Introducing TEADAL

Pierluigi PLEBANI

POLIMI

WORKSHOP ON CO-DESIGNING DATA SPACES: EMPOWERING USABILITY

March 9th, 2023
Bucharest, Romania and Hybrid mode

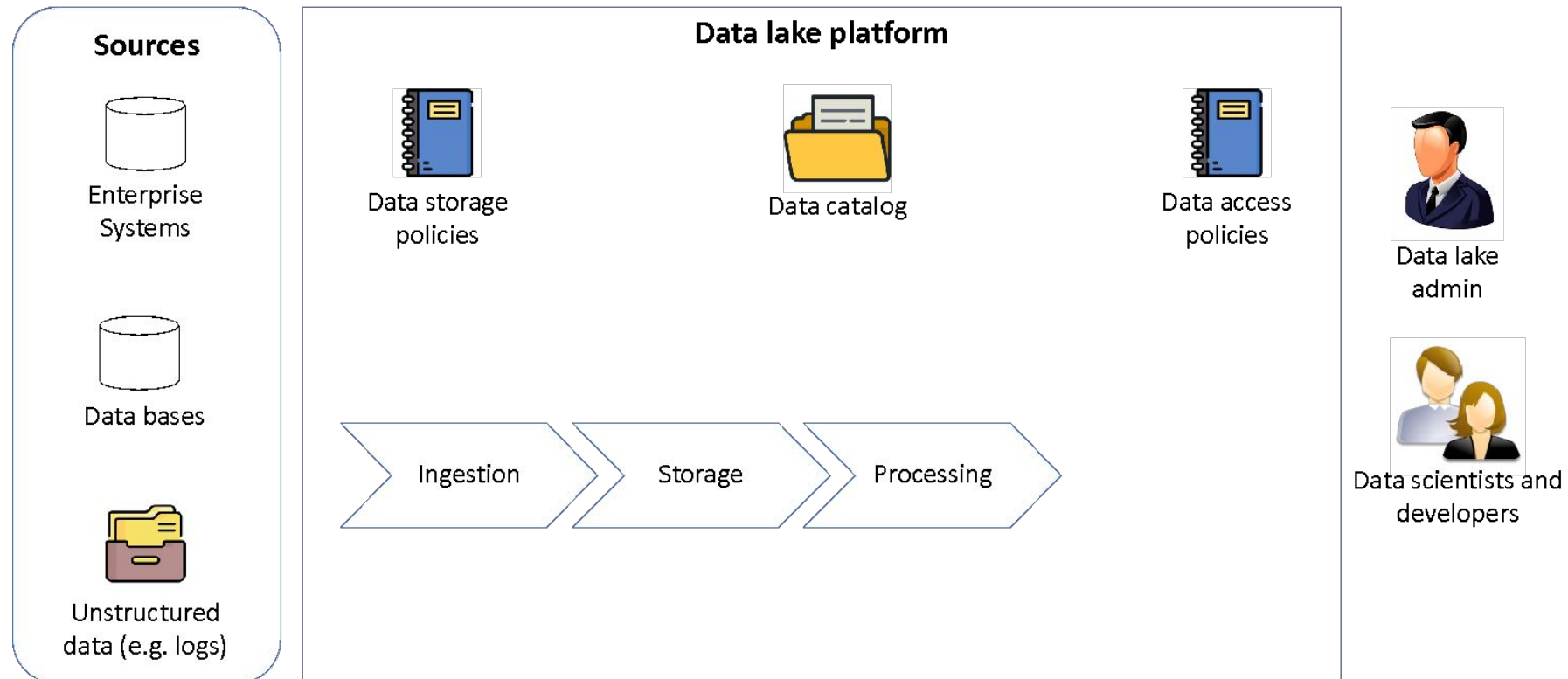*WWW.TEADAL.EU*

Main ambition

To provide <u>key cornerstone technologies</u> that will enable the creation of <u>trustworthy mediatorless federations of data lakes</u> spanning the <u>cloud-edge continuum</u> and, as dynamic constellations of different organizations, to improve a <u>trusted, verifiable, and energy-efficient data sharing</u> as a key driver for fostering a Sustainable European Digital Single Market.
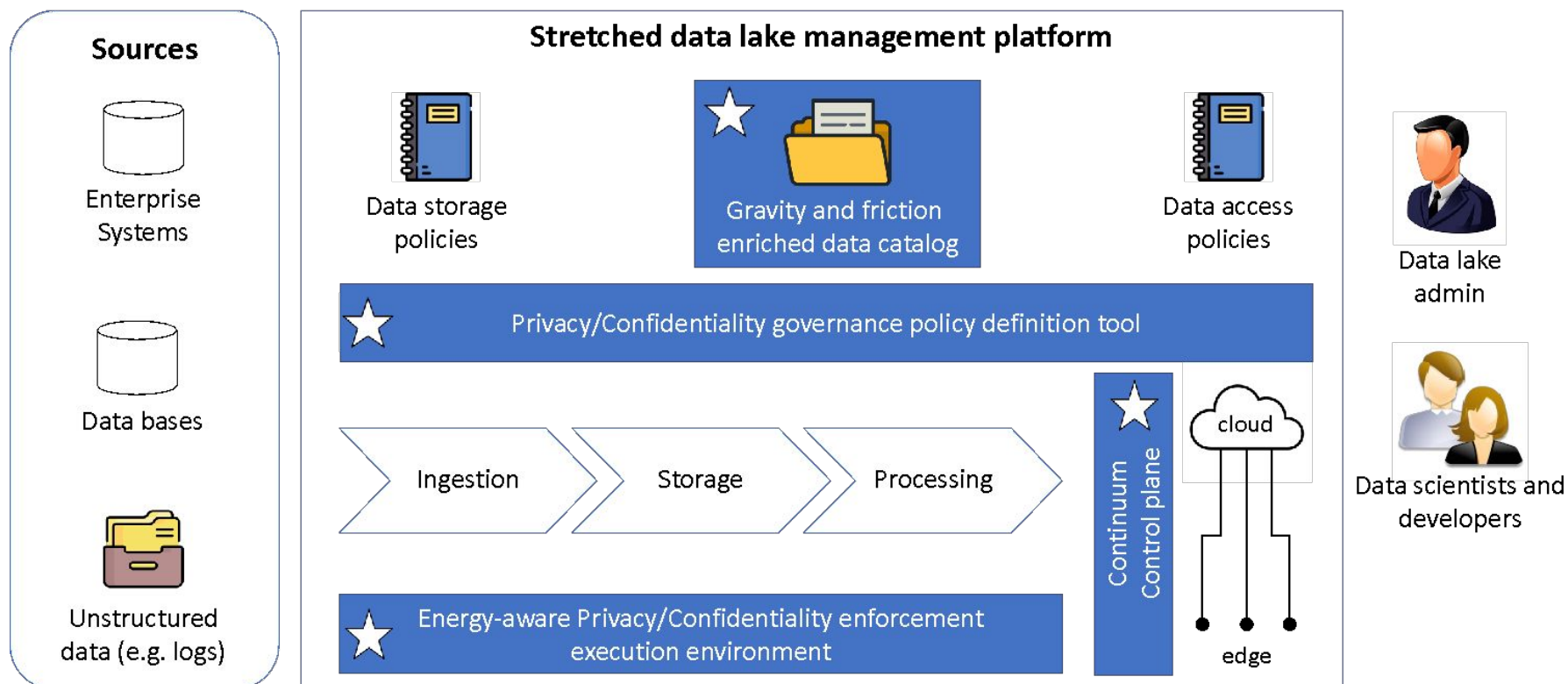
Consortium

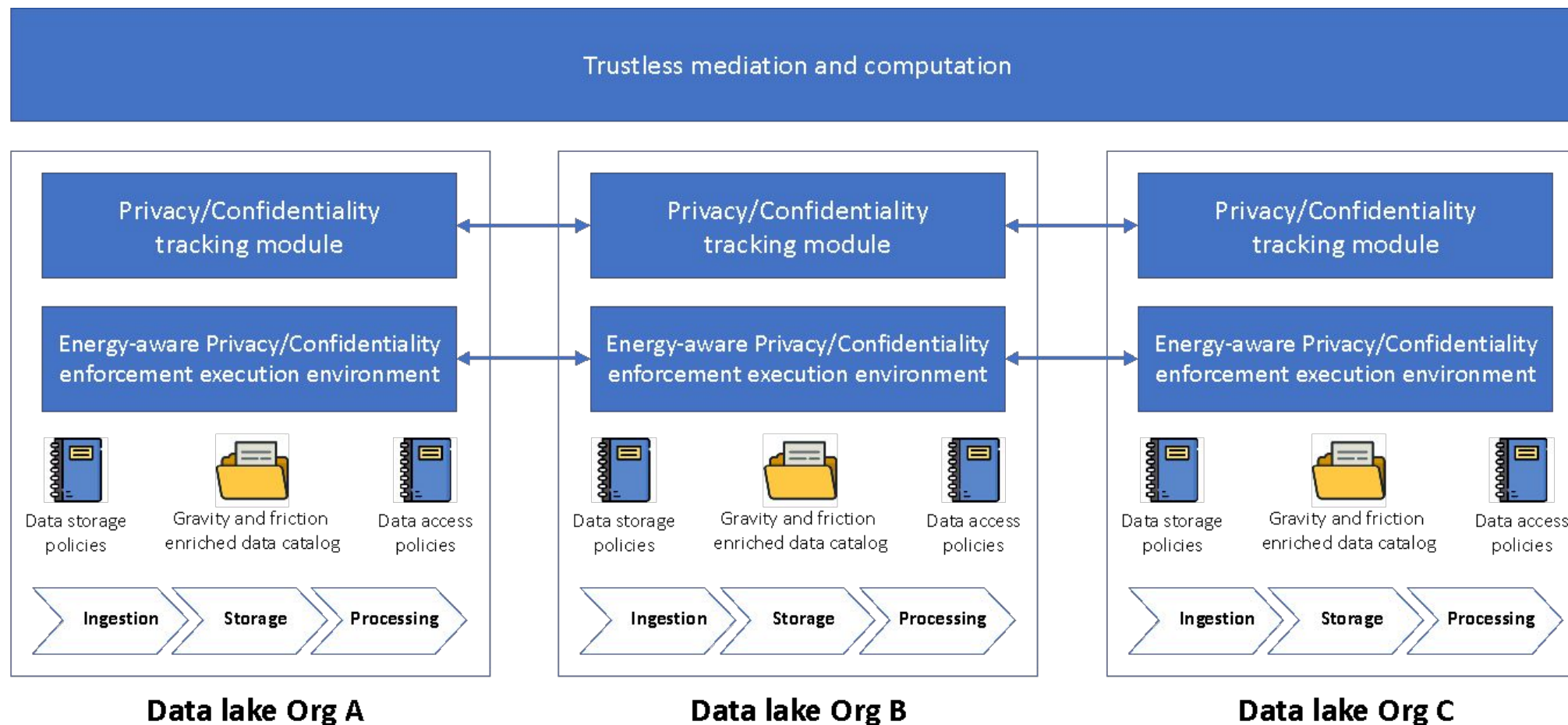| No.	Participant organization name
1	UBIWHERE LDA (Coordinator)
2	POLITECNICO DI MILANO
3	CYBERNETICA AS
4	CEFRIEL SOCIETA CONSORTILE A RESPONSABILITA LIMITATA
5	IBM ISRAEL - SCIENCE AND TECHNOLOGY LTD.
6	TECHNISCHE UNIVERSITAET BERLIN
7	ING BANK N.V.
8	MARINA SALUD, S.A.
9	UNION INTERNATIONALE DES TRANSPORTS PUBLICS
10	AZIENDA METROPOLITANA TRASPORTI E SOSTA CATANIA SPA
11	TECHNISCHE UNIVERSITAET WIEN
12	ALMAVIVA - THE ITALIAN INNOVATION COMPANY SPA
13	MARTEL GMBH
14	TERRAVIEW GMBH
15	ERT TÊXTIL PORTUGAL, S.A.
16	FUNDACIO PRIVADA I2CAT, INTERNET I INNOVACIO DIGITAL A CATALUNYA
17	BOX2M ENGINEERING SRL
18	REGIONE TOSCANA

From data lake …

… to stretched data lake …

Sources
- Enterprise Systems
- Data bases
- Unstructured data (e.g. logs)

Stretched data lake management platform

- Data storage policies
- Gravity and friction enriched data catalog
- Data access policies

Privacy/Confidentiality governance policy definition tool

Ingestion → Storage → Processing

Continuum Control plane

cloud

edge

Energy-aware Privacy/Confidentiality enforcement execution environment

Data lake admin

Data scientists and developers

… That can be federated

TEADAL

Data scientists and developers

Trustless mediation and computation

| Privacy/Confidentiality tracking module | Privacy/Confidentiality tracking module | Privacy/Confidentiality tracking module |

| Energy-aware Privacy/Confidentiality enforcement execution environment | Energy-aware Privacy/Confidentiality enforcement execution environment | Energy-aware Privacy/Confidentiality enforcement execution environment |

Data storage policies   Gravity and friction enriched data catalog   Data access policies

Ingestion → Storage → Processing

Data lake Org A

Data storage policies   Gravity and friction enriched data catalog   Data access policies

Ingestion → Storage → Processing

Data lake Org B

Data storage policies   Gravity and friction enriched data catalog   Data access policies

Ingestion → Storage → Processing

Data lake Org C

S/T methodology - pillars

Data gravity/friction

Trustworthy data sharing

Energy-efficient use of data

Data gravity/friction
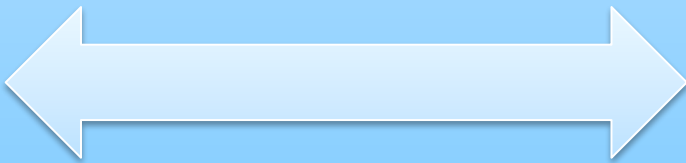
Forces that will guide the data placement
- Along the continuum
- Among the members of the federation

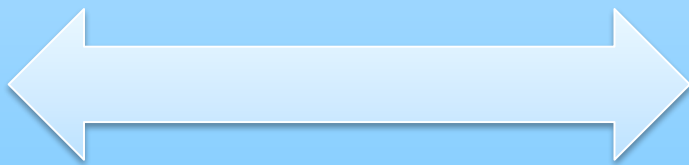They will influence the development of innovative data catalogs where
- Privacy/confidentiality is related to the data location

Trustworthy data sharing

Organizations need to trust each other when sharing data
It is fundamental to balance between
- Data sovereignty
- Effective/Significant data analysis

We propose a Federated data lake architecture that simplifies the policy definition, enforcement, and tracking

S/T methodology - pillars

Energy-efficient use of data
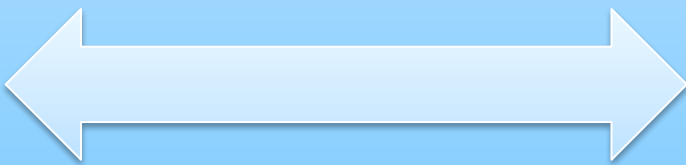
Sustainability must be a first-class requirement when developing our solutions

In all the aspects:
- Ingesting data
- Storing data
- Processing data

Keep always in mind that any proposed solution must be energy efficient somehow

Cloud Federation, common European Data Spaces and AI

Health	Industrial & Manufacturing	Agriculture	Culture	Mobility	Green Deal	Security	Public Administration	Media

High Value Datasets From public sector

- Driven by stakeholders
- Rich pool of data of varying degree of openness
- Sectoral Data Governance (contracts, licenses, access rights, usage rights)
- Technical tools for data pooling and sharing

Marketplace for Cloud to Edge based Services

Cloud services meeting high requirements for data protection, security, portability, interoperability, energy efficiency

Federation of Cloud & HPC Infrastructure & Services

Cloud stack management and multi-cloud / hybrid cloud, cloud governance

Edge Infrastructure & Services

SaaS (Software as a Service)
Software, ERP, CRM, data analytics

PaaS (Platforms as a Service)
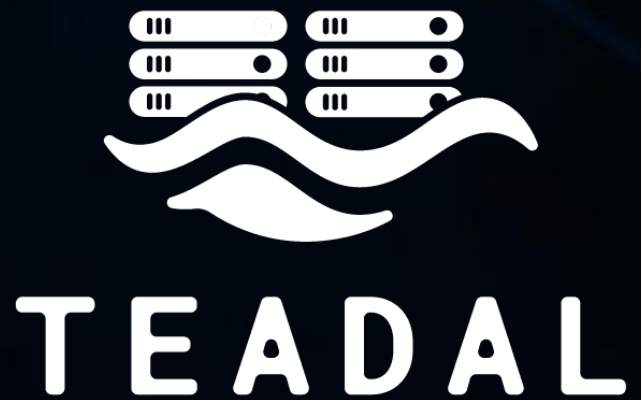*Smart Interoperability Middleware*

IaaS (Infrastructure as a Service)
Servers, computing, OS, storage, network

High-Performance Computing

AI on demand platform

AI Testing and Experimentation Facilities

Scenarios and requirements

Alessio CARENINI

Cefriel

WORKSHOP ON CO-DESIGNING DATA SPACES: EMPOWERING USABILITY

March 9th, 2023
Bucharest, Romania and Hybrid mode

WWW.TEADAL.EU

TEADAL Pilots

TEADAL Pilots represent domains where data sharing is currently difficult for different reasons:

- Regulations (like GDPR or National Access Points for mobility)

- Complex processes requiring data sharing between multiple actors

- Constraints on where data processing must take place

- Constraints on where data should be processed

- Competition between companies requiring complex business policies

The aim of the first iteration of the pilots description and requirements analysis was provide initial insights for the development of TEADAL components

Pilot #1: Evidence-based Medicine

Objective: allow clinical studies to be executed on data coming from multiple hospitals dealing with GDPR patients' consent

Why is it hard?

- GDPR requires patients' consent for each clinical study

- Doctors want to perform data exploration, so data processing requirements are not known a-priori

- Not only data sharing, it requires the implementation of a complex process based on data sharing

- Copying data to another hospital may be forbidden

Which data is shared?

- Clinical data (SQL data - 100GB)

Clinical study with TEADAL
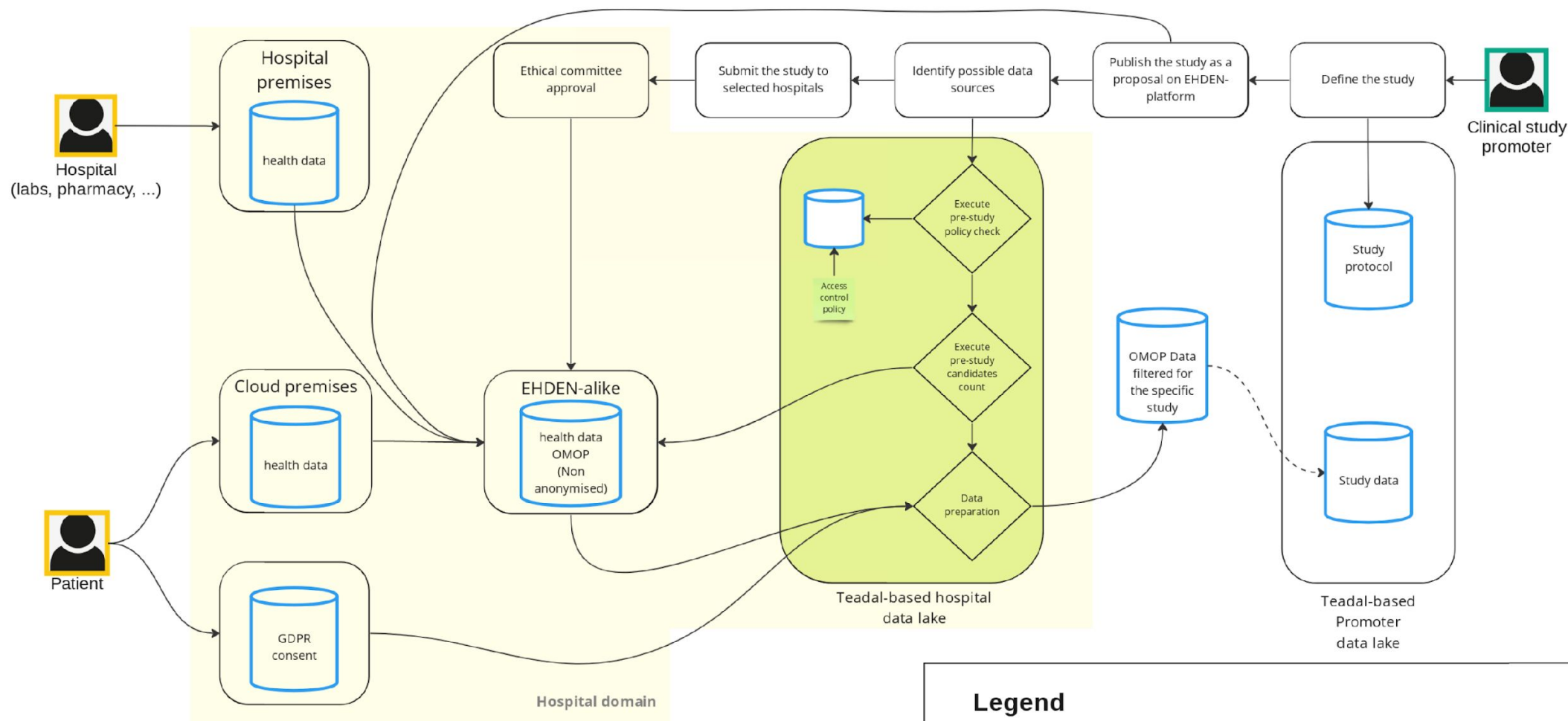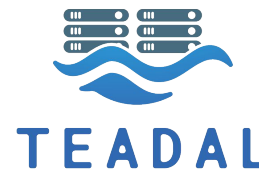
Pilot #2: Mobility

Objectives:

- Allow compliance to EU regulation related to mobility (National Access Points)

- Allow mobility providers to leverage the availability of mobility data on NAPs

Why is it hard?

- Competition between actors usually hinders data sharing (especially delays data)

- EU is forcing mobility providers to share data

- Italy chose a two-layer system (National Access Point and Regional Access Point) to delegate data collection to Regions

Which data is shared?

- Public transport timetables (GTFS updated twice a year usually)

- Real-time information about road traffic and public transport delays (GTFS-RT)

EU regulations compliance through TEADAL datalake
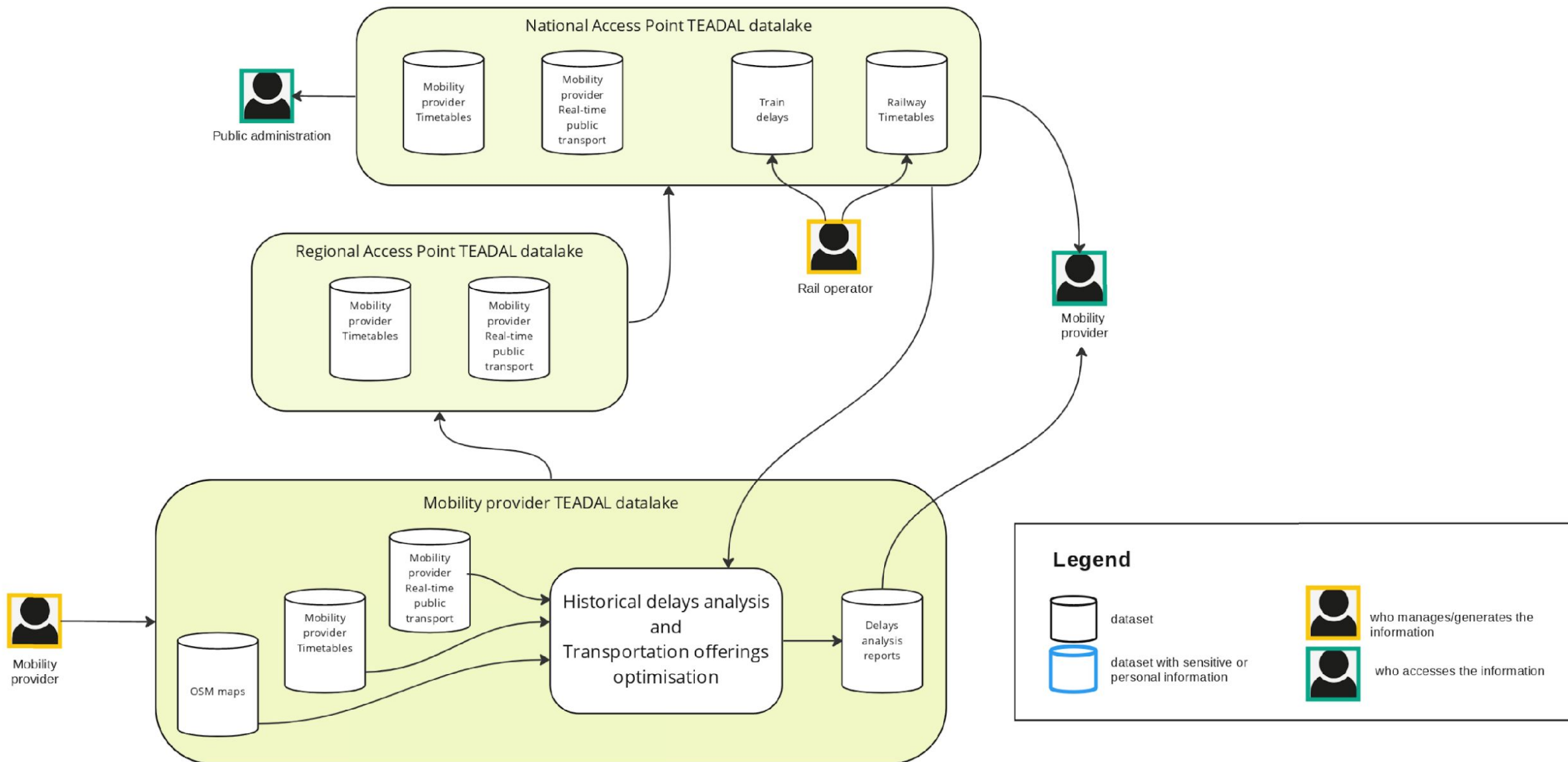
# Pilot #3: Smart viticulture

Objective: Allow vineyard owners to selectively decide who to share their data with, while benefiting from ML-based estimations created using data from all the vineyards

Why is it hard?

- Each vineyard must remain in control of its data

- Policies must be set-up to allow controlled data sharing to selected actors

- (Optional) vineyard data must be stored only on edge

Which data is shared?

- Satellite (images – 260GB update every 8 days)

- Sensors data (JSON - 2GB updated every hour)

- Weather data (binary – 1GB updated every hour)

- Drones and ground-generated data (JSON – 50GB updated every hour)
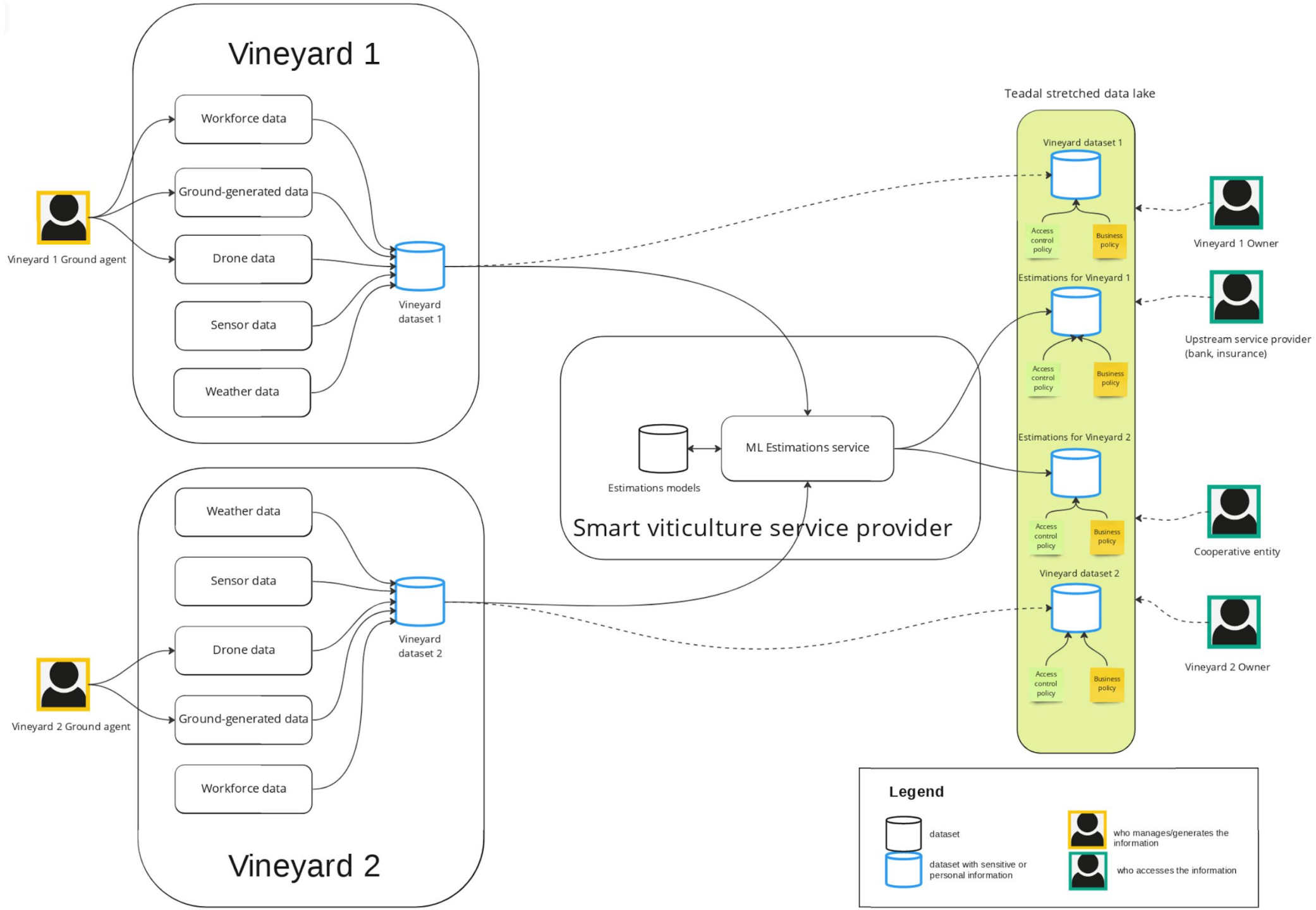
Pilot #4: Industry 4.0

Objective: Allow processing KPIs over data from different manufacturing plants, each with its own data model, and enforce authorisation policies

Why is it hard?

- Processing over large amounts of data

- Different data models

Which data is shared?

- Plant production data (SQL data – 100GB updated every 30 minutes)
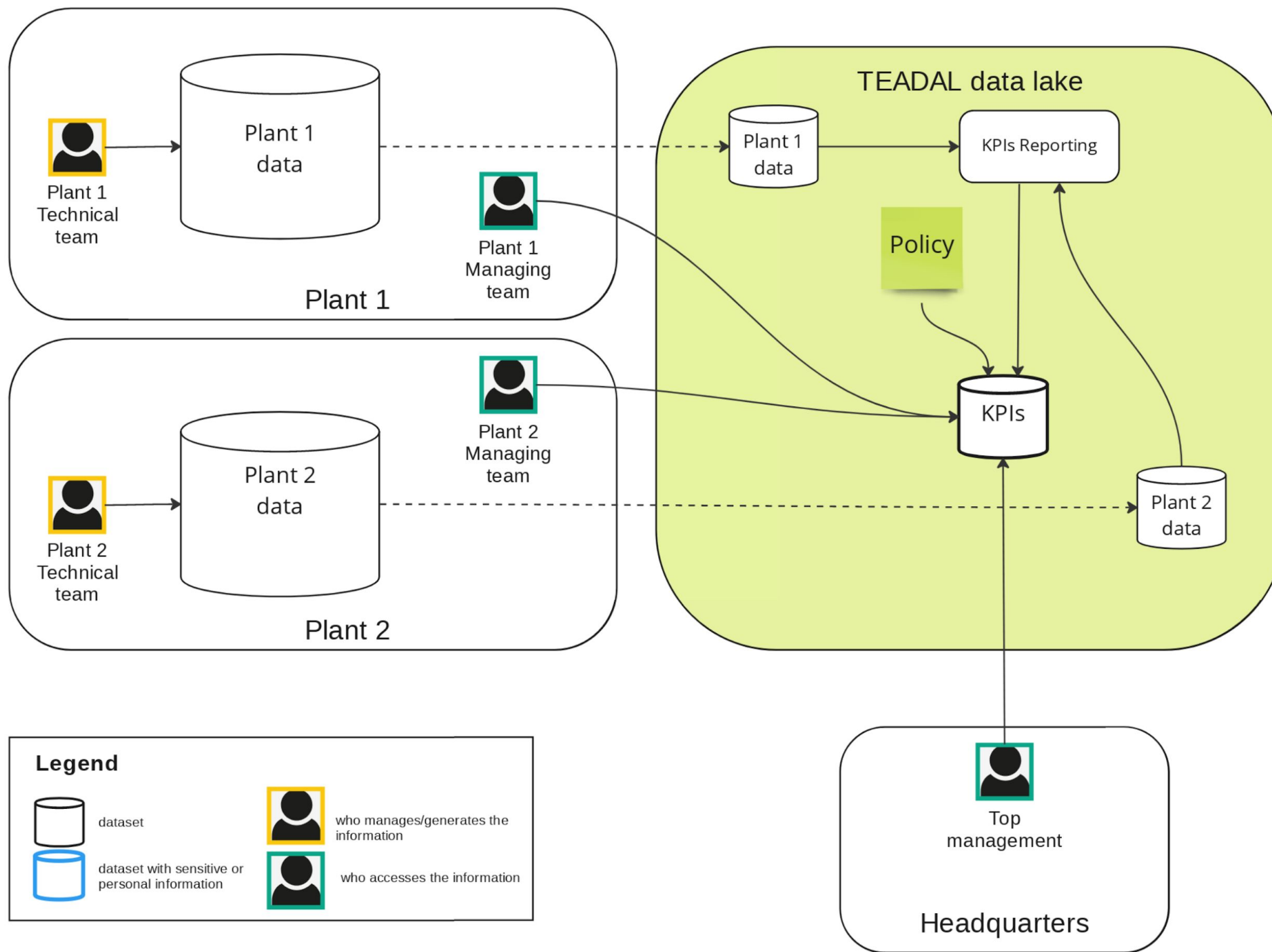
Pilot #5: Environmental sustainability

Objective: Allow reconstructing static and dynamic energy files for public and private buildings and map energy efficiency and air quality patterns by territorial area

Why is it hard?

- Correlate building energy profiles managed by public administrations with sensor data from environment and energy consumptions monitoring deployed by private companies

- Private companies can only benefit of the results of the analysis without accessing to raw data from public administration

Which data is shared?

- Buildings' energy performance certificates (XML - 1M elements updated daily)

- Registry of buildings' thermal and air conditioning systems (SQL – 7M elements updated daily)

- Environmental sensors data (JSON – 400K elements collected in real time and consolidated daily)

- Buildings' energy consumption data (JSON – 400K elements collected in real time and consolidated daily)

Public-Private federation for environmental sustainability
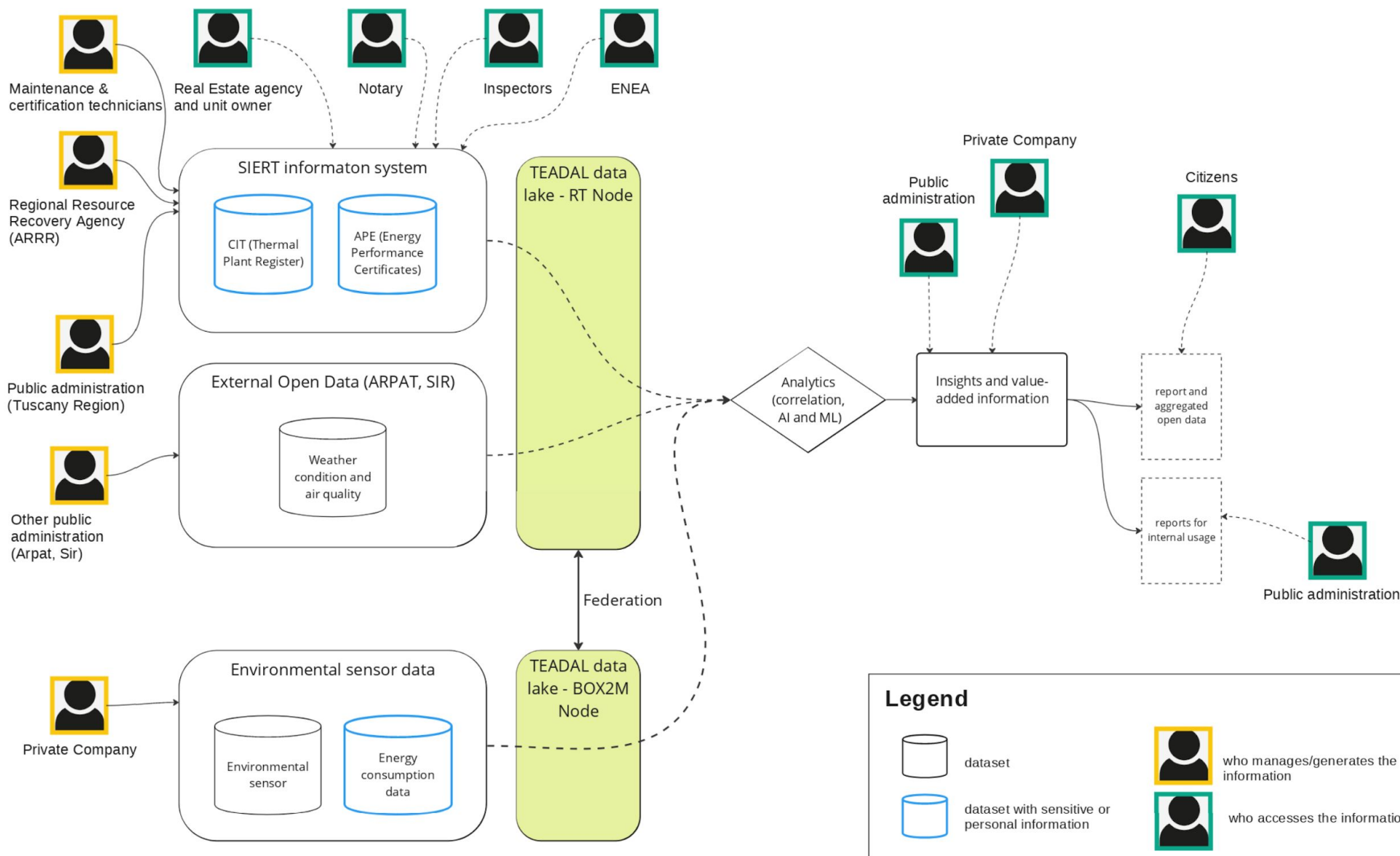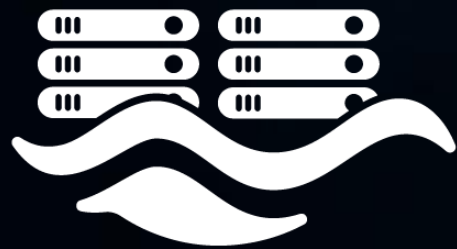
Main features of the pilots

Feature	#1 Medicine	#2 Mobility	#3 Viticulture	#4 Industry 4.0	#5 Env. Sustainability
Federation (Friction)	Y	Y	N	N	Y
Cloud-to-edge (Gravity)	N	N	?	Y	Y
Federated identities	Y	Y	N	N	Y
Efficiency	N	Y	N	N	N
Privacy preservation	?	N	N	N	N
Data tracking	Y	N	N	N	N
GDPR policies	Y	N	N	N	Y
Business policies	Y	Y	Y	Y	Y

Meet the TEADAL Consortium

© Copyright TEADAL 2022-2025
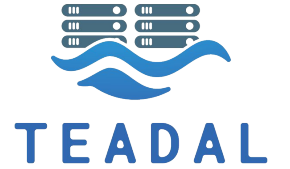
THANKS

TEADAL.EU   @TEADAL_eu   @TEADAL