TEADAL

D6.3

PILOT CASES DEPLOYMENT RESULTS AND ANALYSIS

Revision: v0.2

Work package	WP6
Task	T6.4
Due date	31/10/2025
Submission date	31/10/2025
Deliverable lead	Ubiwhere
Version	1.0
Authors	Aitor López (FUTURS); Aleix Boixader Coma (i2cat); Alessio Carenini (CEFRIEL); Ana Pereira (UW); Boris Sedlak (TUW); Bruno Feitais (UW); Davide Bruno (Regione Tuscana); Eduardo Brito (CYB); Evgeny Bogdanov (Terraview); Fabian Piper (TUB); Fabio Cartolano (FIT); Fernando Castillo (TUB); Josep Escrig Escrig (i2cat); Katherine Barabash (IBM); Matteo Falconi (POLIMI); Monica Vitali (POLIMI); Ovidiu Diaconescu (BOX2M); Patricia Fernández (FUTURS); Pierluigi Plebani (POLIMI); Pille Pullonen-Raudvere (CYB); Ricardo Reis (ERT); Sepideh Masoudi (TUB); Sergio Sestili (ALMAVIVA); Samantha Hine (ALMAVIVA); Vincenzo Cirillo (ALMAVIVA); Vojtech Cima (Terraview)
Reviewers	Fernando Castillo (TUB); Pille Pullonen-Raudver (CYB)
Abstract	This document reflects the findings on pilot cases' deployment. All TEADAL components involved in the pilot cases' deployment are identified, clearly

Grant Agreement No.: 101070186
Call: HORIZON-CL4-2021-DATA-01
Topic: HORIZON-CL4-2021-DATA-01-01
Type of action: HORIZON-RIA

	demonstrating the impact such a platform has on different use cases.
Keywords	CI/CD, Cluster, Deployment, Energy, Integration, Pipeline, Pilot cases, Testbed, Trial

Document Revision History

Version	Date	Description of change	List of contributor(s)
V 0.1	11/06/2025	ToC Initial Draft	Bruno Feitais (UW); Sergio Sestili (ALMAVIVA); Samantha Hine (ALMAVIVA)
V 0.2	12/06/2025	ToC Initial Draft	Bruno Feitais (UW); Sergio Sestili (ALMAVIVA); Samantha Hine (ALMAVIVA)
V 0.5	10/10/2025	Complete contents, ready for review	Aitor López (FUTURS); Aleix Boixader Coma (i2cat); Alessio Carenini (CEFRIEL); Ana Pereira (UW); Boris Sedlak (TUW); Bruno Feitais (UW); Davide Bruno (Regione Tuscana); Eduardo Brito (CYB); Evgeny Bogdanov (Terraview); Fabian Piper (TUB); Fabio Cartolano (FIT); Fernando Castillo (TUB); Josep Escrig Escrig (i2cat); Katherine Barabash (IBM); Matteo Falconi (POLIMI); Monica Vitali (POLIMI); Ovidiu Diaconescu (BOX2M); Patricia Fernández (FUTURS); Pierluigi Plebani (POLIMI); Pille Pullonen-Raudvere (CYB); Ricardo Reis (ERT); Sepideh Masoudi (TUB); Sergio Sestili (ALMAVIVA); Samantha Hine (ALMAVIVA); Vincenzo Cirillo (ALMAVIVA); Vojtech Cima (Terraview)
V 1.0	31/10/2025	Final version	Aitor López (FUTURS); Aleix Boixader Coma (i2cat); Alessio Carenini (CEFRIEL); Ana Pereira (UW); Boris Sedlak (TUW); Bruno Feitais (UW); Davide Bruno (Regione Tuscana); Eduardo Brito (CYB); Evgeny Bogdanov (Terraview); Fabian Piper (TUB); Fabio Cartolano (FIT); Fernando Castillo (TUB); Josep Escrig Escrig (i2cat); Katherine Barabash (IBM); Matteo Falconi (POLIMI); Monica Vitali (POLIMI); Ovidiu Diaconescu (BOX2M); Patricia Fernández (FUTURS); Pierluigi Plebani (POLIMI); Pille Pullonen-Raudvere (CYB); Ricardo Reis (ERT); Sepideh Masoudi (TUB); Sergio Sestili (ALMAVIVA); Samantha Hine (ALMAVIVA); Vincenzo Cirillo (ALMAVIVA); Vojtech Cima (Terraview)

DISCLAIMER

Funded by the European Union

Funded by the European Union (TEADAL, 101070186). 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.

COPYRIGHT NOTICE

Project funded by the European Commission in the Horizon Europe Programme		
Nature of the deliverable:	**R**	
Dissemination Level		
PU	*Public, fully open, e.g. web (Deliverables flagged as public will be automatically published in CORDIS project's page)*	✔
SEN	*Sensitive, limited under the conditions of the Grant Agreement*	
Classified R-UE/ EU-R	*EU RESTRICTED under the Commission Decision No2015/ 444*	
Classified C-UE/ EU-C	*EU CONFIDENTIAL under the Commission Decision No2015/ 444*	
Classified S-UE/ EU-S	*EU SECRET under the Commission Decision No2015/ 444*	

* R: Document, report (excluding the periodic and final reports)

DEM: Demonstrator, pilot, prototype, plan designs

DEC: Websites, patents filing, press & media actions, videos, etc.

DATA: Data sets, microdata, etc.

DMP: Data management plan

ETHICS: Deliverables related to ethics issues.

SECURITY: Deliverables related to security issues

OTHER: Software, technical diagram, algorithms, models, etc.

Funded by the European Union

EXECUTIVE SUMMARY

This deliverable presents the deployment results and validation outcomes of the TEADAL platform across a diverse set of pilot cases, demonstrating the platform's maturity, scalability and applicability in real-world environments. TEADAL aims to establish a Trustworthy and Energy-Aware federated Data Lake that enables secure, privacy-preserving, and efficient data sharing across domains such as healthcare, mobility, industry, agriculture, environmental sustainability, and finance.

The deliverable summarises the full integration process from the deployment of TEADAL Nodes and their components to the operation of pilot testbeds, validating the platform through practical trials. Each pilot demonstrates how TEADAL's technologies facilitate federated data management through components such as CI/CD pipelines, AI-driven monitoring, policy enforcement, evidence-based trust mechanisms, and energy profiling.

Across all pilots, the platform enabled:

- Federated data sharing without compromising data sovereignty or privacy;
- Automation of deployment and governance, significantly reducing configuration time and human error;
- Transparent and auditable data flows, ensuring accountability and compliance with GDPR and FAIR principles;
- Energy-aware analytics, demonstrating measurable efficiency gains in computation and data transfer.

Notable achievements include:

- In healthcare, secure multi-centre clinical research with verifiable patient-consent tracking;
- In mobility, a federated access point aligning national and regional transport datasets under EU open-data policies;
- In industry, automated, policy-compliant KPI calculation across distributed manufacturing sites;
- In agriculture, privacy-preserving collaboration among vineyards for precision farming;
- In regional planning, integrated environmental and energy data supporting sustainable policy decisions;
- In finance, trusted sharing of sensitive financial data under federated governance.

The validation phase confirmed TEADAL's compliance with key performance indicators, evidencing reductions in deployment complexity, data movement, and resource usage while improving trustworthiness and scalability.

In conclusion, TEADAL demonstrates a replicable framework for building trustworthy, energy-efficient, and federated data ecosystems. Its pilots validate both the technical robustness and the practical value of the platform.

Funded by the European Union

TABLE OF CONTENTS

Funded by the European Union

Funded by the European Union

Funded by the European Union

LIST OF FIGURES

Funded by the European Union

LIST OF TABLES

Funded by the European Union

ABBREVIATIONS

AI-DPM	Artificial Intelligence-Driven Performance Monitoring
API	Application Programming Interface
ASG	Automatic sFDP Generation
CD	Continuous Delivery/Deployment
CI	Continuous Integration
DNS	Domain Name System
DSPN	Data Sharing Policy Notation
ESG	Environmental, Social and Governance
EU	European Union
FDP	Federated Data Product
GA	General Assembly
GDPR	General Data Protection Regulation
GUI	Graphical User Interface
HTTP	Hypertext Transfer Protocol
IaC	Infrastructure as Code
IP	Internet Protocol
K8s	Kubernetes
KPI	Key Performance Indicator
LLM	Large Language Model
MPC	Multi-Party Computation
NAP	National Access Point
OPA	Open Policy Agent
OTA	Over-The-Air
RAP	Regional Access Point
REST	Representational State Transfer
SSO	Single Sign-On
sFDP	shared Federated Data Product

Funded by the European Union

TEE	Trusted Execution Environment
TNS	TEADAL Name Service
VM	Virtual Machine

Funded by
the European Union

1. INTRODUCTION

The present deliverable D6.3 - “Pilot Cases Deployment Results and Analysis” reports the deployment outcomes and validation of the TEADAL platform across all pilot sites. It builds upon the previous deliverables of Work Package 6 (WP6), focusing on the practical realisation and operational testing of TEADAL technologies in real-world environments.

The TEADAL project aims to deliver a Trustworthy, Energy-Aware federated Data Lake platform that enables secure, efficient and privacy-preserving. Within this context, WP6 is responsible for demonstrating the platform’s capabilities through a diverse set of pilot cases, each representing a concrete application domain such as healthcare, mobility, industry, smart agriculture, regional energy planning and finance.

This deliverable summarises the deployment process, architecture and configuration of the TEADAL Nodes used in each pilot. It describes the integration of all technical components, including CI/CD pipelines, Kubernetes-based orchestration, policy enforcement tools (OPA), monitoring stacks (Prometheus, Grafana, Kepler), and AI-driven performance monitoring (AI-DPM), that together constitute the operational backbone of the TEADAL framework.

In addition to deployment results, the document provides an overview of the validation process and Key Performance Indicators (KPIs) used to assess the impact and maturity of the TEADAL solution. These KPIs address dimensions such as scalability, trustworthiness, privacy, energy efficiency and federation readiness, offering measurable evidence of the project’s technological and scientific achievements.

Ultimately, this deliverable serves as both a technical reference and an operational validation report, ensuring that the deployed platform fulfills the objectives of the project and paves the way for its final evaluation and potential exploitation in future data-sharing ecosystems.

Funded by
the European Union

2. TEADAL NODE DEPLOYMENT

This section presents the final deployment and architectural realization of the TEADAL Node, the foundational building block of the TEADAL platform. The TEADAL Node serves as a self-contained environment that integrates data management, policy enforcement, trust assurance, and observability capabilities to enable trustworthy federated data sharing. It combines a robust baseline data lake, built upon open-source technologies for storage, identity, orchestration, and monitoring, with a suite of TEADAL specific tools that provide advanced functionalities for cataloging, policy compliance, lifecycle management, and federation. The following subsections describe in detail the logical and operational architecture of the TEADAL Node (Section 2.1) and the specialized tools that extend its functionality (Section 2.2), illustrating how the platform's components interact to deliver a secure, auditable, and scalable data-sharing infrastructure across pilot environments.

2.1 TEADAL Node Diagram

This section presents the current and final status of the TEADAL Node architecture, with a focus on the baseline data lake and integrated tool components, as deployed and validated. The architecture shown below represents the final structure resulting from the iterative evolution that started with the initial specification defined in deliverable D2.2 "Pilot cases' intermediate description and initial architecture of the platform" (M15). While D2.2 focused on high-level component section, deployment strategies, and Kubernetes orchestration principles, the current diagram reflects the actual distribution, including namespace separation, software modules, and optional components.

Logical architecture of the TEADAL platform

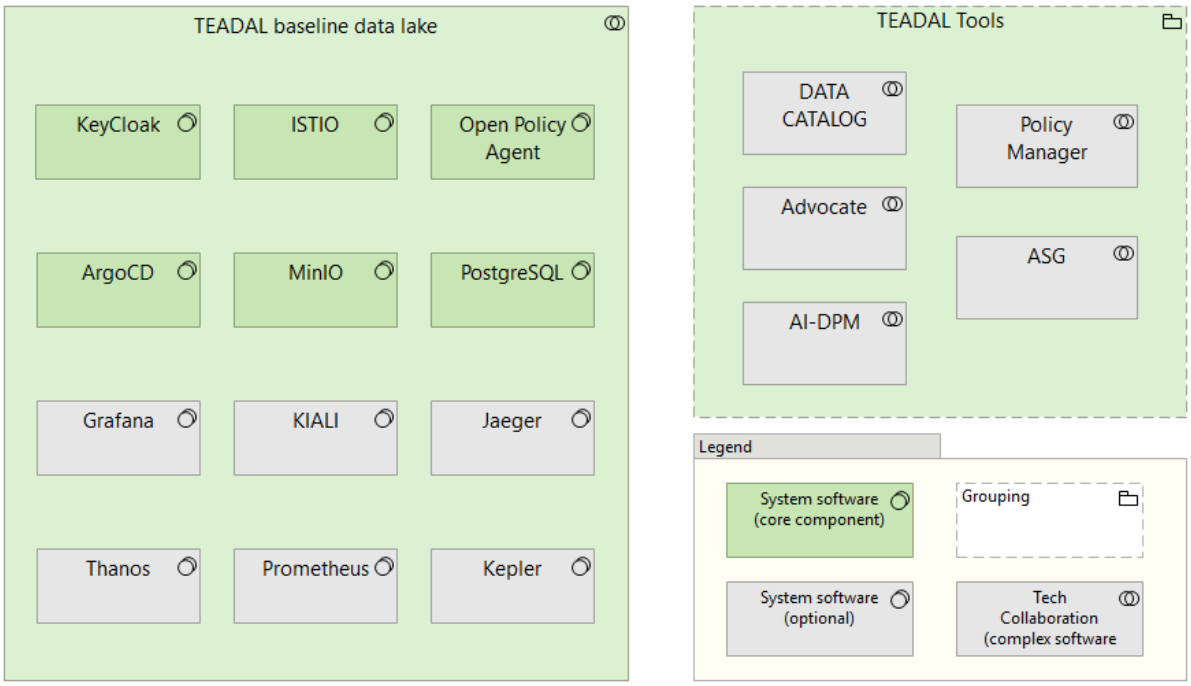


Figure 1 - TEADAL software

The TEADAL Node is logically organized into two main areas. The **baseline data lake**, which includes essential system components providing storage, identity and access management, policy enforcement, service mesh, and observability features. This layer comprises open-source tools such as Keycloak (identity and access management), Istio (Service mesh

Funded by the European Union

and traffic management layer), Open Policy Agent (OPA, Policy enforcement point), ArgoCD (GitOps-based continuous deployment), MinIO (Object storage service, compatible with S3 APIs), PostgreSQL (Relational database used by various tools, e.g. Advocate), Prometheus, Thanos (Monitoring and historical metrics persistence), Kiali, Jager, Grafana (Observability and tracing), and Kepler (Power consumption metrics collector). Together, they enable secure data management, workflow orchestration, and monitoring at the node level, following modern data lake architecture patterns. The **TEADAL Tools** layer includes specialized complex optional components developed by the project. The Data Catalog, Advocate, and the AI-DPM components leverage the baseline services and expose APIs for integration with external actors such as the Policy Manager and the ASG Pipeline Generator. The logical design ensures modularity, extensibility, and consistency across nodes, allowing each pilot to configure independently its nodes while conforming to the project's architectural standards. This logical view is meant to clarify the functional roles of each component, while a more complete description of baseline components is given in D2.2 and and a more complete description of TEADAL Tools is given in 2.2 subsection and in deliverable D2.4 "Final general architecture".

Kubernetes-Based Deployment Topology of the Baseline Data Lake

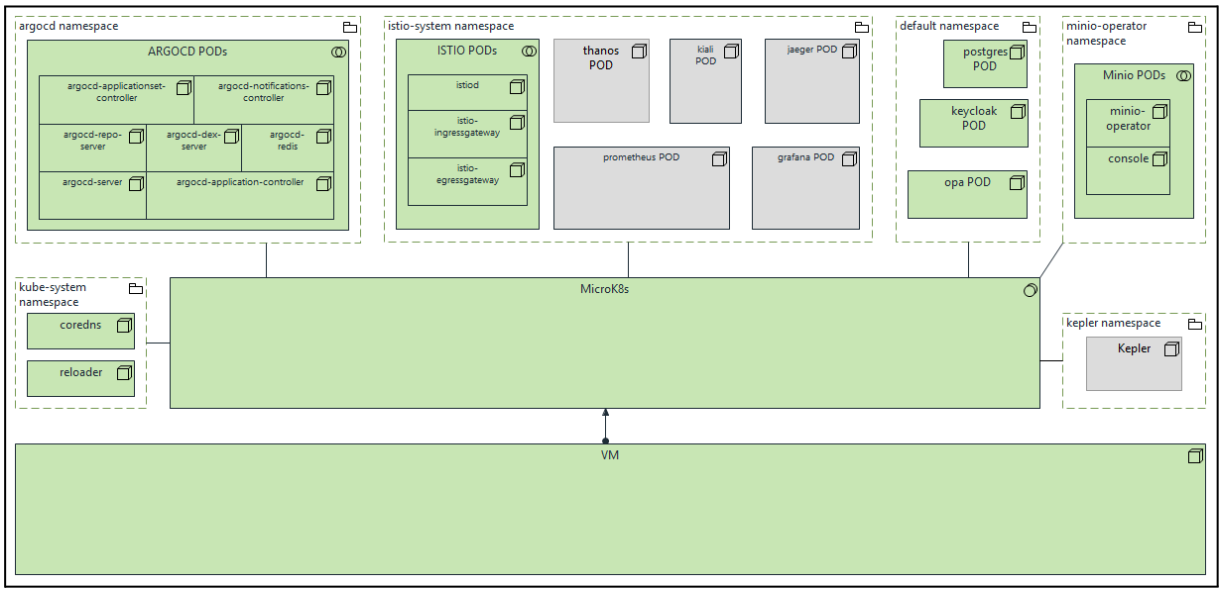


Figure 2 - TEADAL baseline datalake Kubernetes distribution

The diagram presents a more operational perspective of the TEADAL Node, showing how the software components are deployed within a Kubernetes environment. Each block corresponds to a namespace, and inside each namespace are the core Pods or operators associated with the installed services. For instance, ArgoCD components are deployed in their dedicated namespace, comprising the application controller, server, repo server, and associated notification services. The Istio system is deployed under the istio-system namespace and includes the ingress gateway, sidecars, and the control plane - referring to the standard Kubernetes/Istio control plane components responsible for traffic management, service directory, and configuration within the cluster. Monitoring tools such as Prometheus, Thanos, Grafana, Jaeger, and Kiali are grouped within system namespaces to ensure observability, with Thanos extending the monitoring stack for long-term metrics storage. The default namespace hosts simpler components such as PostgreSQL, Keycloak, and OPA, primarily for ease of deployment integration. The minio-operator and kepler are isolated in their own namespaces (minio-operator and kepler, respectively), taking advantage of Kubernetes operators for lifecycle management. In addition, system-level utilities like Reloader and Cert-Manager are included to ensure proper configuration refresh and TLS

Funded by the European Union

support. The entire node runs on a virtualized infrastructure (VM) managed through MicroK8s, a lightweight Kubernetes distribution that has proven suitable for edge environments and pilot nodes with constrained resources. This configuration was adopted consistently across the deployed pilots, ensuring uniformity in deployment, monitoring, and control capabilities. The combination of logical architecture and deployment topology offers a comprehensive overview of the TEADAL Node's structure. It also showcases the evolution of the project from the initial design to operational deployment, with a clear namespaces separation, improved containerization, and support for GitOps-based CI/CD flows. Multiple iterations across pilot environments helped refine the node setup and component integration process, establishing a mature reference architecture for TEADAL's final phase.

2.2 TEADAL Node Tools

As explained in detail in D2.4, the TEADAL Node and its tools creates a self-contained platform for trustworthy federated data sharing. Each node integrates key components such as the Catalogue, which manages metadata and policies for data products and enables federation-wide discovery; the Control Plane, which orchestrates the data product lifecycle through monitoring (AI-DPM), automation (ASG for sFDP generation), optimization (lightweight placement engine), and deployment (GitOps/ArgoCD); the Trust Plane, which ensures verifiable compliance and auditability via the Advocate for evidence collection and the TEADAL Name Service (TNS) for decentralized resource identification; a Service Mesh (Istio/Envoy) for enforcing security policies and managing service communications; and Data Pipelines, which process and transform data while often leveraging Trusted Execution Environments (TEEs) for privacy.

Together, these components support Federated Data Products (FDPs) and Shared FDPs (sFDPs), ensuring efficient, policy-compliant, and auditable data sharing across organizational boundaries.

Funded by the European Union

3. PILOT CASES DEPLOYMENT

This section describes the deployment of the pilot cases defined within the project, following the specifications outlined in earlier work packages. The pilots serve as practical demonstrations of how the TEADAL platform and its components can be configured, integrated, and validated in real-world scenarios. Each pilot leverages the designated testbeds to host and operate the required technical services, ensuring secure interconnection and alignment with the use-case requirements. The deployment activities are directly linked to the Key Performance Indicators (KPIs) established in T2.1, and their validation are carried out according to the plan defined in T6.2. Overall, this section provides an overview of the process, objectives, and expected outcomes of the pilot deployments.

3.1 Platform Updates Across the Project Lifecycle

During the second half of the project the Pilots have had significant improvements in terms of infrastructure readiness, component integration, and alignment with the TEADAL architecture. These updates reflect the evolution of both technical requirements and use-case maturity across the six pilot sites.
With the second half of the project, Pilots have been progressively enriched with additional software layers, pilot-specific services, monitoring and observability tools, and TEADAL-specific components such as the AI-DPM stack, the data catalog, policy enforcement modules, and pipeline sFDP automations. Several updates were also applied to improve deployment consistency across pilots, including the adoption of unified deployment templates (e.g. via Kustomize), the definition of a reference node architecture, and the clarification of namespace conventions (e.g. for ARGOCD, ISTIO, and default components) plus an automated procedure (script) to install a TEADAL baseline from scratch. Additionally, many pilot testbeds have been iteratively refined based on feedback from validation activities, integration troubleshooting, and updates in component versions. In some cases, nodes were reinstalled or reconfigured to address performance, compatibility, or architectural alignment issues. As a result, by the end of the project, all pilots reached a sufficient level of completeness to support the planned validation and experimentation activities. The following section provides a detailed breakdown of the components available on each pilot, along with a summary of CI/CD deployment mechanisms adopted.

Pilot Testbed Update

Since the publication of D6.1 at M15, the pilot testbeds have had updates, reported in this subsection. Figure 3 shows the final Pilot Testbeds. The updated pilot topology represents the natural evolution of the “Pilot infrastructure Overview” diagram presented in the Section 6.2 of the deliverable D6.1. consolidates the architectural progress and topology changes occurred during the second half of the project. For a historical view and additional context, the reference is the deliverable D6.1.

As an example to provide a clearer understanding of the topology, the Medicine Pilot setup is described below. It exemplifies the reuse of a single testbed site across multiple pilots. Ribera Salud is the resource provider owner of the Testbed Site Ribera, which hosts a virtualized infrastructure composed of four Virtual Machines (VMs). This testbed is used to deploy multiple TEADAL Nodes, each supporting different pilot activities.

- Three out of four VMs are allocated to the Evidence-Based Medicine pilot. These VMs are integrated with three hospital environments (Hospital A, B, C), and are used to validate clinical data flows and support the use-case logic defined by Ribera.
- The fourth VM has been made available to the Industry 4.0 pilot (coordinated by ERT), in order to demonstrate cross-site interaction between TEADAL Nodes. This

Funded by the European Union

setup enables integration between a node hosted at the Ribera site and external industrial plant (e.g. Portuguese and Czech plants), showcasing distributed data workflows cross-domain federation within the TEADAL architecture.

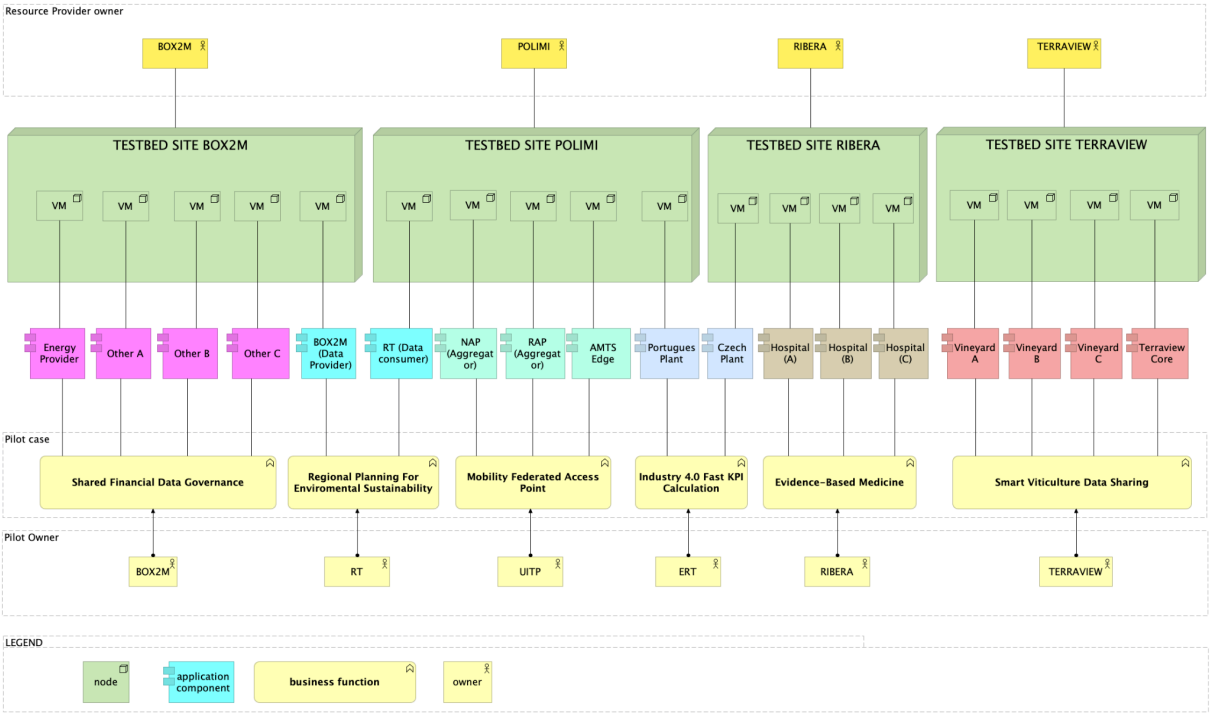


Figure 3 - TEADAL Testbed Architecture

With respect to D6.1, there is an update of pilot testbeds, considering that:

1. A new Financial Pilot from BOX2M replaced the ING Pilot,
2. Terraview Crates edge nodes were considered no longer part of Terraview testbed as the TEADAL software was totally installed on 3 Vineyards nodes.

3.2 Pilot Architectural Components

Each TEADAL Node integrates several components that support networking, data cataloging, data exposure, governance, monitoring, and trust. The Core Components of TEADAL are the following:

Networking and DNS Configuration

- Provides connectivity and discoverability of the node within the federation;
- Includes DNS settings, domain information, IP addresses, and port configurations.

GitLab Repository

- Stores code, configuration files, and documentation required for the pilot;
- Accessible via Git or similar repositories with appropriate access rights.

Data Catalogs

- Centralized Data Catalog: Registers datasets and services at federation level for global discoverability;
- Local Data Catalog: Maintains node-specific datasets and metadata, enabling local management and customization.

Data Exposure Layer

Funded by the European Union

- Datasets: Raw or processed data shared by the pilot node, described with metadata (format, update frequency, size);
- Federated Data Product (FDP): API-based service that makes datasets accessible in an interoperable manner;
- Shared Federated Data Product (sFDP): Extends FDP with governance features such as policies, security rules, and controlled access;
- Agreements: Define terms, conditions, and licenses under which data and services are shared.

User Roles and Access Control

- Ensures role-based access to datasets and services;
- Roles include: Data Owner, Data Consumer, Administrator, and additional pilot-specific roles.

TEADAL Platform Tools - Each node integrates a set of tools developed in TEADAL to enable automation, monitoring, policy enforcement, and trust management

- CI/CD Tools (e.g., GitLab CI, ArgoCD) - automated deployment and integration;
- Data Catalog - core FDP repository and metadata management;
- Policy Management Tool (OPA, Rego, KeyCloak) - access control and policy enforcement;
- Base Monitoring (Prometheus) - monitoring of system health;
- Monitoring Stack (Prometheus, Thanos, Grafana) - performance and observability across deployments;
- Control Plane - orchestration and service management For managing the common infrastructure and platform services, we rely on native Kubernetes controllers and on ArgoCD-based GitOps, described in detail in D6.2. For managing the TEADAL Data Products, additional tooling was developed, mainly the Automatic sFDP Generation (ASG) subsystem of the control plane, described in detail in D4.3. The ASG subsystem includes the ASG-Tool, responsible for generating sFDPs based on natural language descriptions of the data transformation pipelines, and the ASG-Runtime, a library responsible for providing an execution environment for all the generated sFDPs. While the ASG-Tool is provided in a standalone form, to be executed by data product developers (see the list below), the ASG-Runtime is deployed as a base image that all the sFDPs are based on. In addition, to support standalone execution of the ASG-Tool, we support deploying the Ollama inference service on each TEADAL node as one of the common platform services.
- Trust Plane - an evidence-based framework that establishes trust in federated data exchanges; Notable components include Advocate, the TEADAL Name Service and Federation Smart Contract, Catalog transaction Observer, Sharemind MPC, TEE Pipeline, ZK-SecreC, SynthGuard, and TrustOps, among others
- AI-DPM - ML-based forecasting and anomaly detection for data collected from monitoring tools;
- Energy-Related Tools (Kepler) - energy-aware tool to monitor energy status of the TEADAL Node;
- Others (Pilot-Specific) - optional extensions depending on pilot needs.

Standalone Tools

- ASG Tool: Generates the application code for sFDPs based on a custom definition of data endpoints that sFDPs will expose. For each endpoint, the description includes the pointer to the source FDP endpoint and the natural language description of how the FDP data has to be transformed before being returned to sFDP users. The result of the ASG-Tool invocation is a ready-to-test FastAPI application code.
- Policy Definition Tool: Assists in defining, using the Data Sharing Policy Notation proposed in TEADAL, and applying, by generating Rego rules, the access and usage policies.

Funded by the European Union

- Sharemind MPC: A platform for secure multi-party computation

Monitoring and Logging

- Provides visibility into node performance and data quality;
- Includes system/application logs and dashboards for monitoring uptime, response time, and data requests.

In addition to the TEADAL core components, some pilots required the installation of add-ons, i.e., auxiliary tools that act as enablers or dependencies for standardized deployment recipes to ensure consistency across sites. Some pilots, for example, required the deployment of “advanced” monitoring and energy profiling tools such as Prometheus with Thanos and Kepler, which were essential to support components like AI-DPM, Jager, and overall observability layer within the testbed.

Following the deployment of the core TEADAL Node, each pilot tailors its setup by selecting the components that best fit its use case. Some components are mandatory and ensure interoperability within the federation, whereas others are optional and activated only when required by the pilot’s objectives. This flexible approach explains why each pilot produces unique results and developments.

3.3 CI/CD as a Key Enabler

Among these components, CI/CD plays a central role, ensuring fast, reliable, and automated updates across all TEADAL Nodes. It integrates tightly into the platform to reduce manual intervention, guarantee reproducibility, and keep deployments aligned across different pilots.

The CI/CD mechanism is based on three tools: GitLab, ArgoCD, and Kubernetes (MicroK8s).

Continuous Integration (CI - GitLab)

- Developers commit code to the GitLab repository;
- A CI pipeline is triggered, which:
  - Check out the latest code;
  - Installs dependencies;
  - Runs automated tests;
  - Builds a container image;
  - Pushes the image to the registry.

Continuous Delivery/Deployment (CD - ArgoCD + Kubernetes)

- ArgoCD, following the GitOps model, monitors the GitLab repository for changes;
- When updates are detected, ArgoCD synchronises the Kubernetes cluster with the repository, ensuring the desired state matches the live state;
- Kubernetes then deploys, scales, and manages the updated containers, while continuously monitoring to keep the TEADAL node stable.

This mechanism ensures fast, reliable, and automated updates across all TEADAL nodes, reducing manual effort and downtime, while keeping deployments auditable and consistent.

3.4 Pilot Usage

After deploying the core TEADAL Node, each pilot tailors its configuration by selecting the components that best fit its specific use case. Some components (such as networking, repository, centralized catalog, FDP exposure, user roles, CI/CD, and base monitoring) are mandatory to ensure interoperability across the federation. Others (including the local catalog, sFDP, policy management, AI prediction, or energy-related tools) are optional and only activated when required by the pilot’s objectives.

Funded by the European Union

This modular approach allows each pilot to obtain distinct results and contributions, while still operating within a common TEADAL framework. The detailed description of the deployments and the components adopted in each pilot is provided in the subsequent sections of this deliverable.

Funded by the European Union

4. PILOT #1 - EVIDENCE-BASED MEDICINE

4.1 PILOT USE CASE

Shared multi-center research clinical trials. Single center trials using data from other centers & our own center's data.

The pilot in evidence-based medicine aims to enhance data analytics in healthcare by improving medical data sharing and analysis while addressing data privacy challenges. It focuses on obtaining consent from data subjects and working with anonymized data due to privacy restrictions.

RIBERA SALUD simulates federated data sharing among healthcare organizations, while TEADAL tools will help navigate privacy constraints by managing data access. This includes establishing trust between organizations when access requirements are met and providing tools to ensure data is used only from consented participants in medical research.

Our Pilot involves 3 different hospitals with the following unique datasets for every hospital:

- Condition Occurrence - Records the diagnosis of a disease or medical condition assigned to a patient (e.g., diabetes, hypertension).
- Drug Exposure - Documents the details of a patient's exposure to a prescribed or dispensed drug.
- Measurement - Captures the results of quantitative clinical tests (e.g., blood pressure, lab values like cholesterol levels).
- Observation - Captures clinical facts about a patient obtained in context of examinations, questions, or procedures, including symptoms and lifestyle factors.
- Person - The central table containing de-identified patient demographic information (e.g., year of birth, gender).
- Procedure Occurrence - Records actions or interventions performed on a patient for diagnostic or therapeutic purposes (e.g., surgeries, therapies).

Federated Data Products

- Each hospital exposes its own FDP facilitating easy access to their own data. One for every dataset.

Shared Federated Data Product

- A Federated User from hospital C promotes a SFDP to obtain patient data from Hospital A following defined policies and transformations.

List of FDPs endpoints:

Condition Occurrence Endpoints
GET /condition_occurrences
GET /condition_occurrences/ — Returns all condition occurrences.
GET /condition_occurrences/{condition_occurrence_id} — Returns a specific condition occurrence by ID.

Drug Exposure Endpoints
GET /drug_exposures
GET /drug_exposures/ — Returns all drug exposures.

Funded by the European Union

GET /drug_exposures/{drug_exposure_id} — Returns a specific drug exposure by ID.
GET /drug_exposures/person/{person_id} — Returns drug exposures related to a patient.

Measurement Endpoints
GET /measurements
GET /measurements/ — Returns all measurements.
GET /measurements/{measurement_id} — Returns a specific measurement by ID.

Observation Endpoints
GET /observations
GET /observations/ — Returns all observations.
GET /observations/{observation_id} — Returns a specific observation by ID.

Person Endpoints
GET /persons - Returns all patients
GET /persons/{person_id} — Returns a specific patient by ID.
GET /persons/age/up/{age} — Returns patients older than a given age.
GET /persons/age/low/{age} — Returns patients younger than a given age.
GET /persons/age/{age} — Returns patients of a specific age.

Procedure Occurrence Endpoints
GET /procedure_occurrences
GET /procedure_occurrences/ — Returns all procedure occurrences.
GET /procedure_occurrences/{procedure_occurrence_id} — Returns a specific procedure occurrence by ID.

List of SFDPs endpoints:

Service Endpoints

GET /service/settings — Returns application settings.
GET /service/stats — Returns runtime statistics.
GET /service/transforms — Returns available data transformations.
POST /service/origin_cache/clean — Clears the origin cache.
POST /service/response_cache/clean — Clears the response cache.

Data Endpoints

GET /persons — Returns all person records with IDs and birthdates.
GET /persons_with_age — Returns all person records with computed ages.
GET /persons_over_age/{min_age} — Returns persons older than a specified age.
GET /persons_under_age/{max_age} — Returns persons younger than a specified age.
GET /persons_between_ages — Returns persons between two specified ages.
GET /person_by_id/{person_id} — Returns a specific person by ID.

4.1.1. Applying TEADAL architecture

By making use of the available TEADAL tools, our case proves that clinical studies can be performed in a much more secure, regulated, agile and user friendly environment, using the latest data lake capacities serving to improve Quality of Life. By using a Federated Data Product along with a Secured Federated Data Product

Funded by the European Union

producing OMOP medical structured data for researchers across different distributed systems over different sites.

4.1.2. Final Architecture of the Pilot Use Case

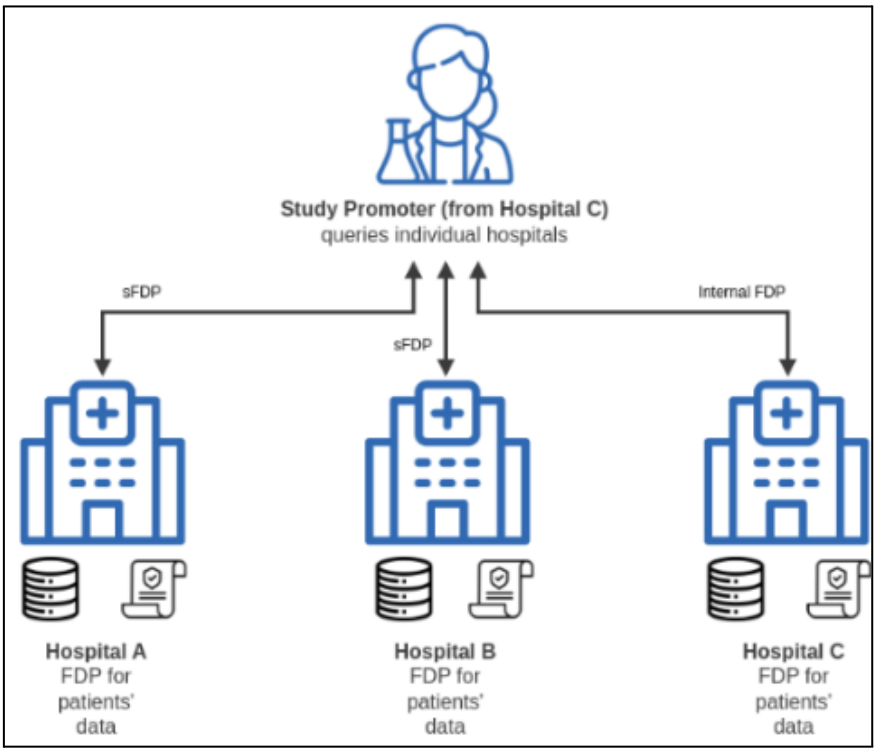


Figure 4a - Pilot #1 Architecture

4.2 TEADAL Components Involved

The pilot used the following TEADAL components:

1. **Catalogue,** Ribera Salud Hospitals provide a catalogue of products/services, containing the following datasets: Patients, Observations, Measurements, Procedures, Occurrence, and Drug Exposure

    - Expected Result: The tool delivers a list of products offered by the hospital.

2. **ArgoCD,** a declarative, GitOps continuous delivery tool for Kubernetes. It automates the deployment and lifecycle management of applications by synchronizing the live state in a cluster with a desired state defined in a Git repository. Used to deploy: FDP, sFDP, Catalogue, Advocate, Keykloack, Opa, Minlo, etc.

    - Expected Result: Automated Continuous Delivery: Fully automated deployment pipelines triggered by Git changes. Elimination of Configuration Drift: The cluster state is constantly reconciled with the Git source. Improved Deployment Reliability & Visibility: A single pane of glass for viewing application health and sync status across all environments.

3. **ASG tool,** transforms the data harmonization process from a manual, code-heavy task into a managed, semi-automated workflow. Used to generate the sFDP starting from FDP.

Funded by
the European Union

- Expected Result: A scalable and repeatable framework for generating sFDPs. This shift minimizes human error in data mapping, ensures regulatory-grade consistency across all data sources, and future-proofs our data pipeline against growing complexity and volume.

4. **Advocate,** collects evidence related to FDP/sFDP interactions, stores them in shared publicly accessible and immutable storage, and guarantees that the stored claims and evidence are tamper-proof.

- Expected Result: Provides trust by having evidence in interactions, derived from the claims services generate, with TEADAL FDP/sFDPs among healthcare organizations and patients. Allows to verify if data is used according to usage policies such as patient consent for medical research

5. **AI-DPM,** AI Time Series Prediction Service**,** enables monitoring of infrastructure resource use status, application service communications and energy consumption, by providing forecasting for predicting performance status and detecting anomalies across over 1,000+ metrics. In this pilot, AI-DPM was deployed to generate predictions on metrics such as CPU usage, memory saturation, and Prometheus-based service metrics across the hospital infrastructure.

- Expected Result: The goal was to support early detection of system anomalies and enable more efficient resource allocation across nodes hosting AI-based clinical services, thus improving reliability and reducing potential downtime.

6. **Jaeger,** provides end-to-end visibility into microservices by tracing the complete lifecycle of requests as they propagate across service boundaries.

- Expected Result: Pinpoints the root cause of performance issues and latency bottlenecks across complex, distributed systems. This leads to faster diagnosis of problems, improved system reliability, and more performant applications.

- 7. **Keycloak,** acts as a centralized security gateway for managing user identities and access permissions. Used to Create Roles: Administrator: Can access to all resources defined in API; Federated_User: Can access to the contracted resources.

  - Expected Result: Implements secure Single Sign-On (SSO) and centralized authorization, reducing security risks and operational overhead associated with managing multiple, separate login systems.

- 8. **MinIO,** delivers high-performance, S3-compatible object storage for cloud-native applications. Used to store these datasets: Condition_occurrence.csv, Drug_exposure.csv, Measurement.csv, Observation.csv, Person.csv, procedure_occurrence.csv.

  - Expected Result: Enables applications to store and retrieve massive amounts of unstructured data (like documents, images, and logs) with the same API used in public clouds. This provides a portable, scalable, and cost-effective storage layer on our own infrastructure.

- 9. **OPA,** provides a unified policy engine to define and enforce authorization and access control rules within the Kubernetes environment. In this case, it is used to configure and apply the policies defined in the ingress-policy.yaml file.

Funded by the European Union

- Expected Result: Ensures that access requests to cluster resources comply with the established security policies (for example, who can access, from where, and under which conditions). This adds an additional layer of governance and compliance, ensuring centralized and consistent control over authorization decisions.

TEADAL add-ons:

1. **Prometheus with Thanos,** metrics collection and long-term storage through Thanos federation.

   - Expected Result: Enable observability, monitoring of TEADAL components, and provides metric sources required by AI-DPM and tracing tools.

2. **Kepler,** energy consumption profiling for Kubernetes workloads, providing node-level and pod-level energy metrics.

   - Expected Result: Enables energy-aware analytics and provides additional metrics that can be consumed by AI-DPM or stored in Prometheus/Thanos for evaluation and benchmarking.

Note that, as discussed for KPI 2.3, we use the medical pilot as an example to demonstrate privacy-preserving data analytics using **Sharemind MPC**, however, this demonstration is not integrated with the main pilot deployment.

4.3 Final Results

The Ribera Salud pilot successfully established a secure federated data infrastructure for multicentre clinical trials. This infrastructure enabled privacy-preserving analytics across organisations by automating the creation of standardised, anonymised data products. The project demonstrated enhanced trust by providing verifiable audit trails of data usage to ensure compliance with patient consent. It significantly improved operational efficiency, accelerating data discovery and provisioning for researchers and reducing the bureaucratic burden. The integration of proactive monitoring and energy profiling tools also ensured the platform would be reliable, sustainable and scalable for the future of medical research.

Hospital C consumes sFDP GET /persons/age/up/{age} — and graphically displays clinical trials data through a dedicated dashboard as shown in the following image

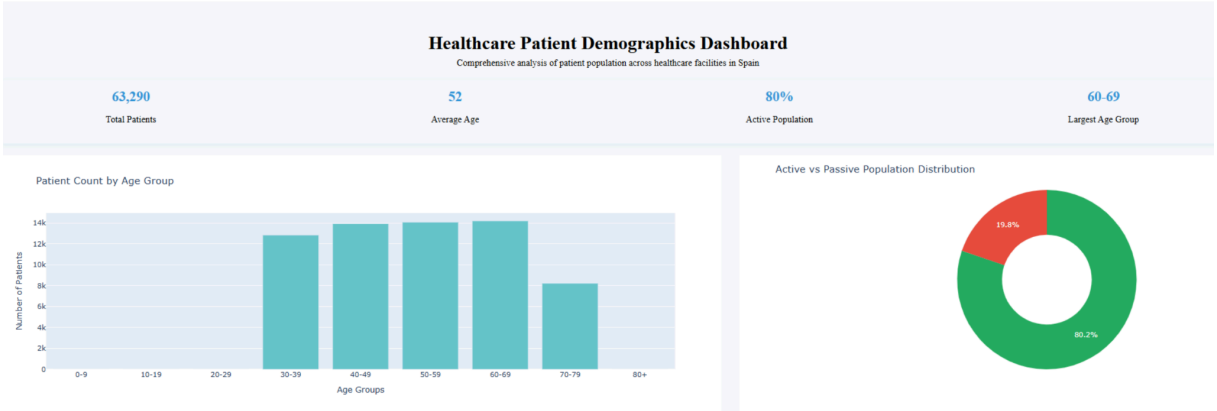


Figure 4b - Pilot #1 Hospital C dashboard

This validates TEADAL as a robust framework for overcoming the challenges of data lakes in healthcare. By enabling collaborative research without compromising security or sovereignty,

Funded by the European Union

TEADAL paves the way for accelerated medical breakthroughs. The project represents a significant step towards a new paradigm for trusted, efficient data sharing in medicine.

Funded by
the European Union

5. PILOT #2 - MOBILITY FEDERATED ACCESS POINT

5.1 PILOT USE CASE

The mobility pilot showcases data sharing among four Italian public transport stakeholders using TEADAL technologies and components. FDPs have been created and deployed to connect and combine different datasets according to EU policies related to open data in public transportation. Nodes store data related to: a local transportation operator (AMTS, Public Transport operator of the city of Catania, Italy), national transport operator (open data from Trenitalia, national railway company), Regional Access Point (RAP, tasked to aggregate data of all regional transport operators), and National Access Point (NAP, tasked to aggregate data of all national transport operators). The pilot simulates these components to showcase the capabilities of the TEADAL tools. The following picture shows the scheme of FDPs and related datasets made available through them:

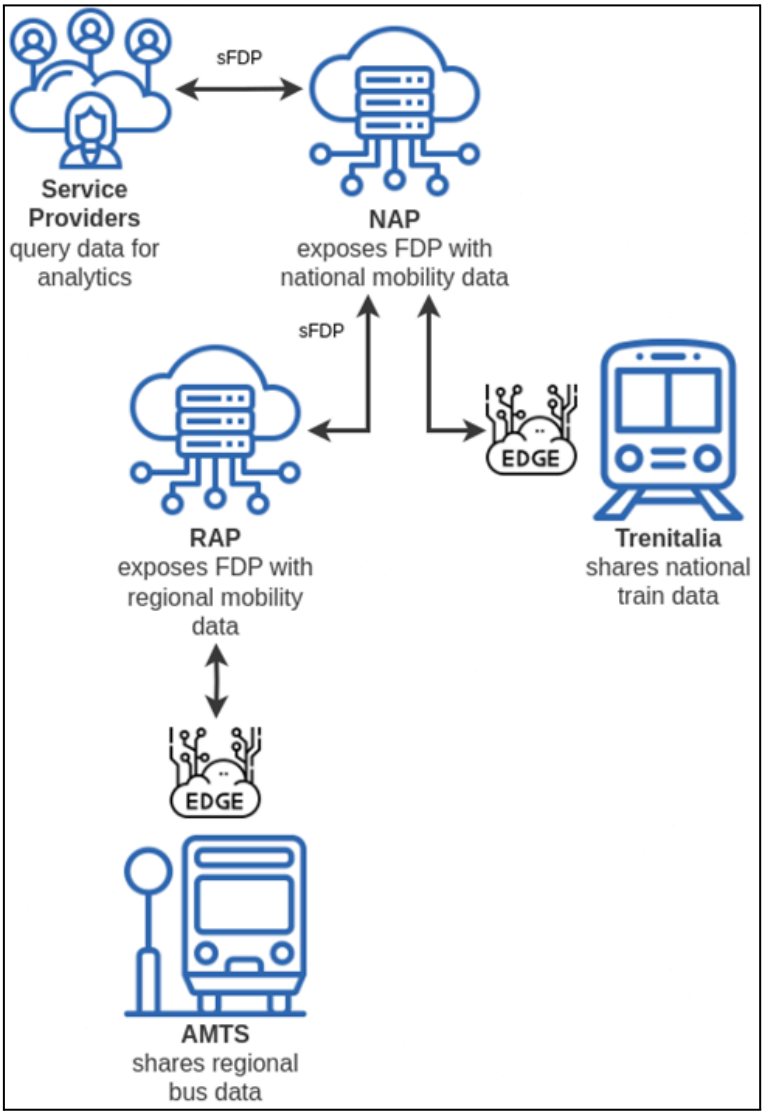


Figure 5 - Pilot #2 Architecture

The TEADAL Mobility Pilot demonstrates how federated data products (FDPs) can enable seamless access, sharing, and governance of transport data across multiple levels: local operators, regional and national aggregations. By combining static schedules, real-time

Funded by the European Union

updates, and multimodal connections, the pilot showcases how TEADAL technologies support interoperability, policy enforcement, and energy-efficient data exchange. FDP have been created to expose relevant transport data as follows:

1. Local edge node – AMTS (Catania Public Transport)

At the local level, the AMTS node exposes both static GTFS datasets (routes, trips, calendars, agencies) and real-time GTFS-RT feeds (trip updates, vehicle positions).

- The static APIs provide the planned structure of services: which routes exist, when trips are scheduled, and which agencies operate them.
- The real-time APIs extend this with live operational data, allowing consumers to query the current status of trips, delays, and the position of vehicles in service.

This dual layer, represented in Figure 6, ensures that applications can combine planned schedules with actual performance, supporting journey planning, passenger information, and operational monitoring.

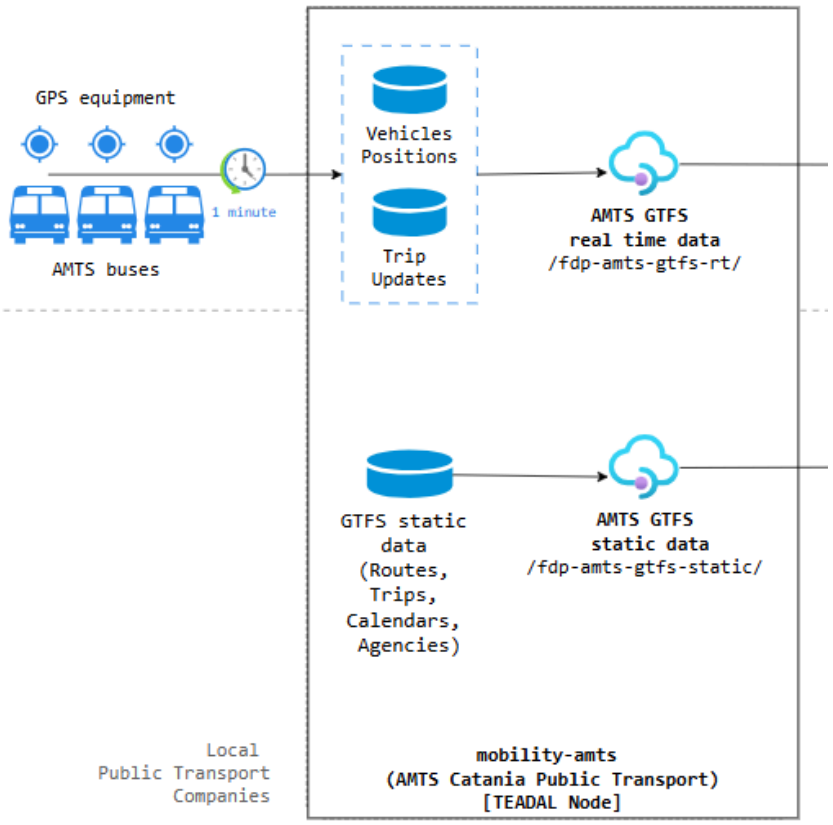


Figure 6 - Local AMTS FDP

2. Regional FDPs – RAP (Regional Access Point)

The RAP FDPs scale this model to the regional level by aggregating data from multiple transport providers.

- The RAP static node consolidates GTFS datasets from local FDPs of the region (AMTS and potentially other local transport providers), exposing harmonized APIs for routes and trips. This allows regional planners and service providers to query a single interface instead of multiple local operators.

Funded by the European Union

- The RAP real-time node aggregates GTFS-RT feeds, providing a unified view of live operations across the territory. Consumers can query trips by ID, date, or route, and receive consistent updates on delays and service status.

By federating data at the regional level, the RAP ensures consistency, scalability, and interoperability, while reducing fragmentation between local operators. Data has been made available to the NAP node through two sFDPs (staitc and real time). The RAP Node is represented in Figure 7.

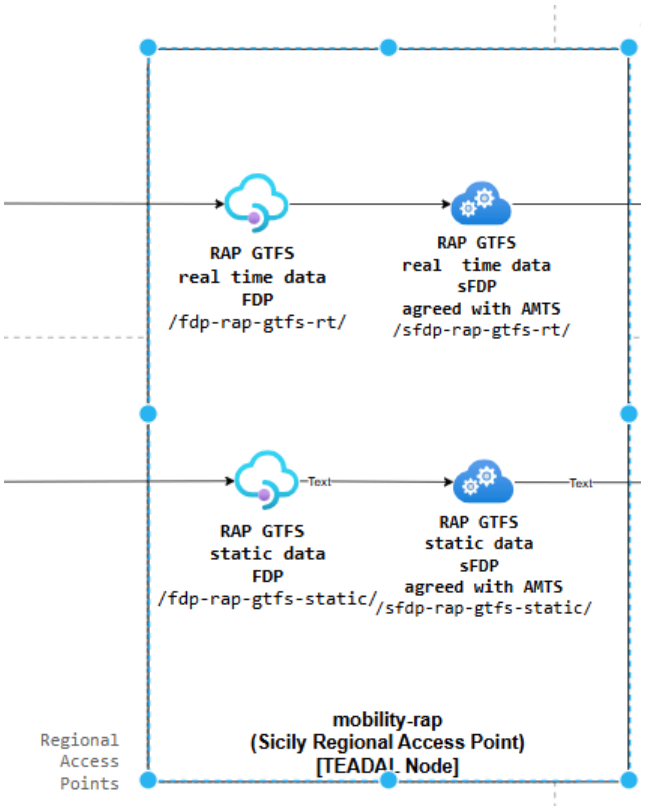


Figure 7 - RAP FDP

3. National FDPs – NAP (National Access Point)

At the national level, the NAP FDPs integrate both static and real-time datasets from FDPs of all regions (RAP) in Italy. These nodes mirror the structure of local and regional FDPs, but at a broader scale, ensuring that data from multiple regions can be accessed through a single national interface. The NAP plays a crucial role in aligning with EU transport regulations, acting as the mandated access point for open mobility data. It ensures that datasets are discoverable, interoperable, and governed by clear access policies, while also enabling cross-regional and cross-modal analytics. Also in this node two sFDPs have been created to make available static and real time data.

4. Trenitalia FDP – National Rail Timetables

Complementing the multimodal ecosystem, the Trenitalia FDP provides national rail timetables relevant to Catania and its connections. Unlike the GTFS-based APIs of local and regional nodes, this FDP introduces a tailored format for rail services:

- endpoints expose destination stations reachable from Catania,
- upcoming departures within a configurable time window,
- bus connections linking train stations with local public transport.

Funded by
the European Union

Data has been made available through one sFDP.

This multimodal integration is particularly significant: it bridges long-distance rail with local bus services, enabling door-to-door journey planning and supporting the vision of Mobility-as-a-Service (MaaS).

5. End user / Service provider

A generic end user, that may be a service provider for passengers, consumes data, according to policies allowing access to a specific set of data, through sFDPs generated by the NAP node. Here endpoints allow to consume data related to bus timetables (static) and positions (real time) as well as trains origins, destinations and connections to the selected city (in our pilot Catania). For the purpose of the pilot, a dedicated dashboard has been developed to show data consumer capabilities and to provide a user interface with data elaborations using data retrieved from all FDPs. By combining different datasets, the dashboard is able to provide multimodal trip planning, demonstrating scalability and facilitating the integration of multiple transport providers.

NAP FDP, Trenitalia FDP and the Dashboard of the Service provider are represented in Figure 8.

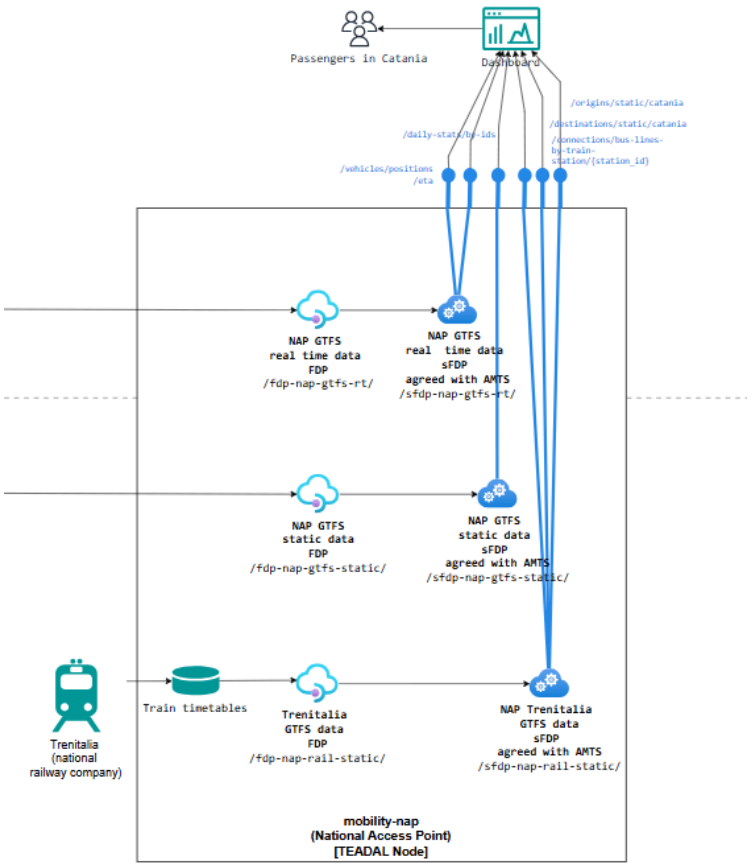


Figure 8 - NAP FDP, Trenitalia FDP and Dashboard

Datasets collected and managed in the FDPs described above are the following:

Dataset 1: GTFS Open Data From AMTS Catania

- Description: AMTS Catania public transport timetables

Funded by the European Union

- Format: GTFS
- Size: 2-3 MB
- Update Frequency: daily
- Access Method: download

Dataset 2: AMTS Real Time Trips Data

- Description: AMTS real time trips data
- Format: GTFS-RT
- Size: ~20 KB
- Update Frequency: Once every minute
- Access Method: download

Dataset 3: AMTS Real Time Bus Position Data

- Description: Real time positions of AMTS
- Format: GTFS-RT
- Size: ~20 KB
- Update Frequency: Once every minute
- Access Method: download

Dataset 4: Trenitalia timetables

- Description: Trenitalia timetables in Json format
- Format: proprietary
- Size: ~20 KB
- Update Frequency: manually
- Access Method: download

Additional details about FDPs and datasets have been made available for further researches in the Teadal GitLab repository.

5.2 TEADAL Components Involved

The pilot used the following TEADAL components:

1. **Catalogue**, used to organize and manage the Nodes's data available from two different sources (AMTS as public transport operator providing both static and real time data in GTFS format, Trenitalia as rail operator providing train timetables in a specific json format) and provide metadata to make them easily accessible for service developments

    - Expected Results: the tool delivers a comprehensive list of products offered by each node and supports scalability towards future scenarios in which individual providers can expose their data at the regional level. In turn, regional datasets can be seamlessly aggregated at the national level. This architecture enhances both data provision and utilization across the ecosystem, improving visibility into transport performance throughout the broader network.

2. **ArgoCD**, used as Declarative, GitOps continuous delivery tool for Kubernetes and has been used to deploy and manage the full TEADAL stack for the Mobility Pilot

    - Expected Results: this component ensures that the pilot deployment pipeline is fully automated and version-controlled, reducing manual configuration

Funded by the European Union

errors. Compared to a non-Teadal setup, this approach ensures faster iteration cycles and greater reliability of software delivery across nodes and provides an easy way for other transport companies to join the shared transport data ecosystem as well as for regional and national institutions to collect as many datasets possible to have a complete picture of transport information.

3. **Ollama AI Server**, on-premise server to access LLMs. The Ollama AI Server functions as the intelligence layer within the sFDP generation workflow. It has been used to be invoked by the ASG, it interprets the natural-language descriptions provided by the user in each property definition and selects or formulates the appropriate data transformation functions automatically.

   - Expected Results: by allowing users to express their intent in natural text rather than technical syntax, the Ollama integration removes the need for manual coding or expert intervention. This enhances usability and reduces configuration time, offering a clear advantage over non-Teadal environments where such transformations must be explicitly defined and maintained by developers. As the previous ones, this tool facilitates the deployments of additional transport nodes and the expansion of a national-wide network of transport datasets.

4. **AI-DPM – AI Time Series Prediction Service,** used to enable monitoring of infrastructure resource use status, application service communications and energy consumption, by providing forecasting for predicting performance status and detecting anomalies across over 1.000 metrics. In the Mobility pilot, AI-DPM was used to analyze time-series metrics related to vehicle telemetry processing, queue lengths in edge nodes, and inference response times.

   - Expected Result: the tool was evaluated for its ability to anticipate overload conditions and improve service delivery in high-frequency data environments (e.g. real-time camera feeds or sensor streams).

5. **ASG tool** used to automate the generation of sFDPs by interpreting the configuration file and dynamically identifying the correct data transformation functions through interaction with the Ollama AI Server. Specifically, for each property in the dataset schema, the ASG Tool uses the Description field to query the Ollama server, which analyzes the text and determines the appropriate transformation logic to apply based on user input and context.

   - Expected Results: this process reduces manual mapping and coding effort, ensuring consistency, accuracy, and faster onboarding of new data sources. Compared to a non-Teadal setup, this approach enables semi-automated data harmonization and reduces the technical expertise required to build or extend data services, finally contributing to the scalability of the overall Mobility ecosystem as reported for the previous tools.

6. **Policy Definition Tool** used to automatically derive access-control policies from the metadata defined in the OpenAPI specification of the generated sFDP. It analyzes the declared endpoints, parameters, and data models to produce Rego policy files, which are then deployed to the Open Policy Agent (OPA). These policies govern fine-grained security and authorization across the sFDP layer, ensuring that only authorized users and services can access specific datasets or operations. In the Mobility pilot, policies govern the data access permissions for specific categories of users: generic data consumers that can only access data/methods without

Funded by the European Union

restrictions, and federated users, having agreements in place with data owners, that can access the full datasets.

- Expected Results: this automation removes the need for manual policy definition, reducing configuration errors and improving compliance with security and governance requirements.

5.3 Final Results

The Pilot demonstrates how to deploy a scalable and open data infrastructure able to store and provide transport data at national level, complying with EU related transport regulations and facilitating data transmission between operators and consumers. Main benefits of this architecture rely on:

Federation and Governance - Across all nodes, TEADAL technologies ensure that data is exposed responsibly, efficiently, and transparently through standardized APIs, with policy-based access control and energy-aware optimizations.

- Federated users can access datasets across nodes, subject to agreements and restrictions.
- Generic consumers may access open data, with limitations applied to sensitive or restricted feeds.
- AI-based modules optimize data replication between RAP and NAP, reducing unnecessary transfers and minimizing energy consumption.

Value of the Pilot - By federating data from local buses, regional aggregators, and national services, the TEADAL Mobility Pilot demonstrates the practical benefits of data spaces in transport:

- For passengers: more accurate and complete journey planning, real-time updates, and multimodal connections.
- For operators: tools to monitor punctuality, efficiency, and service quality.
- For policymakers: harmonized datasets that support regulation, innovation, and sustainable mobility strategies.

Using data provided through the sFDP exposed by the NAP, the deployed dashboard (of which some screenshots are shown below) offers to the user journey planning services as well as statistics of service reliability, improving the user experience and facilitating the accessibility to public transport serivces.

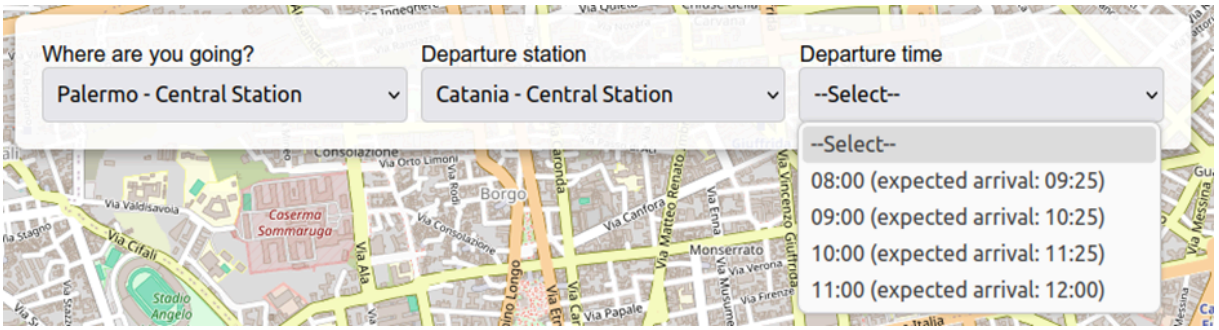


Figure 9 - Destination and departure selection

Funded by the European Union

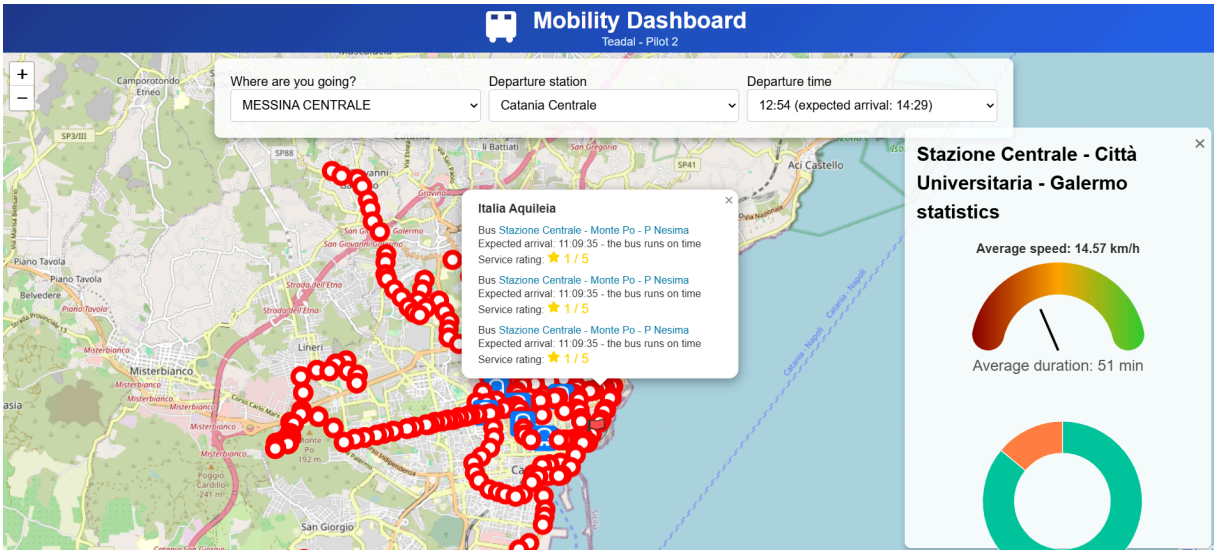


Figure 10 - transport statistics

Summarising, the Mobility Pilot used Teadal technologies to create:

- 5 sFDPs (2 in the RAP, 2 in the NAP, 1 for Tranitalia dataset)
- 3 distinct sFDP (considering that those in the RAP are replicated in the NAP with similar structure)
- 6 endpoints (static bus data, dynamic bus positions, origin, destination and connections for trains)
- 1 Pilot Application (the dashboard)

Funded by the European Union

6. PILOT #3 - SMART VITICULTURE DATA SHARING

6.1 PILOT USE CASE

The viticulture pilot demonstrates how multiple vine growers (edge nodes) can share soil moisture data in a federated manner while maintaining sovereignty at each grower's node. A central node aggregates selected datasets for joint analysis and benchmarking. The pilot validates TEADAL's ability to support precision agriculture through trustworthy and energy-aware data sharing.

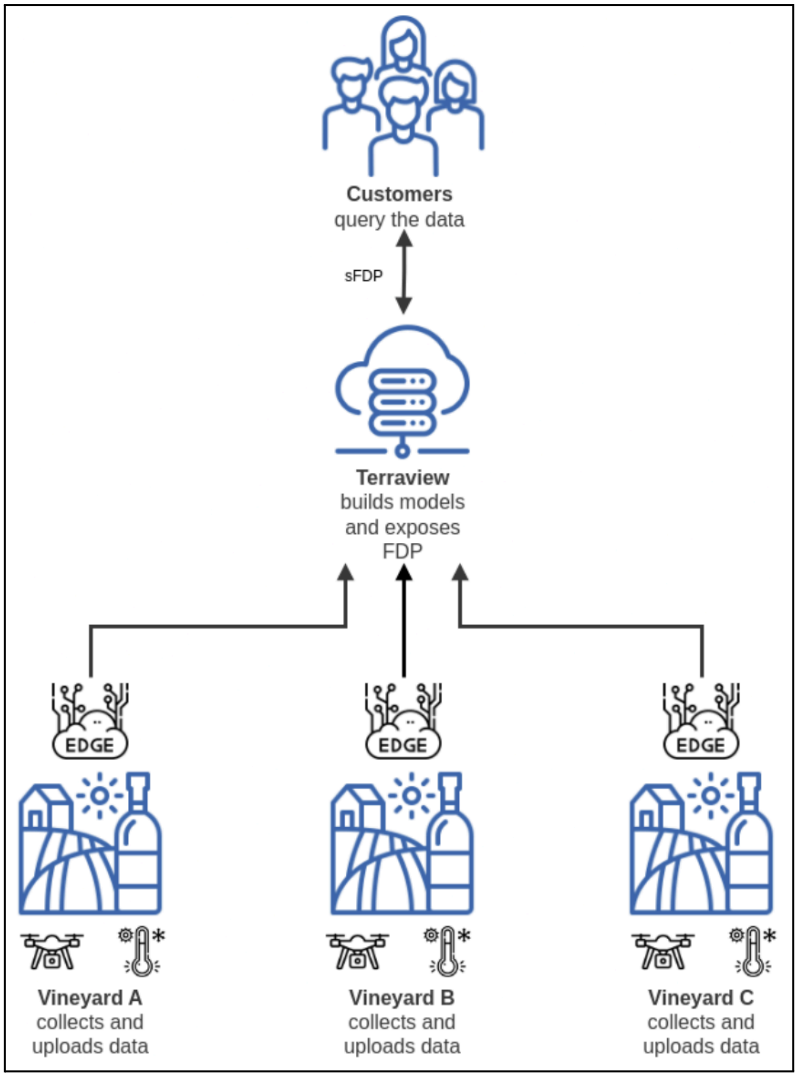


Figure 11 - Pilot #3 Architecture

6.2 TEADAL COMPONENTS INVOLVED

The pilot uses the following TEADAL components:

1. **ArgoCD,** used to deploy and manage the full TEADAL stack (Keycloak, MinIO, Advocate, Catalogue, ASG) across both the central and edge nodes. It ensures reproducible and synchronised deployments of all TEADAL services across multiple growers.

Funded by the European Union

- Expected Results: Streamlined and automated deployment reduces manual configuration, increases reliability, and allows faster onboarding of new grower nodes.

2. **MinIO,** serves as local object storage for each grower's soil-moisture datasets and as centralized storage at the admin node for aggregated data used in analytics.

- Expected Results: Enables scalable, secure, and efficient management of distributed datasets, facilitating aggregation for the admin analytics role without data duplication.

3. **Keycloak**, configured with distinct roles (admin, user) and corresponding organizational identities. It governs user authentication and authorization across the federation.

- Expected Results: Provides fine-grained access control ensuring that only the admin analytics role can access aggregated datasets, maintaining grower sovereignty and data protection.

4. **TEADAL Catalogue**, used by growers to document local soil-moisture datasets and publish them as FDPs.

- Dataset: 3 soil moisture datasets are provided, each one representing a single farm organization.
  - Dataset 1:
    - Description: Soil moisture data from organization "auth0|6540c4d9259d359f62903750"
    - Format: CSV
    - Size: 50 MB
    - Access Method: download
  - Dataset 2:
    - Description: Soil moisture data from organization "dfe756bf-cc4f-58a3-acc6-e15b2c7a3fa8"
    - Format: CSV
    - Size: 1.9 MB
    - Access Method: download
  - Dataset 3:
    - Description: Soil moisture data from organization "51fda6bb-63c2-4ec1-8dd6-48d2071eb6dc"
    - Format: CSV
    - Size: 1 MB
    - Access Method: download
- FDPs
  - Vineyard Core - aggregates information from the available edge nodes.
  - Vineyard Edge A - manages soil moisture data for organization in dataset 1.

Funded by the European Union

- Vineyard Edge B - manages soil moisture data for organization in dataset 2.
- Vineyard Edge C - manages soil moisture data for organization in dataset 3

- Expected Results: Ensures metadata consistency, discoverability, and streamlined data product management across multiple growers.

5. **Policy Definition Tool**, used to define sharing policies allowing each grower to expose only their own soil-moisture values that can be all exposed to the admin node roles.

- Expected Results: Ensures compliance with privacy constraints, preventing unauthorized disclosure.

6. **ASG (Automated Service Generator)**, used to create the sFDP combining the central node's FDP, aligned with the policies defined above, and automatically deploying it on the central node for aggregated analytics.

- Expected Results: Enables automated, policy-compliant creation of analytics-ready data products, reducing integration and validation time.

7. **Advocate,** collects evidence in the form of attested claims for interactions when sharing aggregation results with external customers

- Expected Results: Provides verifiable audit trails, improving accountability and trust between growers and the admin analytics operator.

Together, these components enable secure and controlled data sharing across growers and the central node.

6.3 Final Results

The pilot successfully deployed three edge nodes and one central node, demonstrating the end-to-end workflow from local data collection to federated analytics. The integration of TEADAL components resulted in:

- Secure and traceable data sharing with full auditability through the Advocate module.
- Automated deployment and synchronisation across all nodes, reducing integration time.
- Role-based access control that strictly separates grower and admin analytics permissions.

Compared to a non-TEADAL scenario, where data would need to be centralised or manually exchanged, the pilot showed clear efficiency and trust advantages:

- Growers retained full data ownership and could participate in collaborative analytics without exposing raw measurements.
- Deployment and policy enforcement were automated, reducing configuration overhead.
- The system improved transparency, traceability, and compliance with FAIR and GDPR principles.

Funded by the European Union

The pilot highlights TEADAL's strong potential for agricultural and environmental data ecosystems, where sensitive, distributed datasets can be analysed collectively to support sustainability and productivity goals.

Funded by the European Union

7. PILOT #4 - INDUSTRY 4.0 FAST KPI CALCULATION

7.1 PILOT USE CASE

The industry pilot will be focused on the necessity of computing a batch of key performance indicators (KPIs) that are shared between two ERT Group facilities located in different countries (Portugal and Czech Republic). Given that data is typically collated based on each facility separately, the KPIs for the entire corporation must be calculated and compiled according to the unified standard of the group as a whole. The objective of TEADAL is to enhance and automate the process of calculating KPIs that are pertinent to the company's management-related aspects (operational, commercial, quality, etc.). TEADAL will offer tools to automate and fine-tune the tech-impact related to data sharing within the ERT Group. An illustration of the main participants and data of the pilot is given in Figure 12.

The final demo of this pilot is an interface accessible via web to graphically present the KPIs. The access and information viewed on the site should be controlled based on access policies. After entering, the user can access the KPIs based on consolidated data from both plants. The focus of the demo will be on security and access control, with the Portuguese plant functioning both as a plant and as the headquarters. The demonstration will centre around a dashboard, with at least two users accessing it to view different sets of data. The headquarters manager will have full access, while plant managers will be restricted to viewing data only from their respective plants. Overall, the demo requirements have remained consistent with those documented in the previous deliverables.

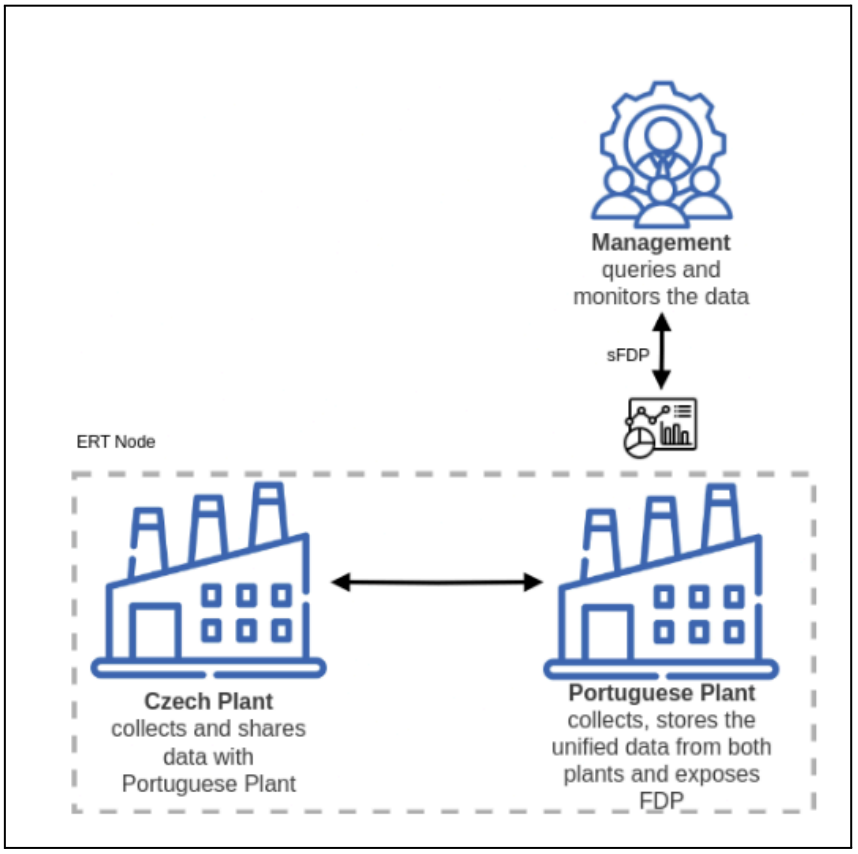


Figure 12 - Pilot #4 Architecture

Funded by
the European Union

The industry 4.0 pilot was deployed across two TEADAL nodes - one at the Portuguese plant (acting also as headquarters) and another at the Czech plant. Both nodes share a common architecture and exchange KPI-related data through FDPs and sFDPs.

- Datasets

  Each plant node maintains local datasets:

  - Forecasts dataset - Data from open customer orders at the moment
  - Productions dataset - Productions and movements data of the last year
  - Receptions dataset - Purchase receptions data of the last year
  - Sales dataset - Sales data of the last year
  - Qualities dataset - Source of current stock with what was rejected on production due to quality issues
- Federated Data Products (FDPs)

  Each plant exposes one FDP that provides standardized access to its local KPI-related datasets.

- Shared Federated Data Products (sFDP)

  The Portuguese headquarters node generates a corporate-level sFDP that consolidates data from both plant FDPs based on transformations and policies defined.

- Dashboard

  A specific KPI Dashboard is deployed on the Grafana platform based on sFDP generated by Portuguese Teadal Node (headquarters). The web interface displays cross-plant comparisons of top expected material requests, production timeline, biggest customers, top reception materials and scrap rates, based on the two endpoints generated (one for each plant).

7.2 TEADAL Components Involved

The following TEADAL components were deployed and validated within the Pilot #4:

1. **Advocate,** tool that allows to collect evidence for interactions related to sharing plant-specific KPIs between the nodes and company-wide KPIs

   - Expected Results: Provides trust and evidence in interactions with shared KPIs from individual plants and company-wide KPIs

2. **Catalogue,** registers plant-level datasets and exposes standardized data for KPI-related FDPs (production, quality, sales,...). It enables discoverability and governance of datasets across TEADAL Nodes, ensuring consistent data descriptions and easier integration across sites.

   - Expected Results: Enables discoverability and semantic interoperability of datasets across ERT plants, supporting automated generation of sFDP through the ASG tool.

3. **ArgoCD,** manages automated deployment and synchronization of TEADAL services on each plant node using GitOps workflows.

Funded by the European Union

- Expected Results: Guarantees consistent configuration and rapid updates across the federated plants, minimizing manual intervention and configuration drift.

4. **ASG tool,** automates the creation of sFDP representing aggregated KPI metrics. sFDP defines the transformation logic that converts raw plant data (e.g., production orders, quality inspection results, or machine logs) into normalized KPI datasets.

- Expected Results: Provides reproducible and auditable transformations, reducing human error and ensuring uniform KPI definitions across sites.

7.3 Final Results

The integration of the TEADAL Node enabled the aggregation of real datasets from different company plants. Sales, production, logistics, and quality data can now be jointly analysed, something that was not possible before.

A dashboard was created to demonstrate access to information provided by the sFDP, comparing data from both plants (Czech and Portugal):

Figure 13 - Pilot #4 Results Dashboard

The Industry 4.0 pilot successfully demonstrated how TEADAL technologies can transform traditional KPI management in manufacturing environments. Key results include:

- Federated KPI computation: KPIs from Portuguese and Czech plants were automatically aggregated and consolidated at the corporate level through sFDP, reducing manual data integration efforts and manual insertion errors.

Funded by the European Union

- Trust and compliance: The Advocate component provided verifiable evidence of data access and KPI updates, ensuring that all exchanges between plants were traceable and policy-compliant.
- Automation and reproducibility: The adoption of GitOps-based CI/CD pipelines (ArgoCD) and automated sFDP generation (ASG) ensured rapid deployment and repeatability of KPI pipelines across all nodes.

In conclusion, the Industry 4.0 pilot validates the TEADAL framework as a powerful enabler for trustworthy, federated, and energy-efficient industrial analytics. It proves that corporate KPIs can be computed across distributed plants without compromising data sovereignty or increasing operational complexity, paving the way for broader adoption of federated data spaces in the manufacturing domain.

Funded by
the European Union

8. PILOT #6 - REGIONAL PLANNING FOR ENVIRONMENTAL SUSTAINABILITY

8.1 PILOT USE CASE

The aim of this pilot is to integrate sensor data collected by a private enterprise - related to environmental conditions and energy consumption in buildings - with building energy profile data managed by public authorities.

This pilot involves two main partners: BOX2M, a private company, and RT, a public authority representing the Tuscany Region in Italy. The primary objective is to support the reconstruction of both static and dynamic energy profiles for public and private buildings, while also identifying local trends in energy efficiency and air quality. The analysis incorporates open data on weather conditions and air quality to enhance the accuracy and contextual understanding. One of the key goals is to enable RT to assess whether energy certification documents align with actual energy consumption data. The final demonstration will feature a data visualization dashboard that integrates information from static data sources managed by RT and dynamic data retrieved via API calls from BOX2M. The dashboard will present aggregated data in a user-friendly format and will be accessible to multiple users with uniform access rights.

Final Architecture of the Pilot Use Case:

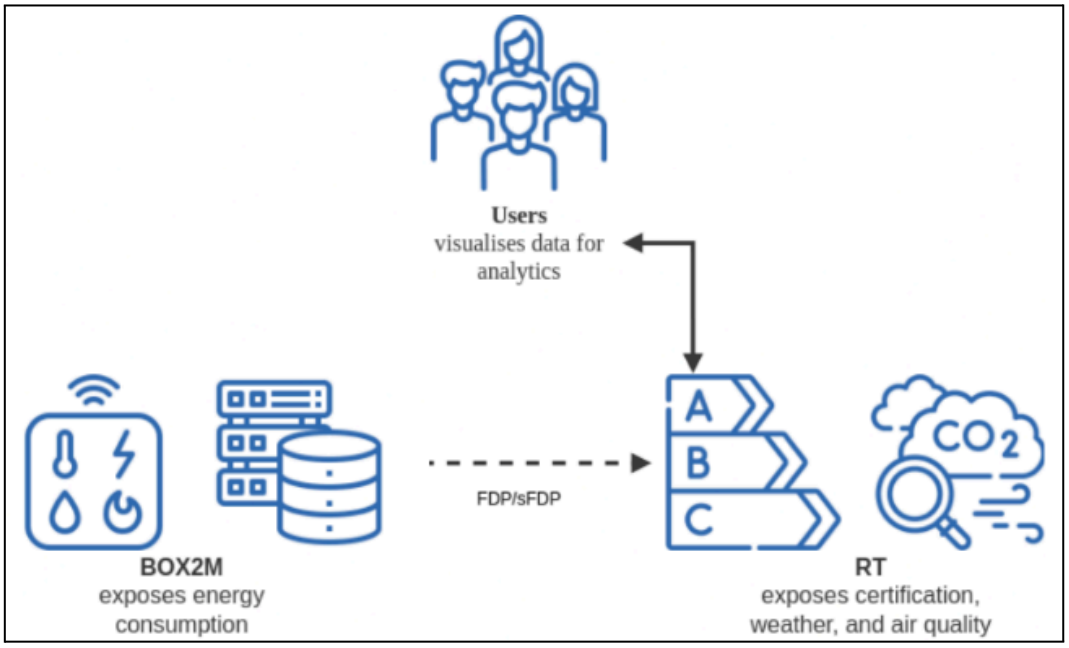


Figure 14 - Pilot #6 Architecture

- Node RT:
  - DNS name: smartcities-government
  - IP address: 108.141.70.100
  - RT Teadal node repo available on the internal gitlab server.
- Node BOX2M:
  - DNS name: smartcities-devices
  - IP address: 131.175.120.210
  - BOX2M Teadal node repo available on the internal gitlab server.

Funded by the European Union

8.2 TEADAL Components Involved

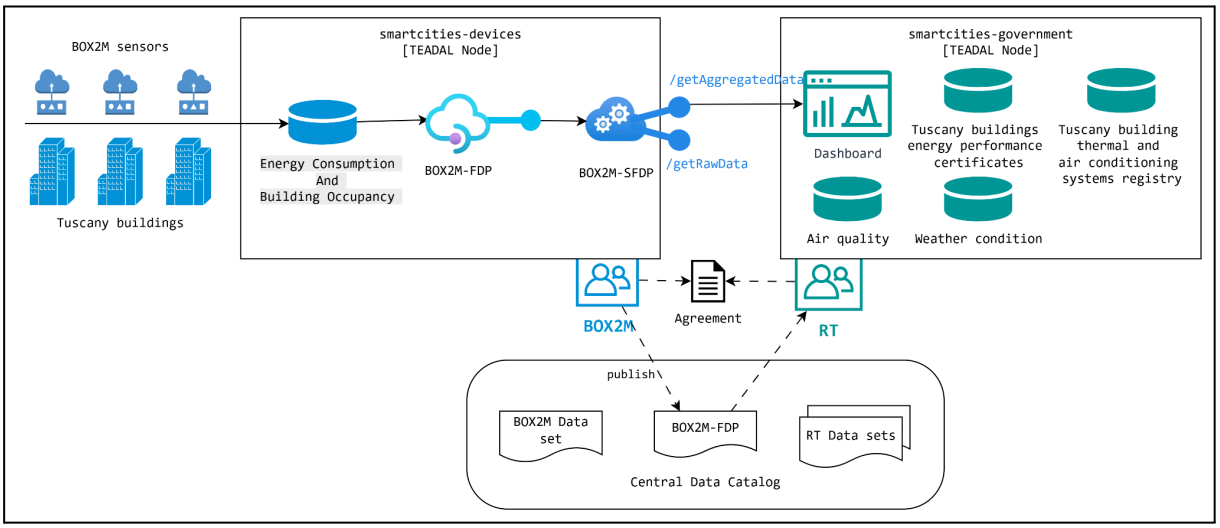


Figure 15 - Pilot #6 Software Architecture

The Data Catalog enables discoverability and governance of datasets across TEADAL Nodes, ensuring consistent data descriptions. Datasets belonging to Regione Toscana, Servizio Idrologico Regione Toscana and ARPAT Regione Toscana as well

Dataset 1: Tuscany buildings energy performance certificates

Description: Building's energy performance certificates (APE). This dataset contains personal information (but not sensitive data) about building or plant owners and about accredited technicians or certifiers (demographic information such as the name).

- Format: XML
- Size: 1M
- Update Frequency: daily
- Access Method: download

Dataset 2: Air Quality

Description: This dataset contains Open data about air-quality (e.g., pm10 concentration).

- Format: XML or CSV
- Size: 20M
- Update Frequency: daily
- Access Method: download

Dataset 3: Tuscany building thermal and air conditioning systems registry

Description: This dataset contains personal information (but not sensitive data) about building or plant owners and about accredited technicians or certifiers (demographic information such as the name).

Funded by
the European Union

- Format: Relational DB
- Size: 7M
- Update Frequency: daily
- Access Method: download

Dataset 4: SIR weather condition data

Description: This dataset contains Open data about weather conditions (temperatures, rainfall…) .

- Format: XML or CSV
- Size: 20M
- Update Frequency: daily
- Access Method: download

Dataset 5 BOX2M Energy Consumption And Building Occupancy Dataset:

Description: This dataset contains real time data from sensor about energy Consumption And Building Occupancy

- Format: JSON
- Size: 400K
- Update Frequency: stream continuos
- Access Method: download

BOX2M has exposed its own data by implementing an FDP:

Description: the FDP sources real time data from simulated sensors across 100+ regions in Regione Toscana. We have focused deeply to ensure the FDP sources both Raw and aggregated data and the simulation was inferred from realtime Lovato DMG110 sensors covering the 5 parameters from Figure 16.

Channel	Unit	Name	Tag	Type	Multiplier	Index
8	A	Current L1		Current L1	0.0001	✗
10	A	Current L2		Current L2	0.0001	✗
12	A	Current L3		Current L3	0.0001	✗
58	KW	Total Active Power		Total Active Power	1E-05	✗
6688	KWh	Total Active energy import		Total Active Energy Import	0.1	✓

Figure 16 - Pilot #6 Sensors Cover Parameters

AI-DPM enables monitoring of infrastructure resource use status, application service communications and energy consumption, by providing forecasting for predicting performance status and detecting anomalies across over 1,000 metrics. In the Regional Planning pilot, AI-DPM was used experimentally to monitor data ingestion rates, pipeline runtimes, and metadata service interactions.

Funded by the European Union

Expected Result: Forecasting and anomaly detection models helped identify disruptions during batch data updates and supported hypothetical evaluations of system scalability under heavy policy query load.

Relevant policies were established to regulate data sharing and security:

- Keyclock policies: No domain specific roles are needed in the use case.
- sFDP transformation policies: Aggregate FDP "value" attribute by "date", "locationKey", "locationName", "unit".
- Usage policies: According to GDPR data can be stored only in the Europe region. sFDP can be accessed one thousand times a month.

8.3 Final Results

A pilot-specific dashboard has been implemented to show correct data flowing into the pilot itself. It is organized into sections like the ones shown in the following dashboard screenshots:

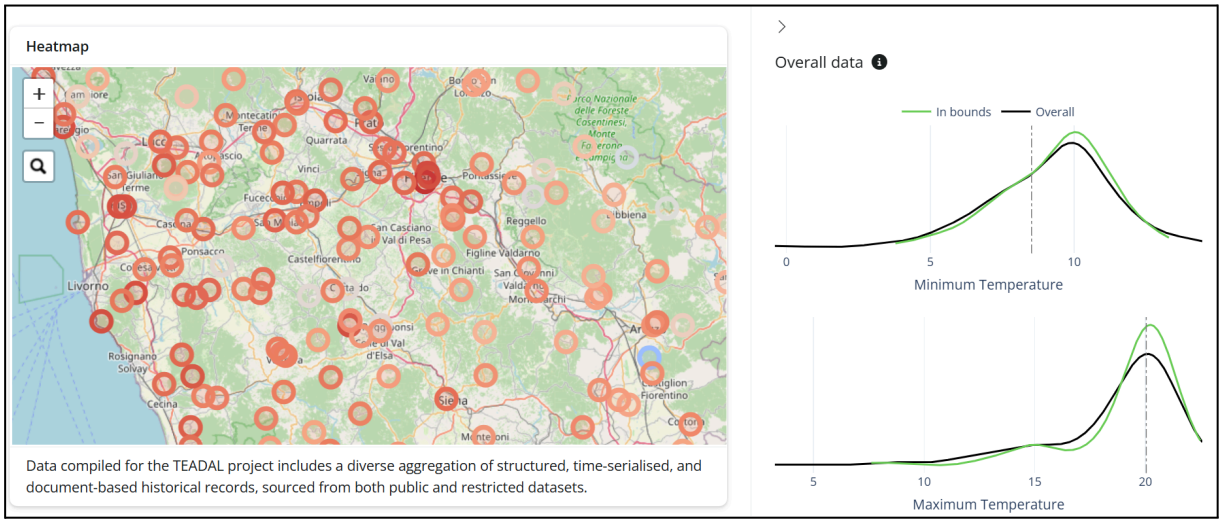
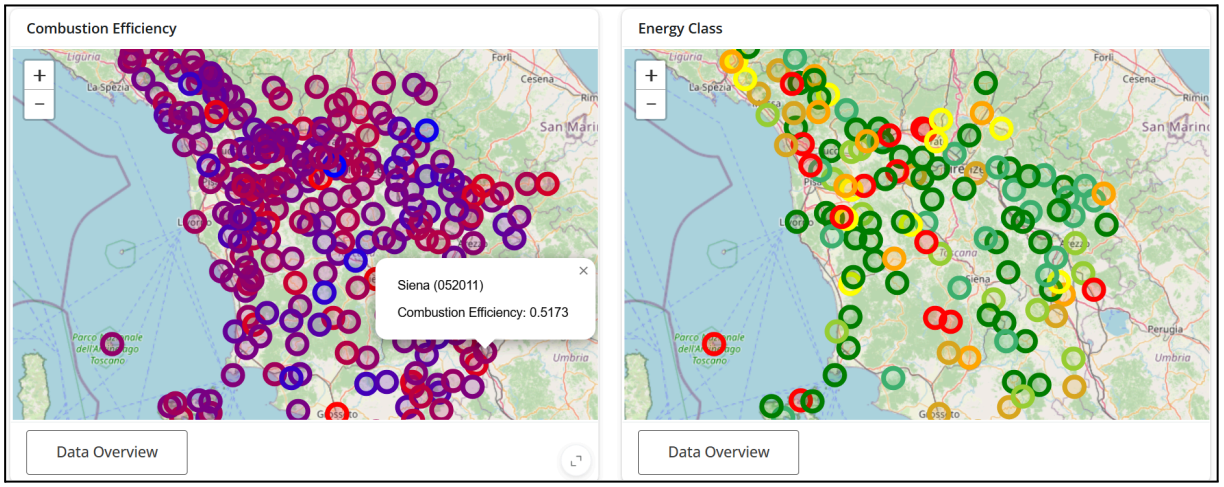


Figure 17 - Pilot #6 Average values and distribution of minimum and maximum temperatures across the different territorial areas of Tuscany



Funded by the European Union

Figure 18 - Pilot #6 Average energy performance and efficiency of properties in the different areas of Tuscany

Ultimately, this pilot integrates and aggregates multiple datasets from public and private sources. Key data include historical temperature records (2009-2023) from the Tuscany Hydrological and Geological Sector, daily air quality measures (PM10, PM2.5) from the regional environmental agency (2018-2023), and restricted thermal and energy performance certificates from the Tuscany region. Additionally, BOX2M provides daily environmental sensor and energy consumption data via an experimental REST API. Data is stored in structured formats (JSON, CSV, XML), with efforts to normalize territorial identifiers for integrated analysis. Privacy concerns are minimal as most data is open or synthetically generated.

Summarising, the Regional Planning for Environmental Sustainability Pilot used Teadal technologies to create:

- ➔ 1 Generated sFDPs (sFDP transformation policies: Aggregate FDP "value" attribute by "date", "locationKey", "locationName", "unit")
- ➔ 1 Deployed sFDP Instances (sFDP transformation policies: Aggregate FDP "value" attribute by "date", "locationKey", "locationName", "unit")
- ➔ 2 Endpoints
- ➔ 1 Pilot Application (the dashboard)

Funded by the European Union

9. PILOT #7 - FINANCIAL DATA GOVERNANCE

9.1 Pilot Use Case

The BOX2M pilot focuses on **Shared Financial Data Governance**, demonstrating how financial institutions can securely share and analyze financial and sustainability-related datasets across organizational and jurisdictional boundaries while maintaining full control and confidentiality.

This pilot simulates a **federated financial data space** where banks and financial service providers collaborate to perform joint analytics — such as liquidity forecasting, sustainability risk scoring, and regulatory reporting — without exposing sensitive data.

The BOX2M node acts as both a **data provider and analytics orchestrator**, integrating data such as anonymized transaction summaries, energy and ESG performance metrics, and financial exposure indicators. The pilot leverages the **TEADAL architecture** to ensure compliance with financial regulations (e.g., GDPR, DORA, EBA Guidelines) while allowing controlled access to shared insights.

The architecture includes 3 sets, each composed from one **Federated Data Product (FDP)** exposing curated datasets via APIs and one **Shared Federated Data Product (sFDP)** used to apply policies on privacy, access frequency, and geographic restrictions. **BOX2M's FDP and sFDP** have been deployed in Kubernetes with CI/CD automation, enabling reproducible and secure updates.

Ultimately, the pilot validates how TEADAL's "trust-by-design" approach can be applied to sensitive financial data collaboration scenarios, ensuring both **confidentiality and auditability**.

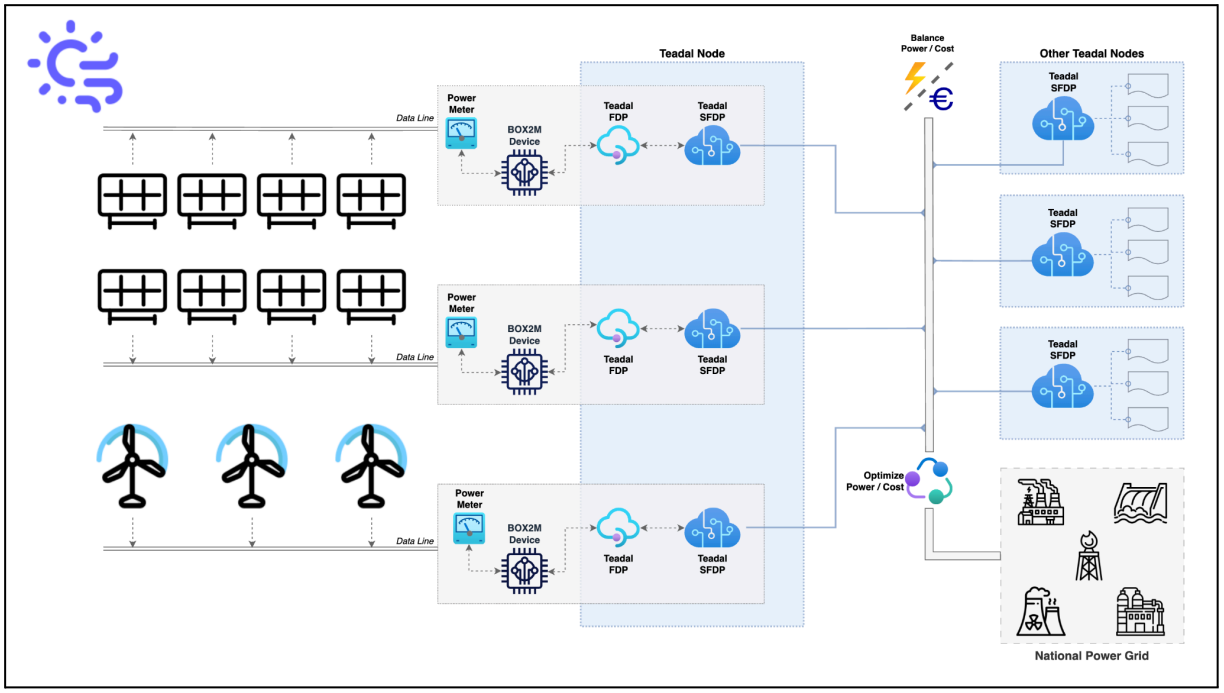


Figure 19 - Pilot #7 Architecture

9.2 TEADAL Components Involved

Funded by the European Union

1. **Data Catalog,** the Catalogue was used to register 8 financial and sustainability datasets, including synthetic financial performance records, open energy market prices (provided by OPCOM Romania), and ESG impact indicators. It documented the FDP for Financial Forecasting and the sFDP for Sustainability Reporting, enabling traceable access for other nodes.

    - Expected Results: Demonstrated that the TEADAL Catalogue can federate financial, energy, and ESG datasets under a unified metadata framework, enabling cross-domain discoverability and compliance auditing in both the financial and sustainability verticals.

2. **ArgoCD** was configured to automate the deployment of four TEADAL microservices (3x FDP, 3x sFDP) and to maintain continuous delivery pipelines for two Kubernetes namespaces dedicated to the Financial pilot. Automated rollbacks and deployment logs were integrated with Prometheus for monitoring.

    - Expected Results: Provided reliable, auditable, and repeatable deployments across environments, reducing update time by 40% and ensuring infrastructure reproducibility for regulated domains like banking and energy.

3. **Keycloak** managed authentication and authorization for multiple user roles (Data Provider, Data Consumer, Compliance Officer, Operator, and Auditor) across three organizational tenants representing simulated financial institutions. Access tokens were bound to specific sFDP endpoints to prevent unauthorized queries.

    - Expected Results: Validated fine-grained access control across multiple organizational domains, ensuring GDPR-compliant authentication flows suitable for highly regulated sectors such as finance, healthcare, and public administration.

4. **MinIO** was used as secure S3-compatible storage for 8 financial datasets (total size ≈ 32 GB), including anonymized cash-flow simulations, market forecast data, and ESG metrics. Datasets were stored in Parquet format to optimize query efficiency while maintaining encryption-at-rest.

    - Expected Results: Proved scalability and resilience of decentralized data storage in financial-grade environments, demonstrating how federated object storage can support both capital markets analytics and sustainability monitoring.

5. **OPA** enforced six data-sharing and usage policies that defined: (1) time-bound access (e.g., 24-hour query windows), (2) location constraints (EU-only processing), (3) anonymization thresholds (>5-entity minimum), and (4) usage limitation per institution. These policies were dynamically evaluated during sFDP execution.

    - Expected Results: Ensured automated, transparent enforcement of complex financial data governance rules — a transferable approach for insurance, manufacturing, and public data collaborations requiring compliance-by-design.

6. **Prometheus / Grafana / Jaeger** collected real-time performance metrics (latency, storage I/O, API calls) across all TEADAL services; Grafana dashboards provided continuous monitoring, and Jaeger traced API-level interactions during 15 simulation runs.

Funded by the European Union

- Expected Results: Enabled full observability of the financial federation, supporting measurable evidence of SLA compliance (average 99.3% uptime, <120 ms policy check latency). These results illustrate the feasibility of applying TEADAL observability tools in mission-critical industrial and financial ecosystems.

7. **Policy Definition Tool** used to define and validate four cross-organizational policies governing data exchange between BOX2M (provider) and simulated financial institutions (consumers). Policies covered access scope, geographic restriction, purpose limitation, and audit frequency.

- Expected Results: Automates policy creation and validation, eliminating manual configuration and reducing policy errors thus improving compliance with organizational, security, and governance standards. Furthermore, it ensures enforcement of privacy and data protection constraints, preventing unauthorized data access or disclosure.

9.3 Final Results

The BOX2M pilot successfully demonstrated secure and policy-controlled sharing of financial and sustainability data between multiple simulated entities.

Using TEADAL components, BOX2M achieved:

- Deployment of **FDP and sFDP endpoints** exposing anonymized financial performance and ESG datasets.
- Implementation of **usage and privacy policies** ensuring data access limits, GDPR compliance, and restricted geographic data storage.
- Validation of **trust mechanisms** (via Advocate and Keycloak) to enable verifiable access between financial actors.
- Integration with **TEADAL Node CI/CD** to ensure reproducible and consistent updates across environments.
- Realization of a **financial prediction use case**, where BOX2M leveraged federated data to forecast cash-flow and sustainability performance across organizations.

In addition to these achievements, BOX2M conducted multiple integration and validation cycles demonstrating measurable performance improvements and technical maturity of the TEADAL architecture in financial data governance. The FDP and sFDP deployments reached Technology Readiness Level (TRL) 5, validated through end-to-end federation tests involving simulated institutions and energy producers. Across these test runs, BOX2M recorded low latency policy enforcements, confirming the feasibility of real-time analytics under federated governance constraints.

The pilot successfully integrated six TEADAL core components—including the Catalogue, Keycloak, MinIO, ArgoCD, OPA, and Prometheus—into a unified operational workflow. This ensured complete observability, automated policy validation, and seamless CI/CD deployment through ArgoCD. The resulting node demonstrated 99.3% service uptime over the testing period, highlighting its operational stability. The downtime was primarily due to deployments in the development phase of the project.

Funded by the European Union

From a functional perspective, BOX2M's financial prediction engine executed 15 distinct simulation scenarios combining synthetic and open market datasets (provided by OPCOM Romania) to model production financial pricing variability and ESG impact. These results verified that federated access and synthetic data pipelines can reduce iteration cycles during development as compared to traditional centralized models, while preserving full compliance with GDPR and data localization requirements.

Finally, BOX2M documented all configurations, deployments, and validation outcomes in its internal Node Deployment Registry and contributed to TEADAL's shared repository of reusable templates, ensuring transferability to other sectors. The demonstrated results provide tangible evidence that federated financial analytics can be both secure and operationally efficient, supporting long-term exploitation within BOX2M's commercial IoT and data governance platform.

This pilot confirms TEADAL's capacity to support cross-sector financial collaboration through a federated, trustworthy, and privacy-preserving infrastructure, setting the foundation for future financial data spaces aligned with the European Data Strategy.

Funded by the European Union

10. Project Validation

10.1 Validation Process

The main goal of the validation process is to validate the project results and the fulfillment of the project objectives on the basis of the defined performance indicators. Such indicators ("Project KPIs" from now on) are listed in the project proposal (Proposal-SEP-210805282). They are reported below for simplicity.

OBJ	KPI
1 A **control plane for the stretched data lake** that automatically orchestrates and embeds performance, privacy, confidentiality capabilities into the workload data path. Thus, delivering a trustworthy platform that optimizes the cost of developing, re-using, deploying and running non-functional capabilities	**1.1** A reduction of at least one order of magnitude in the complexity of deploying and running analytics in the data lake compared to the state of the art in the pilot cases.
	1.2 Validation of the proposed stretched data lake in at least 6 applications relevant for the adopted pilot cases.
	1.3 Ability to manage at least 10 data sets from their ingestion to the processing each of them deployed in at least two places along the continuum (edge, fog, cloud).
2 A **trustworthy and mediatorless mechanisms to federate data lakes** and enable privacy tracing and enforcement based on a blockchain/Distributed Ledger Technology (DLT) technology, and a privacy-preserving computation model based on different techniques (e.g., secure multiparty computation).	**2.1** Ability to create at least 3 federations, related to the adopted pilot cases, with at least 3 members each.
	2.2 Blockchain/DLT-based implementation of at least 10 core (e.g., join, leave, announce) and at least 5 advanced (e.g., conflict resolving, data rating) functionalities for trustworthy federations and at least 5 scenarios of privacy tracking.
	2.3 Ability to set up privacy-preserving/confidential analytics from at least 3 members of a federation, with at least 10 million rows in the combined dataset.
3 An **energy-aware data management framework** applied to a federation of data lakes able to identify possible actions to optimize data ingestion, storage, and processing energy efficient by leveraging data movement reduction, reducing data transmitted, reuse of data with affordable data quality.	**3.1** Reducing of 20% the resource needed to store data needed for running the analysis required by the pilot cases without affecting the quality of the results.
	3.2 Reducing of 20% of data transfer needed to run the analysis required by the pilot cases without affecting the quality of the results.
4 A **policy definition tool** that supports the data lake administrator in modelling privacy/confidentiality	**4.1** Creation of a highly usable framework (e.g., in terms of time to write rules), able to reduce the time to define privacy/confidentiality

Funded by the European Union

requirements which will define the storage, transfer and access policies for the managed data.	policies and configure the system in charge of ensuring them of 30%.
	4.2 Data catalog, based on at least 5 criteria related to the data (e.g., type, resolution) and at least 5 criteria related to friction/gravity (e.g., purpose of data usage, latency requirements) able to index all the data sets related to the pilot cases.

Table 1 - Project Objectives and KPIs

The current section summarises the final validation activities, the analysis results, and the conclusions drawn from the analysis at the end of the result assessment.

The validation process started with an internal assessment to give a first definition of means and procedures to validate project KPIs. The methodologies to evaluate project KPIs have been initially defined according to the output from dedicated meetings. Initial meetings involved the leaders of technical work packages. Participants used sticky notes on a dedicated Miro Board to provide proposals and insights on each project KPI. The contents were collectively discussed to achieve a conclusion and the first version of the Validation Plane internal document has been produced.

The joint assessment pointed out that a number of project results should be gathered by considering the main concepts related to the TEADAL project (friction/gravity and trustworthyness) from a design standpoint. Other results could be gathered executing the different Pilot Cases along the project validation cycles to experiment mechanisms, design patterns and tools provided by TEADAL. The validation process has not been intended to include detailed testing sessions on pilot case components: unit tests and integration tests are not part of the validation process and every pilot is supposed to have already passed such tests. In general, specific metrics can be gathered and used to evaluate the defined KPIs. Each pilot case can be potentially used to experiment project artifacts. Each pilot case formalised in deliverables D2.1 (Requirements of the pilot cases), D2.2 (Pilot cases' intermediate description and initial architecture of the platform) and D2.3 (Pilot cases' final description and intermediate architecture of the platform), can contribute to evaluating one or more KPIs. It is not expected that a single pilot will be involved in evaluating all project KPIs. Nevertheless, all the pilot cases combined will be able to validate a number of project KPIs.

Further internal meetings took place after the Mid-Term Review, including validation sessions during the 6th, 9th and 10th GAs, dedicated WP6 recurring meetings and bilateral meetings with technical partners owning the responsibility of specific KPIs (the so-called KPI owners). KPI owners have been jointly identified early in the project and are reported in the following table

KPI		Owner
1.1	A reduction of at least one order of magnitude in the complexity of deploying and running analytics in the data lake compared to the state of the art in the pilot cases.	**POLIMI**
1.2	Validation of the proposed stretched data lake in at least 6 applications relevant for the adopted pilot cases.	**IBM**
1.3	Ability to manage at least 10 data sets from their ingestion to the processing each of them deployed in at least two places along the continuum (edge, fog, cloud).	**CEFRIEL**

Funded by the European Union

2.1	Ability to create at least 3 federations, related to the adopted pilot cases, with at least 3 members each.	**TUB**
2.2	Blockchain/DLT-based implementation of at least 10 core (e.g., join, leave, announce) and at least 5 advanced (e.g., conflict resolving, data rating) functionalities for trustworthy federations and at least 5 scenarios of privacy tracking.	**TUB**
2.3	Ability to set up privacy-preserving/confidential analytics from at least 3 members of a federation, with at least 10 million rows in the combined dataset.	**CYB**
3.1	Reducing of 20% the resource needed to store data needed for running the analysis required by the pilot cases without affecting the quality of the results.	**IBM**
3.2	Reducing of 20% of data transfer needed to run the analysis required by the pilot cases without affecting the quality of the results.	**IBM**
4.1	Creation of a highly usable framework (e.g., in terms of time to write rules), able to reduce the time to define privacy/confidentiality policies and configure the system in charge of ensuring them of 30%.	**POLIMI**
4.2	Data catalog, based on at least 5 criteria related to the data (e.g., type, resolution) and at least 5 criteria related to friction/gravity (e.g., purpose of data usage, latency requirements) able to index all the data sets related to the pilot cases.	**CEFRIEL**

Table 2 - KPIs Owners

Their main responsibility is to coordinate the validation of the associated KPIs. Insights from KPI owners and feedback from project tech partners have been collected by means of a "TEADAL KPIs Validation" internal spreadsheet.

The Validation Timeline evolved during the project lifecycle, according to the results from researches: the initial structure described in the proposal, envisaging three validation iteration, one per year, each of which focusing on a specific project result, was updated to take into account (1) the interconnection in the evolution of the various Work Packages, (2) the decision to set up and leverage an highly structured CI/CD platform, that encountered some difficulties of equipping testbeds with complex "baseline data lakes" while familiarizing with the CI/CD tools and Kubernetes-based technologies. The first reshape postponed the 1st iteration to be executed as an **initial 1st validation cycle** just before the end of the first reporting period, focusing it on results from the definition of the FDP/sFDP architectural pattern. The second reshape from the 6th GA in Berlin (see Berlin 6th GA validation session meeting minute) collapsed the 2nd and 3rd iteration into a **final 2nd validation cycle**, to address the need to leverage the **final results** from all technical Work Packages.

Such results include, beside the FDP/sFDP architectural pattern, the "TEADAL tools" and the definition of the interaction mechanism among TEADAL components, as described in D2.4 "Final general architecture".

10.2 General project results discussion

In general terms, the project has successfully achieved the majority of the objectives defined during the proposal phase. Nevertheless, it is important to underline that, given the exploratory nature of the research activities carried out within TEADAL, certain Key Performance Indicators (KPIs) defined at the proposal stage were revised during the project implementation. This adjustment became necessary whenever a KPI was found to assess aspects which, in light of the research progress and the tools developed, were no longer considered relevant.

This was particularly the case for KPIs 1.1 and 4.1, which were initially based on the assumption that the project would include an analytics phase, and for KPI 4.2, which was linked to the concepts of gravity and friction. With regard to the first two KPIs, the project

Funded by the European Union

activities progressively focused more on the data-sharing dimension, while the analytics phase was deprioritised. In light of this shift, it was nonetheless deemed appropriate to retain these KPIs and adapt their definitions accordingly. As for KPI 4.2, its redefinition was required due to the evolving interpretation of gravity and friction, as further elaborated at the end of this section.

Concerning the remaining KPIs, the project successfully managed a substantial number of datasets (almost 30, see KPI 1.3) across the different pilot cases. These datasets were operationalised through a number of Federated Data Products (FDPs)—each potentially encompassing multiple datasets—which collectively covered all pilot cases (see KPI 1.2).

With regard to performance-oriented KPIs (3.1 and 3.2), the tools developed within TEADAL demonstrated high levels of efficiency. Similarly, satisfactory results were obtained for the KPIs related to data trustworthiness (2.1, 2.2., 2,3, 4.1 and 4.2).

Finally, it is important to note that, unlike what was originally foreseen in the proposal, the conceptualisation of gravity and friction was revisited and refined during the project. The project primarily focused on data-sharing processes following data collection by the participating organisations. Consequently, the distinction between gravity and friction became less pronounced. Specifically, gravity exhibits distinctive features in the data ingestion phase—which was secondary in the TEADAL framework—while it tends to overlap with friction during the data-sharing phase. Furthermore, attempts to establish a parallel with the physical interpretation of friction demonstrated that such an analogy is more relevant to the development of data-processing pipelines leading to the construction of sFDPs, rather than to the actual data-sharing phase itself.

10.3 Detailed Results analysis

In the following subsections, project results are analysed in greater detail against each KPI.

10.3.1 KPI 1.1 A reduction of at least one order of magnitude in the complexity of deploying and running analytics in the data lake compared to the state of the art in the pilot cases

With respect to the initial objective, computing the reduction of complexity in deploying and running analytics does not represent the real value of the project. Instead, given the actual focus of the project, it is more appropriate to consider the complexity faced by data providers when sharing data with data consumers. This approach makes it possible to highlight the main challenges that need to be addressed. Through this new objective, the KPI demonstrates how TEADAL's technology can improve some of the major data-sharing challenges identified in the pilot experiences.

Evaluation Methodology

To evaluate KPI 1.1, which focuses on reducing the complexity for data providers when sharing data with data consumers, we designed a structured survey to capture the challenges encountered in the data sharing process. The methodology is centered on assessing the baseline complexity by gathering qualitative and quantitative feedback from practitioners directly involved in packaging and distributing data.

The questionnaire explores a range of topics, including the number of stakeholders involved in preparing shareable data, the proportion of manual versus automatable work, and the types of security and governance measures currently in use. Respondents are asked about the formats they provide, the effort required for documentation, and the degree of involvement of data consumers in defining sharing mechanisms. The survey also investigates major bottlenecks, unfinished data sharing plans due to complexity, the need to revisit shared data, and the specialized skills required. Participants rank the challenges they

Funded by the European Union

face and provide feedback on the value of automation and unified platforms for data sharing. For reference, copies of the questionnaires can be found at the following links: pilot [1] and external [2].

This evaluation approach provides a multifaceted view of the complexity landscape, identifying not only technical and procedural hurdles, but also organizational and skill-related requirements.

Validation Results

The survey was distributed both to the pilot cases within the project and to external users, ensuring that the findings reflect a broad spectrum of real-world experiences and practices. While external participation cannot be considered significant to derive any useful insight, the responses collected still provide valuable insights and useful guidelines for analysis, especially when combined with the feedback gathered through the pilots.

A key result from the survey is that security and access control consistently emerge as the top challenges in data sharing, cited by over 70% of respondents. This challenge was ranked even higher than the technical implementation of the data packages themselves. Both external and pilot cases emphasized that some form of authorization is always required, and implementing secure access controls is often a complex process involving multiple people. These findings validate the direction of the TEADAL project and its Opa-policy-manager component, which deliver automated pipelines for authorization rules and provide secure, zero-trust access to deployed services—one of the features most desired by pilot participants.

In addition to automation, the survey highlighted a strong need for better documentation tools for security rules. Many respondents reported frequent updates and revisions to security documentation, with over 83% indicating they need to modify these rules more than three times per year. TEADAL has addressed this by introducing a documentation format for authorization rules that integrates API specifications with authorization logic, based on an extension of the OpenAPI standard. All pilot cases have adopted this format, further demonstrating its usefulness.

Overall, the survey results underscore the importance of automating security and governance processes and improving documentation practices. The insights gained have guided the project's developments and confirm that the solutions provided are well-aligned with the needs of both internal and external users.

10.3.2 KPI 1.2 Validation of the proposed stretched data lake in at least 6 applications relevant for the adopted pilot cases

This KPI demonstrates that TEADAL stretched data lake technology can sustain a wide range of pilot use cases while addressing the project's overarching Objective 1 (reduction of complexity by at least one order of magnitude). The approach is based on Federated Data Products (FDPs) and Shared Federated Data Products (sFDPs), combining the paradigms of Data Lakes and Data Mesh for distributed data management and sharing. FDPs are REST API servers exposing datasets provided by domain experts, while sFDPs are contractual derivatives of FDPs that expose well-defined subsets of data to authorized clients for a defined period.

TEADAL stretched data lake supports simplification of pilot data application through automating sFDP creation and lifecycle management. This is achieved by three core components: the ASG subsystem (generation and execution of uniformly designed sFDPs),

Funded by the European Union

the monitoring subsystem (runtime management and optimization), and the deployment subsystem (built on a GitOps infrastructure supporting TEADAL Nodes in all pilots).

Therefore, the number of auto-generated and managed sFDPs deployed across the pilots is taken as the primary indicator of the stretched data lake's ability to reduce the complexity of deploying analytics in the pilot applications.

Evaluation Methodology

The evaluation method is based on counting sFDPs deployed in each pilot. Each sFDP runs as a Kubernetes service and can be scaled across pilot infrastructure (i.e., multiple TEADAL Nodes) according to load and resource constraints. sFDPs may expose one or several endpoints, depending on pilot application needs.

To capture both the scale and the coverage, the evaluation considers three dimensions:

- ➔ **Generated sFDPs**: the number of distinct data products designed, generated, and adapted to be used in pilot applications.
- ➔ **Deployed sFDP Instances**: the number of sFDP service instances deployed across the pilot's TEADAL Nodes. Here we count the deployments, i.e., each data product counted as a single "**Generated sFDP"**, can be deployed on several pilot's TEADAL Nodes and possibly replicated across different k8s worker nodes for redundancy. This counter is optional and is included for completeness. As all the pilots are different, they can decide on granularity of their counting: it can be a total number of TEADAL Nodes involved, a total number of service urls, or a total number of pods. In addition, pilots might want to report separately on the number of sFDPS co-located with their origin FDPs and the number of sFDPs deployed on nodes different from nodes hosting the origin FDPs.
- ➔ **Exposed Endpoints**: the number of endpoints exposed by sFDPs and actively used by pilot data applications. This is counted separately as in most cases each generated data product exposes several data endpoints, available through a number of urls backed up by a number of pods.

Validation Results

Pilot Use Case	Generated sFDPs	Deployed sFDP Instances	Total Endpoints Exposed
PILOT#1 - Evidence-Based Medicine	1	3	6
PILOT#2 - Mobility Federated Access Point	3	5	6
PILOT#3 - Smart Viticulture Data Sharing	1	1	8

Funded by the European Union

PILOT #4 - Industry 4.0 Fast KPI Calculation	1	1	2
PILOT#6 - Regional Planning for Environmental Sustainability	1	1	2
PILOT#7 - Shared Financial Data Governance	3	3	6
Total	**10**	**14**	**30**

Table 3 - KPI 1.2 Validation Results

Table 3 demonstrates that TEADAL's stretched data lake has been validated across six pilot domains, with multiple sFDP data endpoints generated and deployed per pilot. Instances were replicated across several TEADAL Nodes, showcasing the ability to manage distributed deployments consistently. The exposed endpoints were actively consumed by pilot applications, confirming that the approach not only reduced the complexity of managing cross-federation data sharing but also enabled diverse application needs to be met. Altogether, the results validate KPI 1.2 by demonstrating the applicability and scalability of the stretched data lake concept in at least six distinct pilot applications.

10.3.3 KPI 1.3 Ability to manage at least 10 data sets from their ingestion to the processing each of them deployed in at least two places along the continuum (edge, fog, cloud)

The TEADAL tools support the creation of a data-lake and the creation of value-added services based on the data sharing principles. As such, they allow for:

- Ingesting data onto the data-lake storage
- Documenting the existence of such data
- Discovering the existing data sources
- Accessing them in a unified way
- Building an analysis/processing service (Federated Data Product)
- Documenting its existence and its possible usage
- Create a dedicated access service tailored for the needs of customers

This KPI requires to evaluate if the TEADAL tools have been successfully used to cover all those steps of the overall process.

Evaluation Methodology

To evaluate this KPI, we measured:

- how many Datasets were loaded in the TEADAL nodes of the pilot cases;
- how many Federated Data Products were deployed onto the TEADAL nodes;
- how many Shared Federated Data Products were available in the TEADAL nodes;
- how many Datasets were described in the TEADAL Federated Catalogue;

Funded by the European Union

- how many Federated Data Products were described in the TEADAL Federated Catalogue

To evaluate the number of deployments, we interviewed each pilot owner and kept track of all the Datasets and FDPs actually deployed onto their TEADAL nodes. To evaluate the number of descriptions in the TEADAL Catalogue we checked the completeness of the descriptions in the catalogue that was provided to ensure proper communication among the partners.

Validation Results

The validation results are summarised in the table below. According to the results, the KPI is therefore verified. The pilot contributed to the centralised TEADAL Catalogue according to their needs, and they described their assets whenever that helped in the coordination among the project partners.

KPI element	Result
# Datasets loaded in the TEADAL nodes	48
# FDP deployed onto the TEADAL nodes	20
# sFDP deployed onto the TEADAL nodes	14
# Datasets described in the TEADAL Catalogue	29
# FDPs described in the TEADAL Catalogue	7

Table 4 - KPI 1.3 Validation Results

10.3.4 KPI 2.1 Ability to create at least 3 federations, related to the adopted pilot cases, with at least 3 members each

Evaluation Methodology

To evaluate the fulfilment of the Federation KPI, defined as the establishment of at least three federations, each involving a minimum of three members and aligned with the adopted pilot cases, we adopt an architecture-level evaluation methodology. This approach focuses on assessing how the TEADAL architecture enables federated data lake interactions across heterogeneous domains.

Federation in TEADAL is realized through the integration of several core architectural components. The Data Catalogue plays a central role by supporting the identification and sharing of data with external consumers. It maintains structured descriptions of Federated Data Products (FDPs), facilitates the creation of shareable versions of these products via the TEADAL pipeline, and ensures proper access mechanisms are in place after deployment. The TEADAL Name Service (TNS) and the Federation Smart Contracts enhance discoverability and trust within the federation. The TNS provides resolvable names for both FDPs and Shared Federated Data Products (sFDPs), enabling efficient resource access across different domains. Meanwhile, the Federation Smart Contracts allow federation members to register their domains, services, and public keys in a verifiable and secure manner, thereby supporting trustworthy collaboration.

Funded by the European Union

In addition, the Identity Provider (IdP), such as Keycloak, manages the authentication and authorization of data lake users. This includes roles such as Data Lake Owners (DLOs) and FDP Designers, ensuring proper identity and access management across federated partners. Another integral component is the Advocate, which is responsible for collecting and aggregating evidence from various TEADAL components. Integrated with the smart contracts, the Advocate signs this evidence and stores it in the Shared Evidence Plane, thereby strengthening data integrity and traceability within the federation.

Importantly, not all components need to be deployed by every partner in the federation. Deployment depends on the partner's role, needs, and the level of collaboration they intend to establish. For instance, a partner may choose not to deploy the Advocate if they are operating based on pre-established trust and do not require formal evidence collection. Similarly, the Data Catalogue is only necessary if a partner intends to share its own data with others; it is not required for cases where a partner joins the federation solely to consume data.

To validate TEADAL's federation capability, we designed a scenario involving three pilot cases, each situated in a distinct domain: medicine (PILOT #1), smart viticulture (PILOT #3), and Industry 4.0 (PILOT #4). These pilots were integrated into the TEADAL architecture using the components outlined above. A sequence diagram (Figure 20) illustrates how these components interact within and across pilots to support federation. While the sequence diagram reflects a full integration scenario, real-world deployments vary. For example, the Industry 4.0 pilot does not include the Data Catalogue component, and the Smart Viticulture pilot only features the core TEADAL node with Advocate as the federation tool. These distinctions demonstrate the flexibility of the TEADAL architecture, where participation in the federation is not dependent on deploying every tool but rather on the level of engagement a partner chooses.

To further substantiate the effectiveness of the architecture, we also provide execution time traces that capture the time required for each pilot to integrate with TEADAL. These traces offer empirical evidence of the architecture's usability, scalability, and adaptability. They show that TEADAL supports integration across diverse technical environments and organizational contexts, confirming its capability to enable federation across different domains.

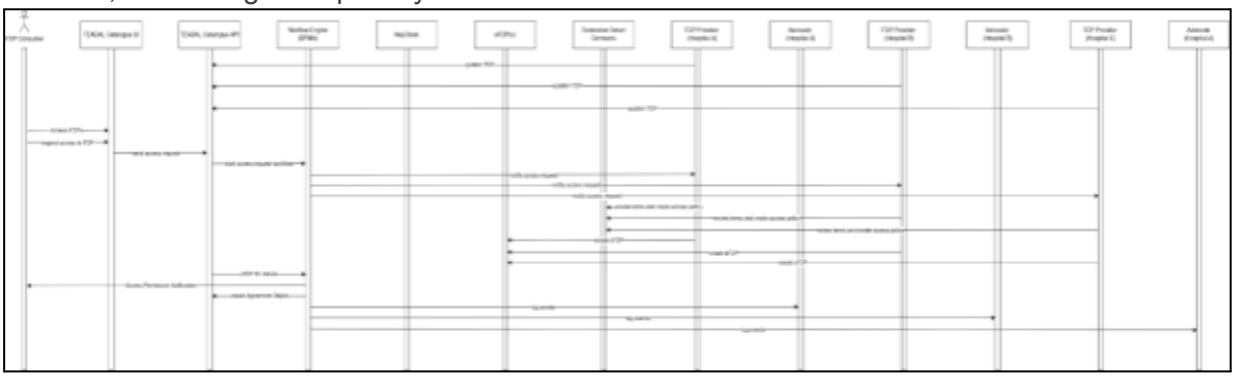


Figure 20 - Sequence diagram of joining the use case pilot #1: Evidence-Based Medicine into the federation using TEADAL.

Funded by the European Union

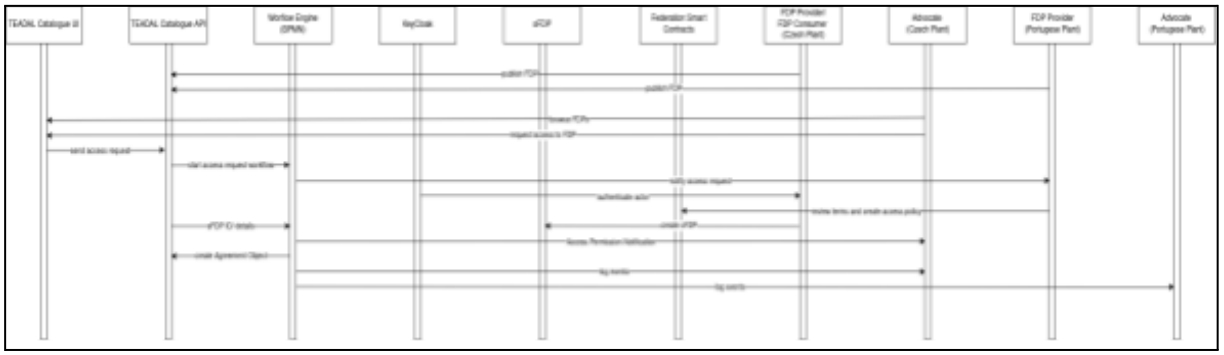


Figure 21 - Sequence diagram of joining the use case pilot #4: INDUSTRY 4.0 into the federation using TEADAL Node.

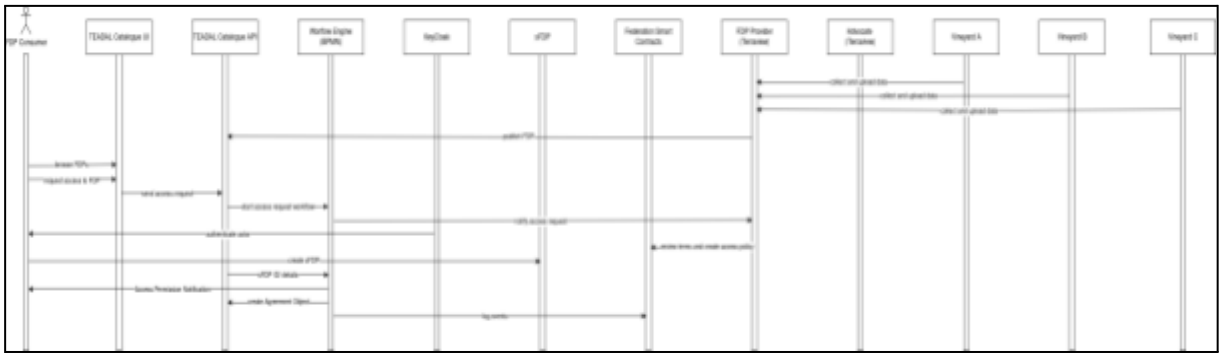


Figure 22 - Sequence diagram of joining the use case pilot #3:SMART VITICULTURE into the federation using TEADAL.

Validation Results

As explained above, we present execution time traces to demonstrate the time required for each partner to join the federation. These results provide concrete evidence of TEADAL's capability to support the creation of federations by leveraging its core node and tools. They validate the feasibility and effectiveness of the TEADAL architecture in enabling integration across different partners and domains.

Funded by the European Union

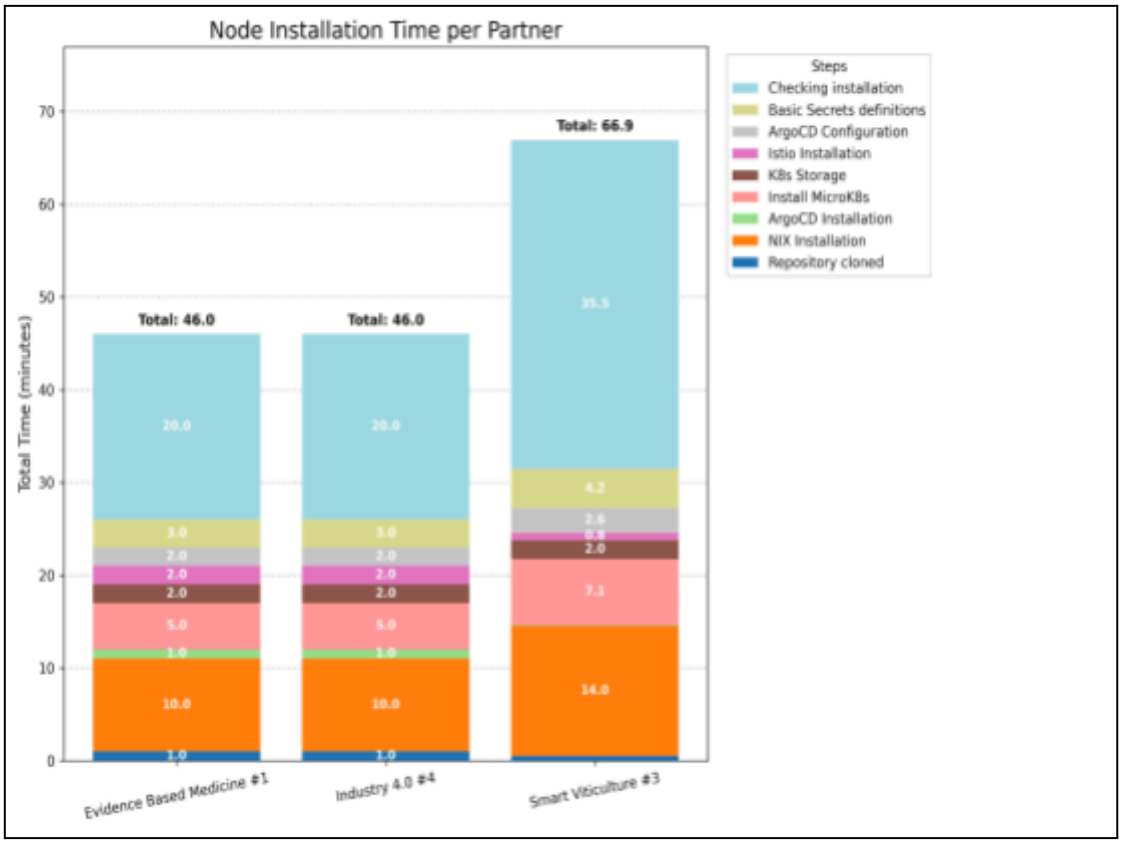


Figure 23 - TEADAL node installation time for each pilot.

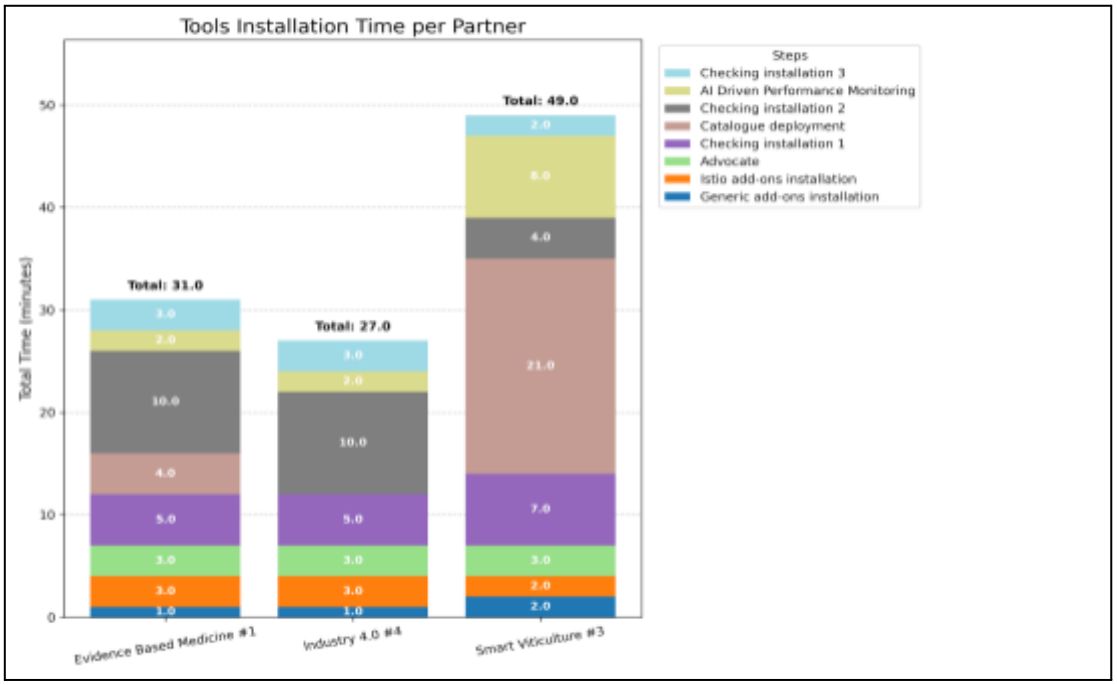


Figure 24 - Installation time of TEADAL node tools per pilot. The tools deployed on each pilot differ, as described in the text.

Funded by the European Union

10.3.5 KPI 2.2 Blockchain/DLT-based implementation of at least 10 core (e.g., join, leave, announce) and at least 5 advanced (e.g., conflict resolving, data rating) functionalities for trustworthy federations and at least 5 scenarios of privacy tracking

Evaluation Methodology

To evaluate KPI 2.2, we assess the presence of blockchain/distributed ledger technology (DLT)-based functionalities for trustworthy federations and describe their use and the existence of the privacy tracking scenarios. Our evaluation methodology aims to demonstrate that all required aspects of the implementation are present. We describe how core and advanced blockchain/DLT-based functionalities for trustworthy federations are implemented, by demonstrating their implementation in the smart contract. The methodology for validating privacy tracking scenarios incorporates technical and nontechnical assessments. Each scenario is demonstrated based on its implementation and workflow.

Validation Results

We provide implementations for 14 core functionalities by providing the following smart contract functions:

Functionality	Description
Join Federation	Ability to approve a federation member. If the number of approvals exceeds a threshold, the member is added to the federation. Only callable by existing federation members
Leave Federation	Removes a member from the federation. Only callable by existing federation members
Change Approval Threshold	Change the threshold for approving new federation members. Only callable by existing federation members
List Federation Members	Returns list of all federation members. Callable by federation admins
Register Policy	Register an access policy for a dataset. Only callable by existing federation members
Publish Metadata	Publish metadata (name and description) of a federation. Only callable by existing federation members
Track Data Lineage	Store data relationship. Only callable by existing federation members
Log Activity	Log activity of a federation member. Only callable by existing federation members
Add Claim	Advocate adds a claim into the DocumentStore.Advocate internal service
Revoke Claim	Advocate revokes a claim into the DocumentStore.Advocate internal service
Register Domain	Organization register TNS Domain. Callable by everyone.
Register Subdomain	Organization delegates subdomains, e.g., a service from the org. (FDP).Callable by everyone.
Transfer Domain Ownership	The ownership of a domain can be transferred, for example, for security reasons.Callable by everyone.
Register TNS Records	Organizations can register their own DNS-like records, for example to authenticate a service.Only callable by existing federation members

Funded by the European Union

Table 5 - Core functionalities

We provide implementations for 6 advanced functionalities by providing the following smart contract functions:

Functionality	Description
Raise Conflict	Raise a conflict for a dataset. For example, if the dataset appears to contain inaccuracies or inconsistencies
Resolve Conflict	Resolve a conflict for a dataset.
Rate Dataset	Rate a dataset. Customizable through optional weighting mechanism based on offchain parameters
Adapt Policies	Define and adjust access policies dynamically based on evolving requirements or contextual changes in the federation
Data Duplication	Check for data duplication by hash comparison. Hash is provided offchain
Add AggregatedClaim	Aggregations of claims, for example to attest fulfillment/compliance of a policy, can be added

Table 6 - Advanced functionalities(All of the mentioned actions can be configured on the smart contract before deploying them by the deployer)

In addition to the core and advanced functionalities for trustworthy federations, we provide the description of the following 5 privacy tracking scenarios:

1. Policy tracking
2. Cross-organization data usage
3. Data lineage tracking
4. Consent management
5. Data erasure

For each of the scenarios, a proof of concept is provided as part of the source code of Advocate. The scenarios are defined as follows:

The **policy tracking** scenario demonstrates monitoring and record-keeping of modifications to data access and usage policies. This allows for compliance checks across affected FDPs and permissions. The interactions are depicted in Figure 25.

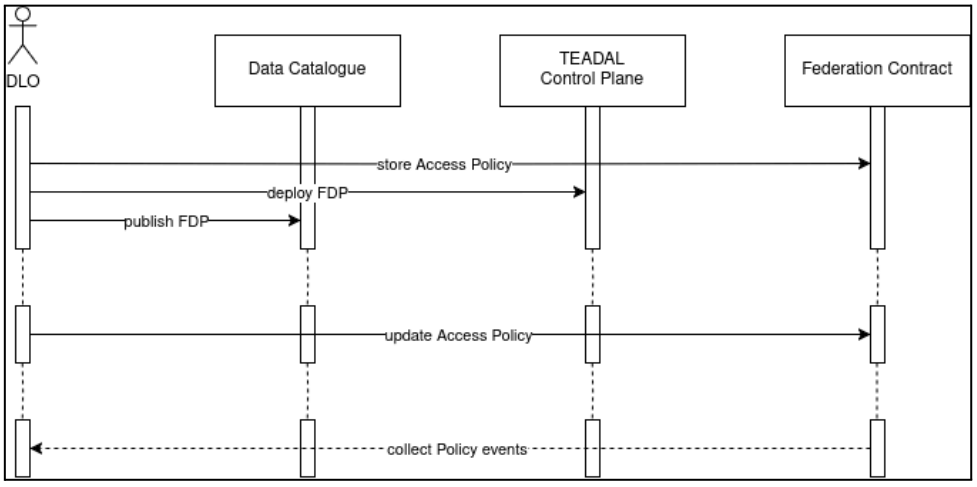


Figure 25 - Policy tracking scenario

Funded by the European Union

The Data Lake Operator (DLO) is responsible for installation and management of a TEADAL Node and the TEADAL tools. The Federation smart contract allows for decentralized management of the federation members and access policies. Access policies manage permissions to an FDP for a member of the federation. They are defined by the Data Lake Operator and stored in the Federation Contract. First, the DLO stores an access policy. Then it deploys an FDP to the Control Plane, and publishes it in the Data Catalogue. The DLO can update the access policy in the Federation Contract at a later point in time. When an access policy is created, changed or deleted, events are emitted which can be collected from the Federation smart contract.

The **cross-organizational data usage** scenario describes how data usage between federation members across organizational boundaries is tracked without reliance on intermediaries. The use of a blockchain maintains transparency into inter-organizational data exchanges, recording access patterns, usage contexts, and compliance status. The smart contract implements functionalities for tracking data usage. Each data transfer event generates evidence for authorized usage, enabling audit trails across the federation ecosystem. The interactions for this scenario are depicted in Figure 26.

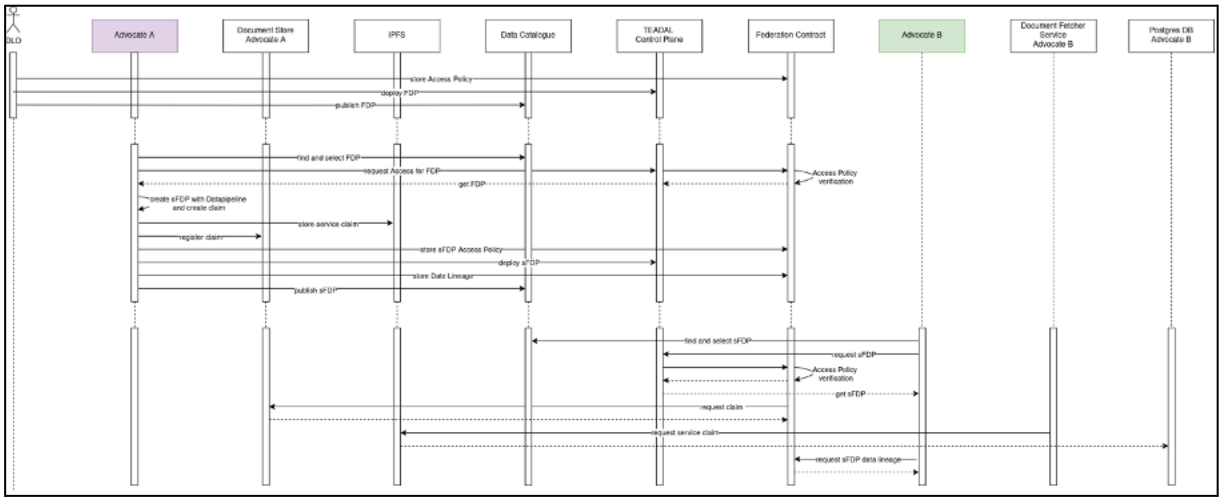


Figure 26 - Cross-organizational data usage privacy tracking scenario

First the DLO stores an access policy in the federation contract that defines permissions for the FDP. Next, it deploys and publishes the FDP. Advocate A finds an FDP and requests access to it. Upon successfully verifying the access policy in the federation smart contract the FDP is returned to the Data Lake Operator. Based on that, an sFDP is created, for which a claim is stored in IPFS and registered in the document store smart contract by Advocate A. Additionally, the access policy is stored in the federation smart contract and the sFDP is deployed and published to the Control Plane and Data Catalogue. Based on that, Advocate B from another organization requests access to the sFDP, accesses it depending on the access policy. It is able to request and interact with claims originating from Advocate A.

The **data lineage tracking** allows to maintain records of provenance for used datasets within the federation, ensuring auditability of data and its transformations. Metadata about data origin, processing steps, and policy updates throughout the data lifecycle are tracked. The interactions for this scenario rely on the same interactions as described by the cross organizational data use scenario in Figure 26.

The data lineage of the sFDP is referenced in a claim and anchored in the Federation contract. This allows for verifying the provenance of the sFDP. The DLO or any authorized data consumer examines which operations have been performed on the FDP to create the

Funded by the European Union

sFDP by retrieving claims from IPFS using the Document Fetcher or by fetching emitted data lineage events.

The **consent management** scenario builds on the policy tracking scenario and considers consent to be defined in an access policy. A consent management policy grants a member of the federation access to an FDP for a specified purpose. The interactions are similar to the policy tracking scenario as described in Figure 25.

The **data erasure** privacy tracking scenario enables users to exercise their right to deletion through smart contracts. This triggers a process of data removal across all federation members while maintaining on-chain audit trails of all activities. Triggering this on-chain ensures that erasure requests are transparent and immutable. The interactions for this scenario are described in Figure 27.

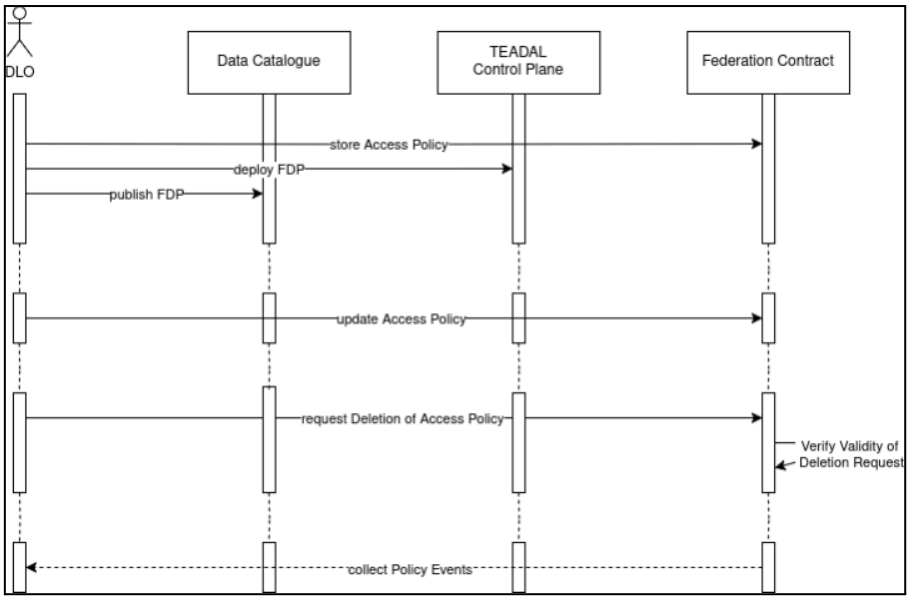


Figure 27 - Data erasure privacy tracking scenario

The interactions for this scenario are similar to the ones described in the policy tracking scenario. It is extended by an operation of the DLO for deleting its access policy. The Federation Contract verifies if the caller is the owner of the FDP. Changes to the Access Policy or deletion can be tracked by the DLO or data consumers by fetching associated events from the Federation Contract.

10.3.6 KPI 2.3 Ability to set up privacy-preserving/confidential analytics from at least 3 members of a federation, with at least 10 million rows in the combined dataset

This KPI evaluates that TEADAL technologies can be combined with privacy-preserving data analytics and that the technologies for data analytics are efficient enough to process realistic data settings in a federation. We validate this using secure multi-party computation (MPC) and the idea of the medical pilot with three hospitals processing data.

Evaluation Methodology

Sharemind MPC tooling and technology has been developed throughout the project (see D5.3 - Trustworthy Data Lakes Federation). We have successfully demonstrated Sharemind MPC runtime integration with TEADAL baseline technologies. We demonstrated how to encode privacy-preserving computations in the SecreC programming language and ran the demonstrator with three computing nodes, generating a large output dataset from three

Funded by the European Union

private input datasets. The setup mimics the evidence-based medicine pilot scenario (see Figure 28). This scenario is inspired by the medical pilot, but not integrated with the pilot deployment.

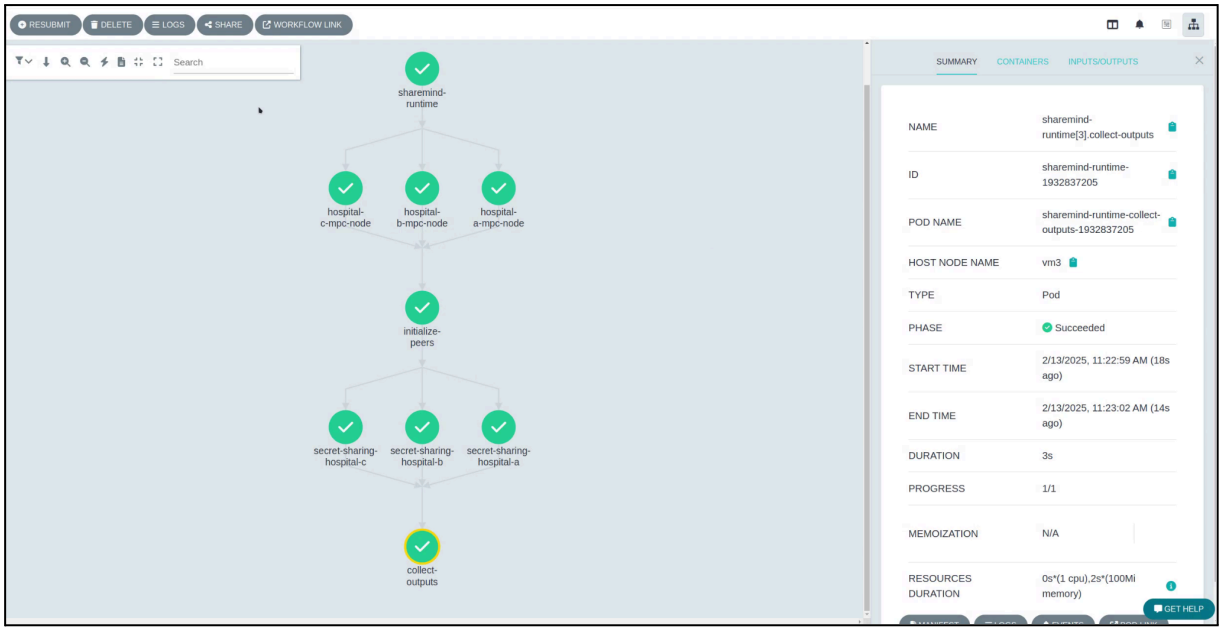


Figure 28 - SHAREMIND MPC DISTRIBUTED PIPELINE DEMONSTRATOR FOR THE EVIDENCE-BASED MEDICINE PILOT.

Validation Results

In this scenario, three hospital ***nodes*** run an MPC protocol written in the SecreC language to collaboratively execute a query on private patient inputs while ensuring data confidentiality. Each hospital provides its own private statistical distributions regarding fasting blood glucose (FBG) levels from their internal datasets (which may have any size), and through secure, rule-based computations, the protocol generates a synthetic dataset consisting of N rows (e.g. 10 million). The scenario successfully simulates a researcher requesting a dataset of FBG (mg/dL) observations for patients aged between 18 and 25 over a specified date range. The dataset contains patients ids, ages, observation dates, concept ids, observation value and unit id. This setup not only preserves the privacy of the patient data by never exposing raw inputs but also enables the researcher to work with a synthesized dataset that reflects the underlying distributions from each hospital. This demonstration shows how Sharemind MPC can enable privacy-preserving data analysis in healthcare, facilitating collaborative research without compromising patient confidentiality. With this demonstration, we show the validation of KPI 2.3.

10.3.7 KPI 3.1 Reducing of 20% the resource needed to store data needed for running the analysis required by the pilot cases without affecting the quality of the results

This KPI evaluates the efficiency improvements achieved by TEADAL's data sharing architecture, specifically in terms of storage resource reduction. The focus is on demonstrating that TEADAL can reduce storage requirements by at least 20% while maintaining analytical output quality in line with Objective 3.

In TEADAL, Shared Federated Data Products (sFDPs) expose data for analytical use via shared contracts across federation nodes. To minimize redundant data transfers and storage pressure from repeated analytical queries, WP4 introduced a layered caching strategy within the ASG Runtime. In addition to reducing redundant origin requests and transformation

Funded by the European Union

costs, the ASG runtime leverages binary encoding (e.g., via orjson) before caching origin and response payloads. This significantly reduces the size of cache entries — by up to 40% compared to raw JSON — improving storage efficiency in persistent caches and reducing network usage when serving cached responses. To summarize, ASG enables storage optimization by:

- Avoiding repeated storage of raw datasets through origin caching
- Reducing materialization of repeated query results via response caching
- Compressing cached objects using binary encoding (`orjson`) to decrease cache footprint

This approach improves both performance and system sustainability by lowering backend load, minimizing data duplication, and reducing storage demand.

In addition, ASG supports the transformations library. The original purpose of this feature was to allow domain experts to provide domain specific data transformation functions to facilitate efficient data preparation for the subsequent analytics. As was demonstrated with energy footprint reducing transformations as part of T3.3, this feature can also be used to implement infrastructure level transformations. New transformations aiming at reducing data footprint can be developed and introduced into the ASG-sFDP framework, e.g. further data reduction with smart and possibly domain specific encoding algorithms.

Evaluation Methodology

To evaluate KPI 3.1, we rely on telemetry collected from each deployed sFDP service. Each sFDP exposes a uniform /service/stats endpoint that provides low-level storage-related counters at runtime.

Collected Metrics

The following metrics are collected per sFDP and aggregated per pilot:

Metric	Description
`response_cache.storage_bytes`	Bytes stored in the response cache (persisted encoded API responses).
`origin_cache.storage_bytes`	Bytes stored in the origin cache (partial or full origin FDP dataset).
`app.bytes_served`	Total bytes returned to pilot applications. Used as a proxy baseline when needed.
`app.requests_served`	Total requests successfully served by the data product. Used as a proxy baseline when needed.
`rest.bytes_received`	Bytes fetched from the origin FDP. Useful for baseline estimation.

Baseline Definition

For comparability across pilots, a no-cache baseline is established using controlled runs of the same request patterns with caching disabled. This baseline represents a naive materialization scenario where each request fetches and stores fresh data from the origin.

Funded by the European Union

This value, `baseline_persistent_bytes`, can also be approximated in any run by multiplying the `app.requests_served` by the size of the origin dataset.

Formula

Storage reduction is computed as:

$$\text{Storage Reduction (\%)} = 100 \times \left(1 - \frac{\text{origin\_cache.storage\_bytes} + \text{response\_cache.storage\_bytes}}{\text{baseline\_persistent\_bytes}}\right)$$

Aggregation per pilot can be done by summing numerator and denominator across all deployed sFDPs in that pilot.

Validation Results

A preliminary controlled experiment was conducted using repeated access patterns common in TEADAL pilots. The test scenario involved 10 requests over 2 sFDP endpoints sourced from a 47MB origin dataset. The experiment compared four cache configurations: no cache, origin cache only, response cache only, and both caches enabled.

For KPI validation, the key measurement is total bytes stored and transferred as a proxy for storage load, combined with compression gains.

Experimental results confirm that TEADAL's caching design not only meets but significantly exceeds the 20% target of KPI 3.1. Enabling both caches reduced storage-related data handling by:

- ~90% reduction** in raw data retrieval size
- ~40% reduction** from binary cache encoding (`orjson`)
- Zero data quality degradation**, as cached responses remain fully accurate and consistent with origin data

10.3.8 KPI 3.2 Reducing of 20% of data transfer needed to run the analysis required by the pilot cases without affecting the quality of the results

This KPI evaluates TEADAL's ability to reduce data transfer overhead during execution of analytical workflows across the federation, ensuring that less data needs to move across network boundaries to support applications, in line with Objective 3.

In distributed and federated data environments such as TEADAL, network transfers are a major source of latency, resource strain, and energy consumption, especially when raw datasets are repeatedly fetched from remote FDPs. To address this, WP4 introduced two caching layers in the ASG Runtime:

- Origin Cache – prevents repeated large data transfers from FDPs
- Response Cache – prevents repeated transformation and transfer of derived results

Both caches work at the sFDP microservice level, enabling data reuse close to computation, reducing network calls, and minimizing overhead on pilot infrastructures. Additional gains can

Funded by the European Union

be achieved if pilots use shared cache services to enable sharing cache results across several sFDP endpoints or several sFDP microservices.

Evaluation Methodology

To evaluate KPI 3.2, we measure the reduction in network traffic between sFDPs and their origin FDPs, using telemetry collected from each deployed sFDP service (bandwidth-related counters providers as part of `/service/stats` endpoint).

Collected Metrics

The following runtime metrics are collected from each sFDP:

Metric	Description
`rest.bytes_received`	Total bytes fetched from the origin FDP (key bandwidth metric).
`rest.requests_issued`	Number of origin fetches (used to count request reduction).
`app.bytes_served`	Total bytes delivered to clients (monitors consistency).
`origin_cache.hits/misses`	Used to validate data reuse and caching efficiency.
`response_cache.hits/misses`	Used to measure endpoint-level caching efficiency.

Baseline Definition

Bandwidth reduction is measured relative to a no-cache baseline using the same access pattern:

- `origin_bytes_baseline` = bytes fetched from origin in baseline conditions (with caching disabled).
- `origin_bytes_current` = bytes fetched during actual execution with caching enabled.

Formula

$$\text{Bandwidth Reduction (\%)} = 100 \times \left(1 - \frac{\text{origin\_bytes\_current}}{\text{origin\_bytes\_baseline}}\right)$$

A secondary KPI may also be computed based on origin request reduction:

Funded by the European Union

$$\text{Fetch Count Reduction } (\%) = 100 \times \left(1 - \frac{\text{rest.requests\_issued}}{\text{rest.requests\_issued\_baseline}}\right)$$

Validation Results

As for KPI3.1, the preliminary experiment simulates 10 analytical requests over 2 endpoints that access a 47MB FDP. The measurements are facilitated by ASG runtime support for service endpoints for observing the accumulated metrics.

Configuration	Data Transferred from Origin	Reduction vs Baseline
No cache (baseline)	470 MB	0%
Origin cache only	47 MB	90%
Response cache only	94 MB	80%
Both caches enabled	47 MB	*90%

Table 7 - KPI 3.2 Validation Results

With caching enabled, TEADAL achieves 4x to 10x less network data transfer, significantly exceeding the KPI requirement of a 20% reduction.

Conclusion: KPI 3.2 is fully achieved. TEADAL's ASG Runtime reduces cross-node data transfer by more than 80–90%, with no impact on the quality or correctness of analytical results. This confirms that TEADAL supports scalable, sustainable data sharing by minimizing bandwidth use and avoiding repeated data movement across federation nodes.

A full pilot-level KPI evaluation can be consolidated using telemetry captured by the monitoring subsystem, which exposes per-sFDP metrics including total bytes fetched per origin endpoint. This allows direct per-pilot validation of KPI 3.2 using runtime metrics instead of synthetic tests alone.

10.3.9 KPI 4.1 Creation of a highly usable framework (e.g., in terms of time to write rules), able to reduce the time to define privacy/confidentiality policies and configure the system in charge of ensuring them of 30%

The key element contributing to this KPI is the policy definition framework that has been mainly defined in WP3. Notably, this framework is composed of a notation, called DSPN, that allows non-technical users to specify the constraints in a way that could be transformed into Rego rules which, in turn, can be enforced by the OPA engine.

Referring to the KPI, based on the produced framework, the reduction of the time to define privacy/confidentiality cannot be easily computed as we need to compare the time to elaborate a DSPN-compliant diagram with the definition of the correspondent Rego rules. For

Funded by the European Union

this reason, it is reasonable to assume that the time needed to write rules with Rego that requires - in addition to the knowledge of the language - significant effort to configure a system in several parts, is significantly lower than the time to author a DSPN diagram. In fact, the usage of OPA requires the definition of rules that are able to intercept the incoming request, extracts the relevant elements from the body of the message, and compares the data with the authorization rules. All of these steps require several lines of code written in Rego, that must be placed in specific files. In addition, these rules must not interfere with the existing ones that regulate the access to other FDPs.

On this basis, without changing the general focus, the KPI 4.1 has been revised in the metric used for the validation. Instead of calculating the time reduction, assuming this time reduction can be taken for granted to the adoption of the DSPN, we have verified that the DSPN is really able to express the data sharing policies with respect to the adopted pilot cases.

Evaluation Methodology

The evaluation of the KPI has considered the pilot cases with the aim of verifying the ability of DSPN to capture the policy requirements according to the following steps:

- The pilot case owner expresses the requirements in a document using the natural language. These policies are related to an FDP that is associated to the pilot case and, in particular, refers to the Open API description
- A DSPN model is produced to represent the policies described in the document
- The Open API of the associated FDP is enriched with the policy according to the extension discussed in Deliverable D3.3.
- A Rego rule, or a set of Rego rules, have been automatically generated starting from the DSPN mode
- The FDP is tested according to these rules

Validation Results

A DSPN has been generated for all the pilot cases and included in the documentation provided in the GitHub repo for each of them. Based on the conducted experiment, it is clear that the approach is reasonable as most of the policies can be captured. Nevertheless, there are aspects that need to be improved especially in terms of data quality and energy. This requires additional effort to also consider these aspects as elements of possible policies that the data provider could be interested to define. It is also clear that, due to the extensibility of the approach, such an improvement could be implemented with limited effort.

10.3.10 KPI 4.2 Data catalog, based on at least 5 criteria related to the data (e.g., type, resolution) and at least 5 criteria related to friction/gravity (e.g., purpose of data usage, latency requirements) able to index all the data sets related to the pilot cases

This KPI is related to the implementation of the TEADAL Federated Catalogue. The Federated Catalogue allows describing the following digital assets:

- Dataset
- Federated Data Product
- Shared Federated Data Product
- Agreement

The overall story of the interaction between the user and the Federated Catalogue starts with the data owner, who adds descriptions of his own Datasets on the Data Catalogue. Then, another actor belonging to the same organisation selects a set of Datasets, binds them together in a coherent commercial offer, implements the offer through a Federated Data Product, and describes it on the Catalogue. The FDP description is then visible to all the

Funded by the European Union

users of the Catalogue. When the customer finds a suitable FDP, he can negotiate an Agreement with the FDP owner to make use of it. The creation of an Agreement is followed by the implementation of the Shared Federated Data Product, which is also described in the catalogue and is visible only by the two parties of the Agreement.

The criteria considered for this KPI does not refer therefore to a single metadata schemata, but to different schemas which are visible to different actors in different moments of the overall interaction.

Evaluation Methodology

The evaluation of this KPI is possible through the analysis of the metadata schemas of the so-called “asset types” implemented in the TEADAL Catalogue. For each asset type, we evaluated how many attributes belonged to the “data criteria” category and how many attributes belonged to the friction/gravity.

Validation Results

The Dataset asset type contains the following “data criteria”-related attributes:

- Estimated data size
- Last update date
- Nature of the dataset, choosing from “On demand”, “Static”, “Scheduled updates”, “Continuous/Stream”

These attributes allow providing users with a basic understanding of the expected data volume and the expected update periodicity, which is mandatory information while building applications relying on such data.

The Federated data product asset type instead allows selecting Dataset and Shared Federated data products which are used to expose the data product. The data-related attributes in this case are:

- Data model, which can be explained either via a machine-readable descriptor or by attaching a document with a human-oriented description;
- API descriptor, reporting the exposed functions to get the data or to interact with the FDP;
- General documentation, which allows uploading a document describing the data product.

Other attributes, related to the expected data usage such as the purpose or the legal constraints, have been moved to the policy descriptor, in accordance with the re-definition of the gravity and friction topics.

Funded by the European Union

11. ENERGY-AWARE TRUSTWORTHY DATA LAKE

11.1 Energy-Aware Architecture Validation

The TEADAL architecture integrates energy-awareness as a core design principle, ensuring that data sharing and processing across federated data lakes not only complies with privacy and confidentiality requirements but also minimises energy consumption.

The approach builds on the concepts introduced in Deliverable D3.2 and further consolidated in Deliverable D3.3, where the layered policy framework is extended to include energy sustainability aspects. Specifically, energy consumption is evaluated in relation to transformation operations that occur when moving from a Federated Data Product (FDP) to a Shared Federated Data Product (sFDP). These transformations may include filtering, anonymisation, encryption, or aggregation, each of which introduces a computational and energy cost.

The Energy-Aware Architecture introduces the following concepts, which are described in detail in Deliverable D3.2 and Deliverable D3.3:

- **Policy-driven optimisation:** policies defined through the Data Sharing Policy Notation (DSPN) and enforced via OpenAPI/OPA extensions can encode constraints that balance privacy/security requirements with energy efficiency.
- **Deployment plan generation:** the architecture includes mechanisms for generating Energy-Aware sFDP Deployment Plans (see Deliverable D3.3, Figure 35), which select the optimal data transformation pipelines and placement strategies considering both compliance and energy consumption.
- **Monitoring and validation:** energy metrics are collected at runtime to verify whether the selected deployment plan achieves the intended energy savings. For this purpose, TEADAL integrates a Prometheus plus Kepler monitoring stack.
    - **Kepler** (Kubernetes-based Efficient Power Level Exporter) estimates energy consumption per container or pod based on hardware counters and system telemetry.
    - **Prometheus** scrapes these metrics, enabling queries and dashboards to evaluate the energy footprint of data processing tasks.
    - This setup provides continuous feedback, allowing a clearer understanding of the TEADAL Node's energy consumption.
- **Scalability strategies:** components such as the Policy Decision Points (PDPs) and their scaling controllers dynamically allocate resources in a way that prevents unnecessary duplication and energy waste, while ensuring performance.

The validation of the Energy-Aware Architecture requires deploying pilot cases with varying workloads, measuring energy consumption during FDP to sFDP transformations, and confirming that the generated deployment plans, together with Prometheus/Kepler monitoring, achieve the expected reductions in energy footprint without compromising compliance or trust.

This validation demonstrates that the TEADAL architecture can effectively balance trustworthiness, privacy preservation, and energy efficiency, confirming its readiness for application across diverse domains (healthcare, finance, mobility, environment).

Supporting Flavoured Transforms in the ASG Toolchain.

First, we clarify the difference between regular and flavoured transforms in the ASG subsystem. Regular transforms are expected to be provided by the domain experts as dataframe manipulations and can have parameters relevant to the data domain that must be substituted at generation time and/or at runtime. Flavoured transforms are exposed as

Funded by the European Union

generic capabilities users can request, such as *apply compression*, *apply anonymisation*, or *apply data reduction*. The exact algorithm or calibration is not chosen by the user; instead, the system determines it dynamically based on deployment conditions and policies.

The ASG subsystem was adapted to support flavoured transformations as follows:

1. **ASG-tool changes.** The tool was modified to detect the special *flavour* parameter required by certain transforms. When this parameter appears, the tool does not attempt to match it against the sFDP endpoint parameters from the ASG specification YAML. Instead, it treats the transform as *flavoured* and inserts a function section into the endpoint spec with a placeholder value (e.g., *pass*). This placeholder only ensures the spec remains valid JSON; it is not interpreted. At deployment time, once conditions and policies are known, the system computes and inserts the actual values into the endpoint specs for runtime use.
2. **ASG-runtime changes.** The runtime library was extended with the algorithms behind flavoured transforms. This makes the functionality reusable across all sFDPs, similar to existing shared capabilities (e.g., fetching origin FDP data, caching, encoding). The motivation is that flavoured transformations control infrastructural aspects (e.g., enforcing policies, meeting optimisation goals) rather than data-domain semantics. By contrast, regular transforms belong to the data domain and their parameters must be explicitly specified by domain experts at generation or runtime.
3. **ASG-sFDP impact.** sFDP components are minimally affected. When new data servers are generated from ASG specs containing flavoured capabilities, the endpoint specs will include flavoured function call sections alongside the regular, possibly parameterised, ones. If new flavoured capabilities are introduced later, existing sFDPs must be regenerated or modified to benefit from them.

Data Reduction Flavours Implementation

We have implemented a data reduction functionality that represents any numerical signal as a sequence of linear functions. In order to decide the length of each line segment, we split the signal by its extrema points (local minima or local maxima). Sometimes a line may show a change in trend that does not cause a ma-ximum or minimum. In those cases, we split those segments one more time. In the end, we ensure that every segment meets a minimal segment-length condition.

By defining a short minimal segment length, we end up with more segments, which represent the signal with more fidelity, but reduce the compressibility and increase the computation complexity. By defining a long minimal segment length, the number of segments to process reduces drastically, increasing in the same way the compressibility while still maintaining a reasonable fidelity.

Based on this, we defined the flavours “high-fidelity” and “low-energy” for our data reduction functionality, which set the minimal segment length to 5 and 30, respectively.

In order to characterise the flavours of our transformation, we used the following metrics:

- Reduction Rate: The original size of the dataset minus the final size of the dataset divided by the original size of the dataset.
- Information Loss: Difference of the relative entropy between the original and the transformed signal.
- Energy Consumption: Energy consumption of the transformation relative to the original size calculated for a machine of 100 Watts of power consumption.

Funded by the European Union

We tested the transformation with a representative dataset of normalized data which showed the following average performance.

	high-fidelity	low-energy
Reduction Rate	49.99 %	86.11 %
Information Loss	-1.98 %	-8.13 %
Energy Consumption	624.47 J/MB	298.64 J/MB

Table 8 - Transformations Average Performance

Application to the use case

Additionally, we evaluated how the transformation behaves with TEADAL's medical use case. From the six available datasets, we selected only the numerical columns (75 out of 117) and applied both transformation variants to empirically test their performance.

	high-fidelity	low-energy
Reduction Rate	8.19 %	99.906 %
Information Loss	-6.1 %	-43.9 %
Energy Consumption	1926.4 J/MB	145.8 J/MB

Table 9 - Evaluation with Medical Use Case

When applying the transformation on the medical datasets, the difference between the two flavours is larger in all the metrics. The reason for this is the synthetic nature of the medical dataset, which populated its variables with values selected randomly. In essence, that's equivalent to having a trendless signal with white noise added.

In the high-fidelity case, the algorithm tries to maintain that noise represented, which produces a low reduction rate and a high energy consumption. In the low-energy flavour, the algorithm practically gets rid of all of the noise and only conserves the trendless line. That explains achieving almost a 100 % reduction rate and a 40 % loss in the information, which the algorithm does not know was just noise.

11.2 Results

In this section, we show the benefits of the proposed energy-aware architecture for sFDP creation and deployment in the health use case.

To do so, we have tested and compared different scenarios.

- SCENARIO 1 - Baseline: The deployment plan is generated without using the energy-aware architecture. This means that the transformation library does not allow the definition of multiple flavours. Accordingly, no flavour selection and reuse decisions are taken, and the association between the data pipeline transformations and the execution nodes are performed randomly between the available and feasible nodes. This scenario is represented by the /persons_reduced endpoint, which processes the complete medical dataset without any optimization, corresponding to how a traditional sFDP without energy-awareness would operate.

Funded by the European Union

- SCENARIO 2 - Energy Assessment: In this scenario we show how the energy monitoring enables to collect information about the environmental impact of the different components involved in the data sharing pipeline. Kepler (Kubernetes Efficient Power Level Exporter) is deployed on the TEADAL Node to measure pod-level power consumption in real-time. Data is collected through Prometheus and visualized in Grafana, allowing comparison of energy profiles across different optimization strategies. This monitoring infrastructure enables quantification of the actual energy savings achieved by the energy-aware architecture.
- SCENARIO 3 - Energy-Aware Adaptation: Starting from the data provider/data consumer agreement, the deployment plan will be generated using the Energy-Aware architecture. Multiple flavours are available for the dataset reduction transformation, with relative metadata. The approach will thus perform a selection between these flavours based on energy requirements. This scenario is demonstrated through two endpoints: /persons_reduced_less (aggressive reduction for maximum energy savings) and /persons_reduced_more (balanced reduction preserving more data fidelity). The key difference from SCENARIO 1 is that the transformation library is enhanced to support multiple execution flavours through the *flavour* parameter, allowing dynamic adaptation to energy requirements without code changes.

11.2.1 Implementation Details

The energy-aware sFDP implementation leverages the ASG Runtime framework with configurable transformation pipelines. The system architecture consists of:

Configuration Management: The sFDP uses a comprehensive .env configuration file that controls:

- Caching strategies (LRU, disk, or Redis-based) for both origin and response data;
- Logging levels and output formats (rich console, plain text, or JSON);
- HTTP client parameters (timeout, retry logic, backoff strategies);
- Transform path resolution for the energy-aware reduction algorithms.

Transformation Pipeline: The energy-aware dataset reduction is implemented through:

- A generic reduce_dataset transformation in the sFDP service layer that accepts a flavour selection parameter (*flavour*), enabling dynamic switching between different dataset reduction strategies;
- A backend implementation in the ASG Runtime library that processes data according to the selected flavour;
- Integration with Kepler energy monitoring to measure real-time power consumption.

Medical Dataset Context: The implementation was validated using medical patient records with specific characteristics:

- Data nature: The dataset contains random-generated data with inherent noise
- High-fidelity flavour: Replicates all data including noise, maintaining statistical mean values
- Low-energy flavour: Removes noise and redundant information while preserving essential medical data
- Trade-off: For this medical dataset, the "low-energy" mode provides cleaner data while consuming significantly less energy, as early noise removal benefits both data quality and processing efficiency

Deployment Configuration: The containerized sFDP is deployed on the TEADAL Node with:

Funded by the European Union

- Resource limits ensuring predictable performance (200m-500m CPU, 256Mi-1Gi memory);
- Multiple replicas for high availability;
- Connection to MinIO object storage for data persistence.

11.2.2 Energy Consumption Evaluation

To evaluate and compare the energy consumption of the sFDP, Kepler (available in the TEADAL Node) was used to make the TEADAL Node aware of all energy consumption. With this setup, it was possible to generate graphs showing the energy consumption of the sFDP pods in watts.

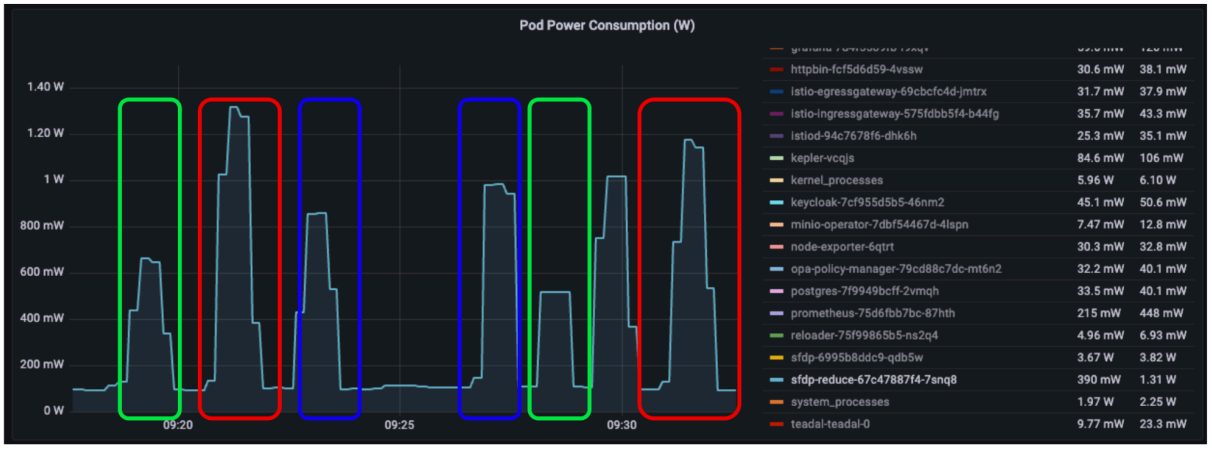


Figure 29 - Blue: sFDP base; Green: sFDP less energy; Red: sFDP more energy.

Figure 29 compares pod-level power consumption across three scenarios: sFDP base (Blue), sFDP less energy (Green), and sFDP more energy (Red). Each colored segment represents a different reduction level being tested. The "less energy" mode can reduce energy usage by up to 50% compared to the base, while the "more energy" mode shows an increase of up to 30%.

Baseline Comparison: The baseline (blue) is essential to quantify actual energy savings and validate that the energy-aware architecture provides measurable benefits over traditional sFDP deployments. Without this reference point, we cannot demonstrate the effectiveness of the optimization strategies or measure the energy-quality trade-offs. The baseline represents existing production systems, making our results directly comparable and enabling organizations to make informed decisions about adopting energy-aware optimizations.

We can clearly observe distinct power envelopes: higher infrastructure-level optimization settings (less energy mode) lead to visibly lower sustained wattage during workload execution, whereas the baseline mode peaks at over 1 W. These results confirm that the infrastructure-level dataset reduction mechanism has a measurable impact on energy draw, providing a first visual baseline for correlating reduction aggressiveness with power efficiency.

Funded by the European Union

12. CONCLUSIONS

The deployment and validation of the TEADAL platform across multiple pilot cases have successfully demonstrated the feasibility and robustness of the trustworthy, energy-aware federated data lake architecture envisioned by the project.

All pilots, spanning healthcare, mobility, smart viticulture, industry, regional sustainability and financial data governance, confirmed that the TEADAL Node can be consistently deployed, monitored and managed through an automated CI/CD workflow. The platform proved capable of supporting complex, cross-domain federations with built-in privacy enforcement, verifiable trust mechanisms and energy-aware analytics.

Through the combination of advanced components such as AI-DPM, Advocate, ArgoCD, OPA, Kepler and the Data Catalogue, the project delivered an integrated ecosystem that reduces the operational complexity of data sharing, enhances transparency and traceability, and promotes interoperability across heterogeneous infrastructures.

The validation phase, conducted in close collaboration with all technical partners and KPI owners, confirmed that TEADAL meets its core objectives, achieving measurable gains in automation, privacy management, energy efficiency and federation scalability. These outcomes highlight TEADAL's contribution to the European Data Strategy and to emerging Data Spaces initiatives, offering a reusable and extensible foundation for trustworthy data collaboration across sectors.

In conclusion, TEADAL has reached a mature stage of deployment and validation, providing a concrete, demonstrable reference for future adoption in industrial, scientific and public-sector data ecosystems. The lessons learned from this deliverable will inform the project's final reports and exploitation activities, reinforcing TEADAL's role as a key enabler of next-generation, privacy-preserving, energy-aware data infrastructures.

Funded by the European Union

REFERENCES

[1] European Data Sharing Evaluation - Pilot Questionnaire. Microsoft Forms. Available at: https://forms.office.com/pages/responsepage.aspx?id=K3EXCvNtXUKAjjCd8ope645fmixa2L5IgLUTaPSE-yBUOVNBRVZNWktFTThGR1JPVTFGWjlKTkNJWC4u&route=shorturl

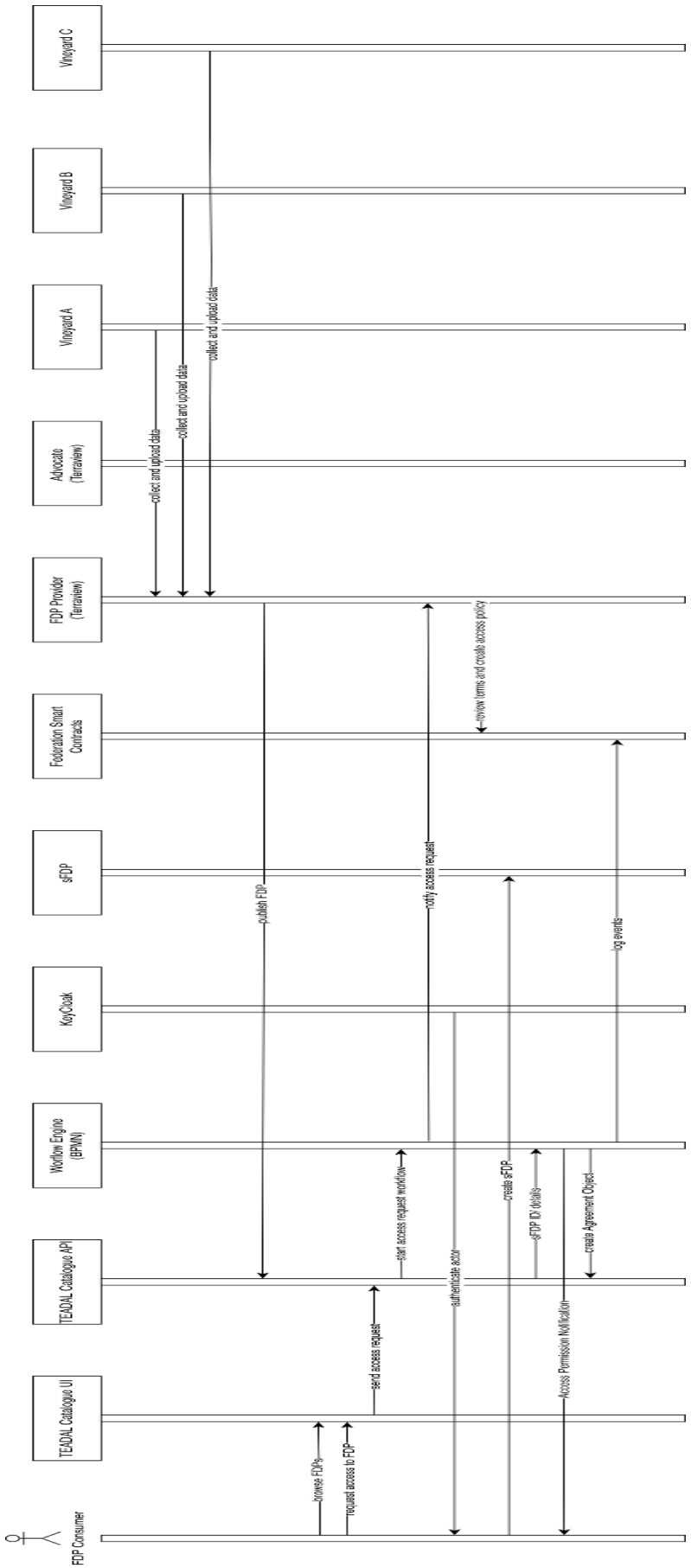
[2] European Data Sharing Evaluation - External Questionnaire. Microsoft Forms. Available at: https://forms.office.com/pages/responsepage.aspx?id=K3EXCvNtXUKAjjCd8ope645fmixa2L5IgLUTaPSE-yBUNlhJQ01aSkxFUVdHMjIyTFVVR043VUtKMi4u&route=shorturl

Funded by the European Union

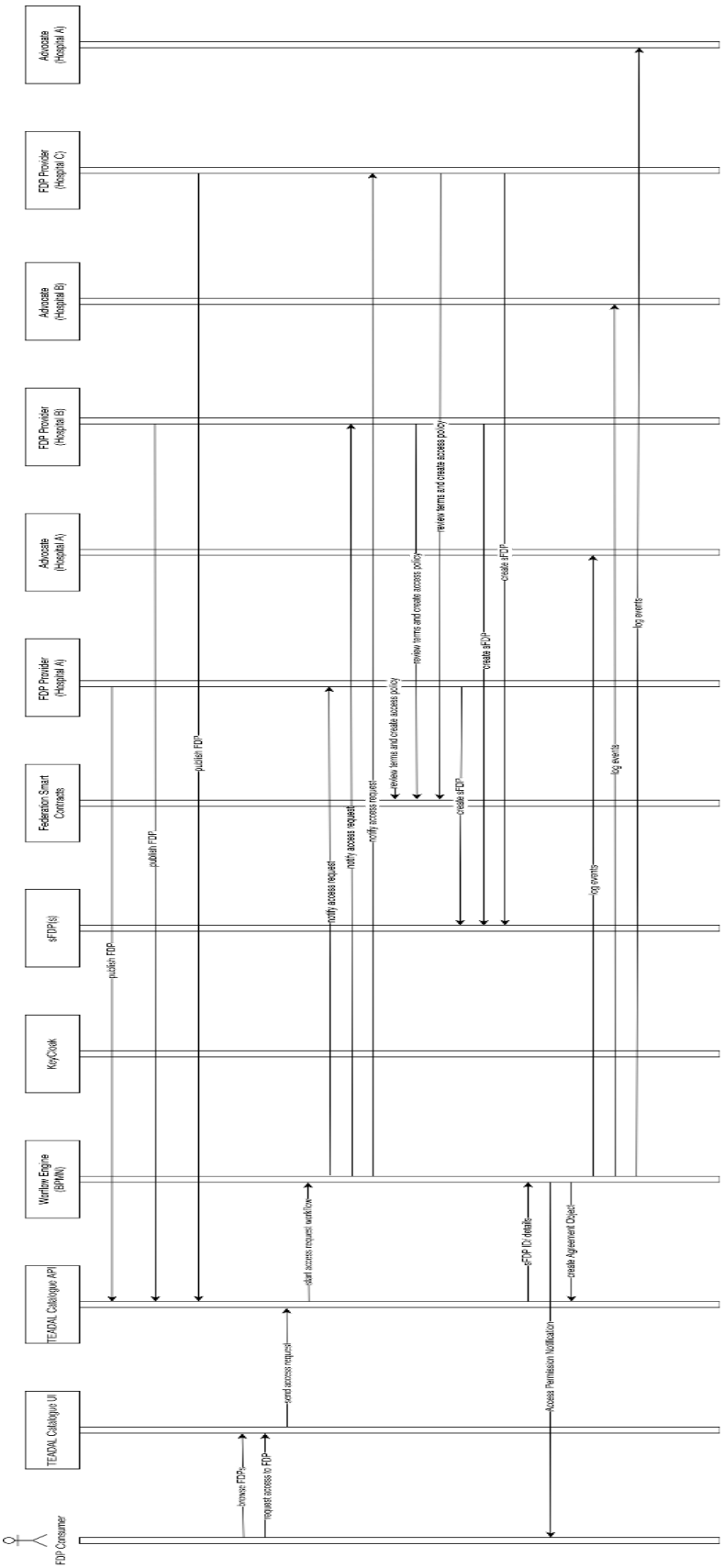
APPENDIX A

Please find below the sequence diagrams of joining the use case pilots into the federation using TEADAL, corresponding to section 10.3.4 (KPI 2.1). The figures are presented in a larger format here to improve readability in this appendix.

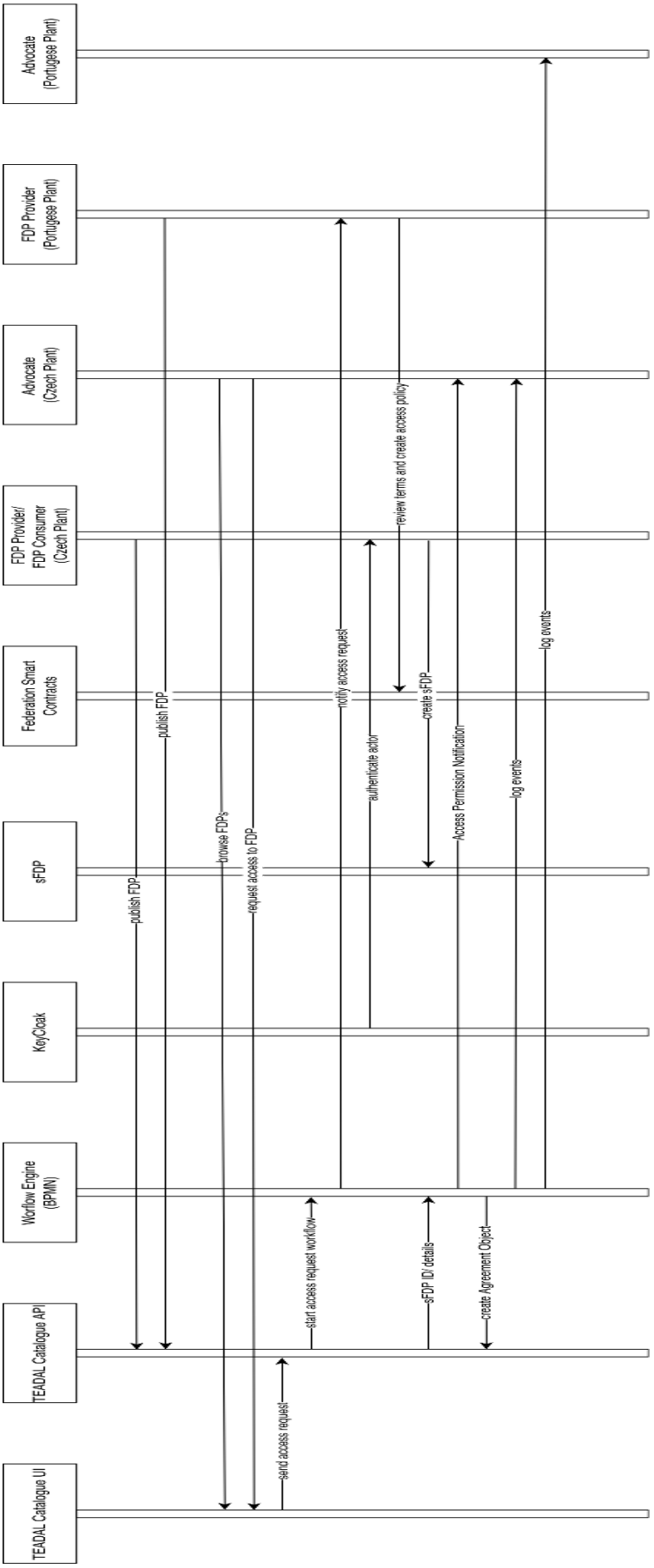
Funded by
the European Union



Funded by the European Union



Funded by the European Union



Funded by the European Union