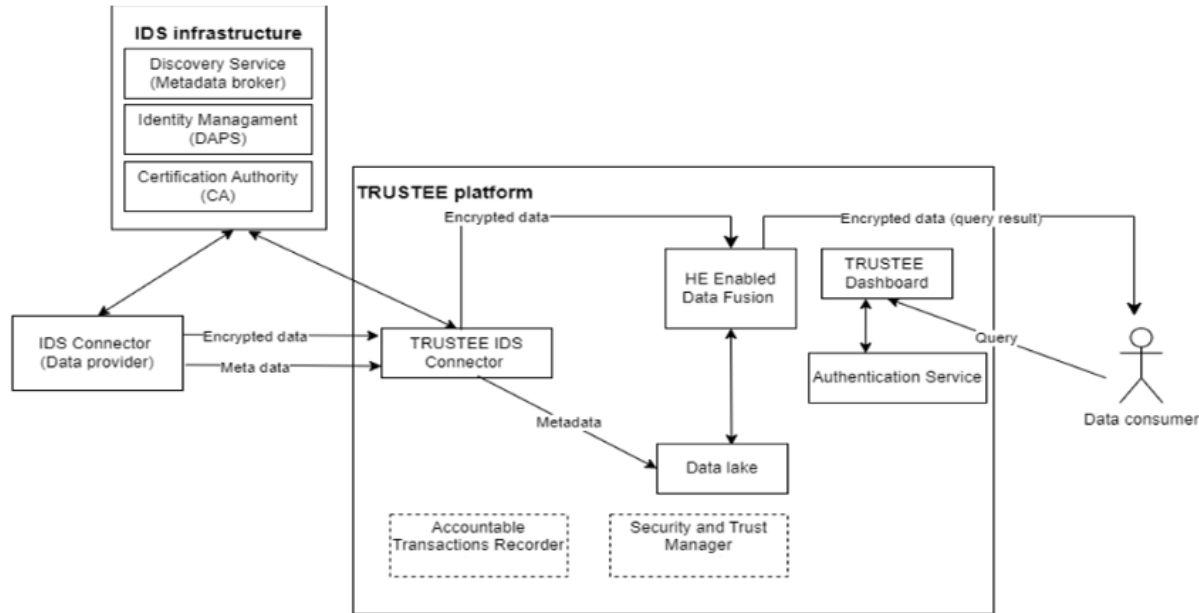


TRUSTEE

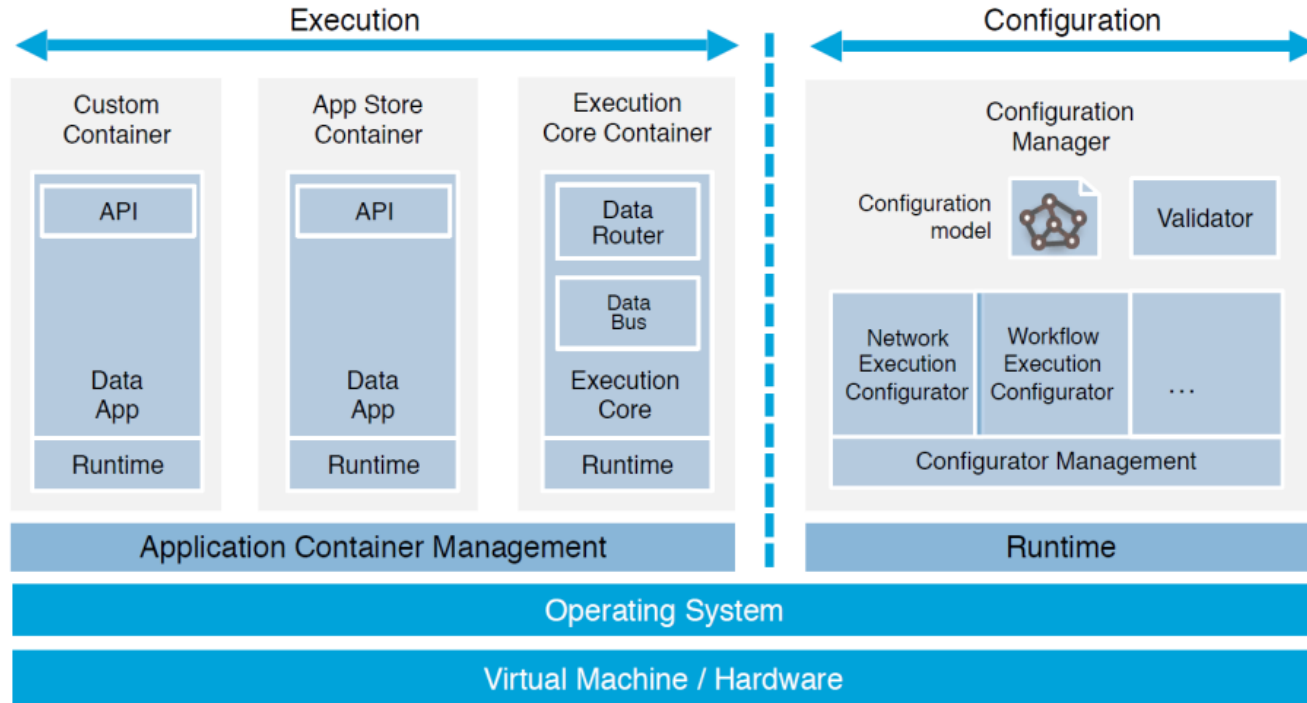




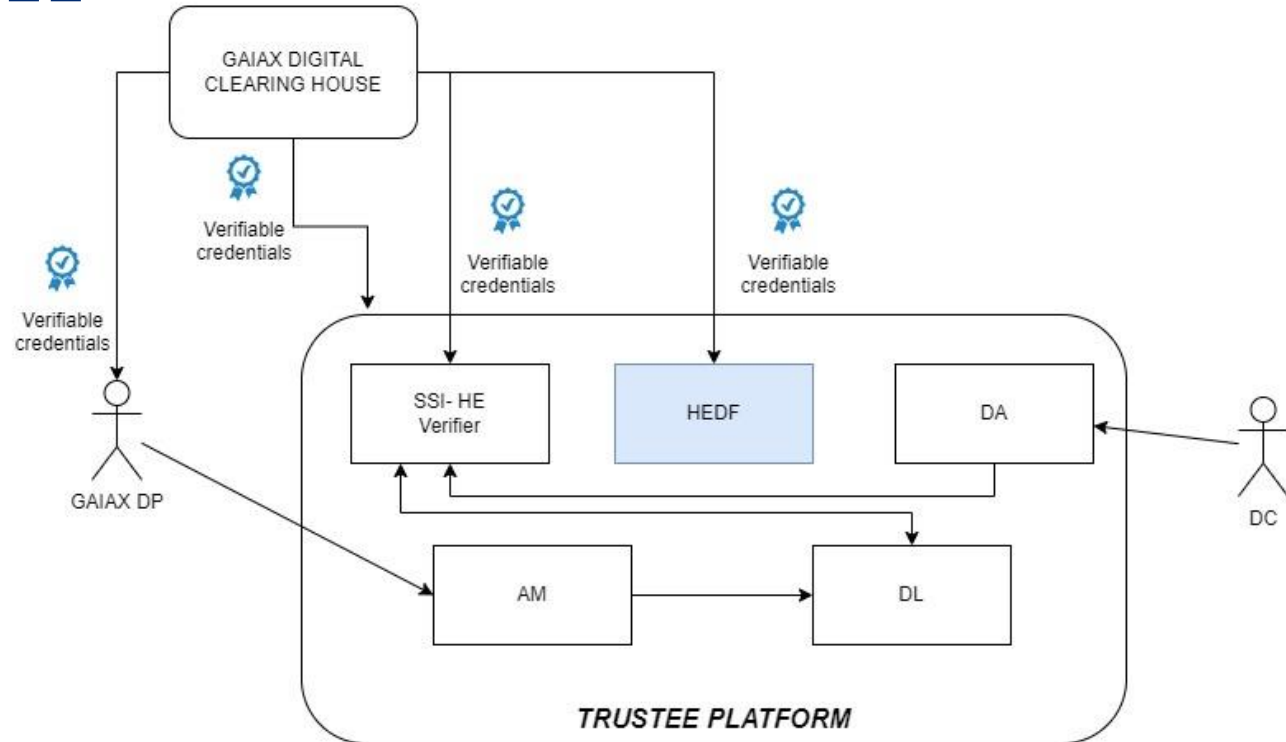
IDS connector in integration with TRUSTEE



IDS connector architecture



GAIAX integration with TRUSTEE





Follow our progress



<https://horizon-trustee.eu>



<https://www.linkedin.com/company/horizon-trustee/>



<https://twitter.com/TrusteeHorizon>



<https://www.youtube.com/@horizon-trustee6566>

Thank you!

IN COLLABORATION WITH



INCLIVA | VLC
Biomedical Research Institute



UNDER THE AUSPICES OF

