TEADAL

Scenarios and requirements

Alessio Carenini

Cefriel

TRUSTEE Workshop

Bucharest, 9/3/2023

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 stored
- 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

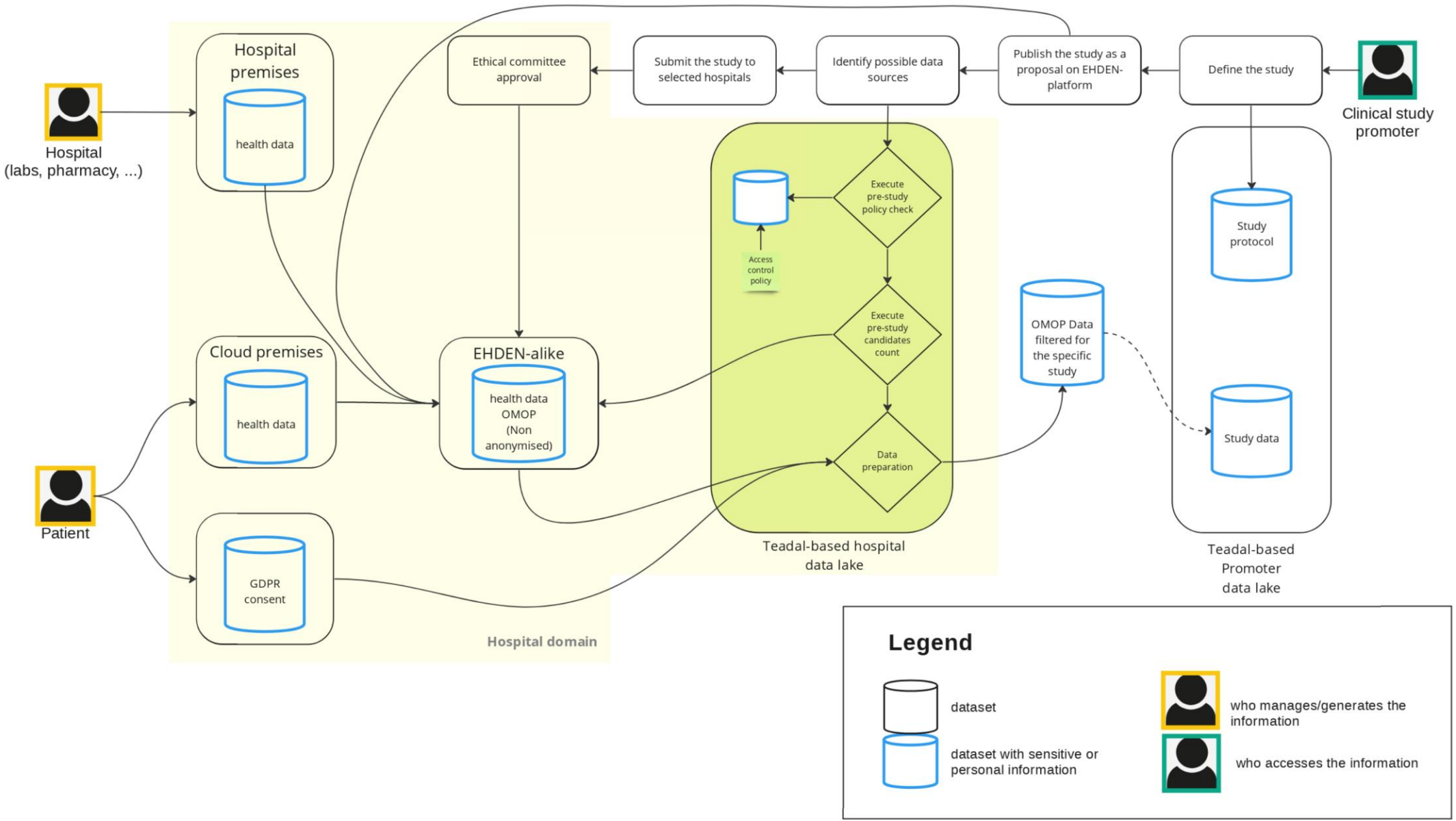
Why is it hard?

- 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

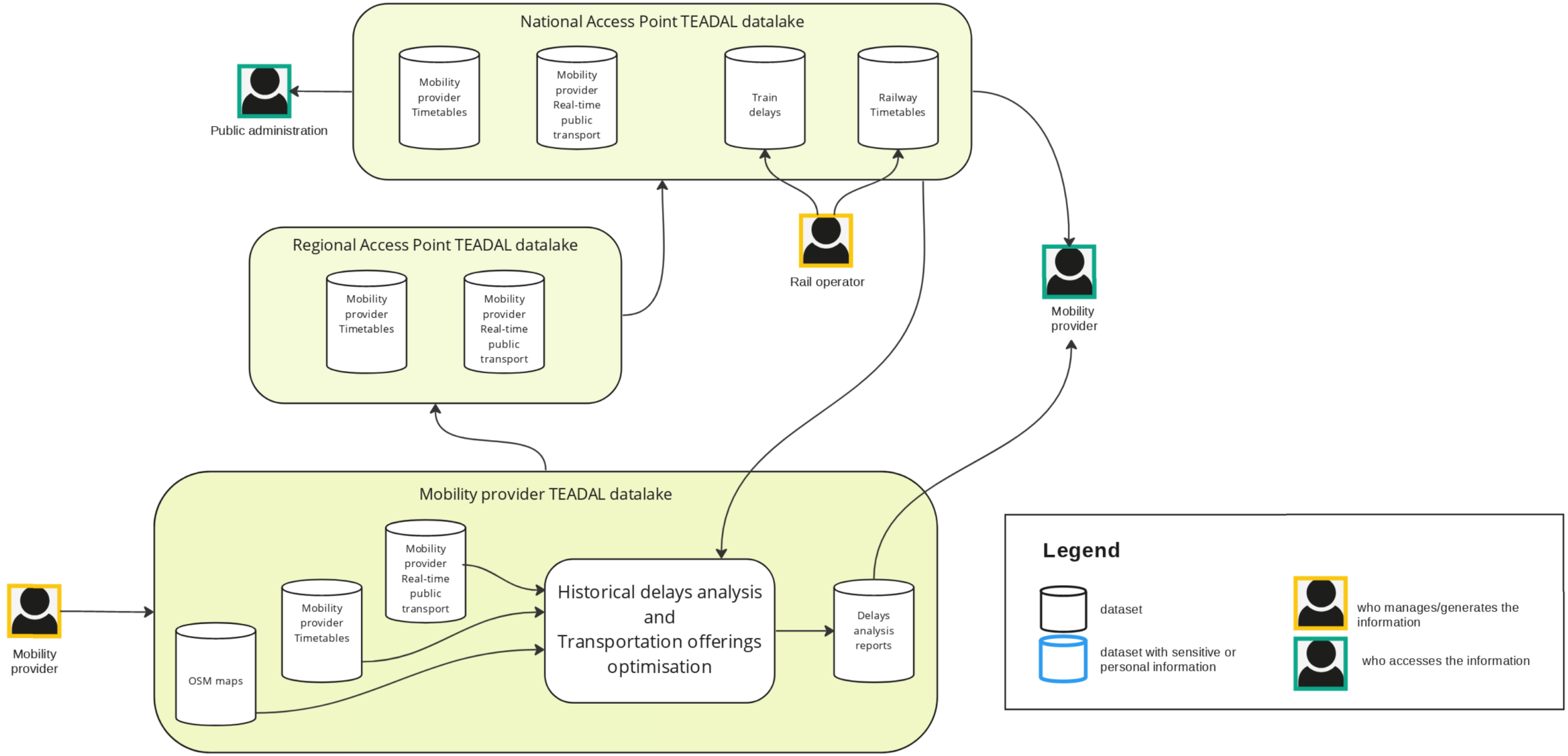
Why is it hard?

- 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

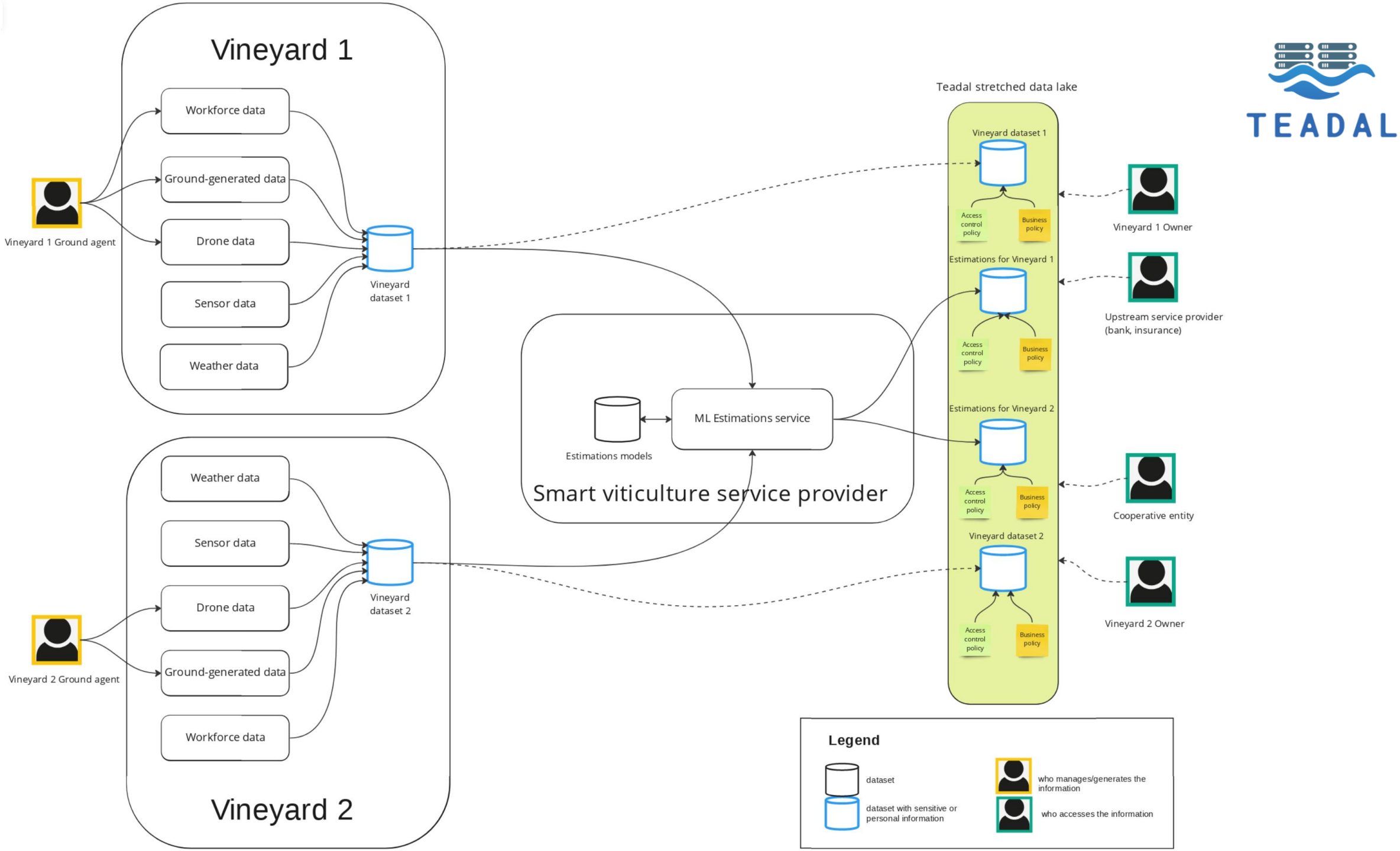
Objective: Allow vineyard owners to selectively decide who to share their data with, while benefiting from ML-based estimations created using data from all the vineyards

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)


Vineyard 1
Workforce data
Ground-generated data
Drone data
Sensor data
Weather data
Vineyard 1 Ground agent
Vineyard dataset 1
Weather data
Sensor data
Drone data
Ground-generated data
Workforce data
Vineyard 2 Ground agent
Vineyard dataset 2
Vineyard 2
Estimations models
ML Estimations service
Smart viticulture service provider
Teadal stretched data lake
Vineyard dataset 1
Access control policy
Business policy
Estimations for Vineyard 1
Estimations for Vineyard 2
Vineyard dataset 2
Vineyard 1 Owner
Upstream service provider (bank, insurance)
Cooperative entity
Vineyard 2 Owner
TEADAL
Legend
dataset
dataset with sensitive or personal information
who manages/generates the information
who accesses the information

Pilot #4: Industry 4.0

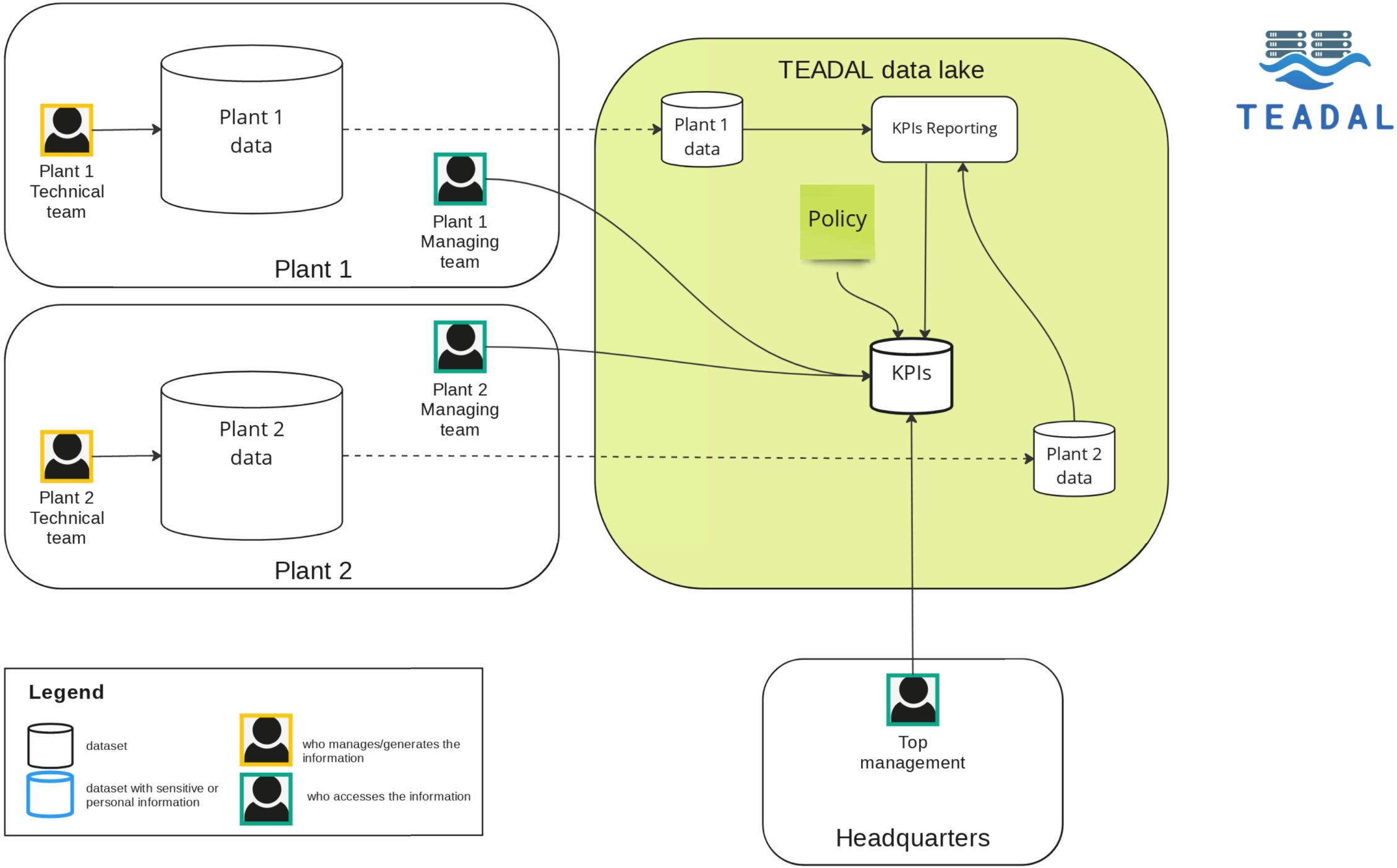
Objective: Allow processing KPIs over data from different manufacturing plants, each with its own data model, and enforce authorisation policies

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)


Plant 1
data
Plant 1
Technical
team
Plant 1
Managing
team
Plant 1
Plant 2
data
Plant 2
Technical
team
Plant 2
Managing
team
Plant 2
TEADAL data lake
Plant 1
data
KPIs Reporting
Policy
KPIs
Plant 2
data
Top
management
Headquarters
Legend
dataset
dataset with sensitive or
personal information
who manages/generates the
information
who accesses the information

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

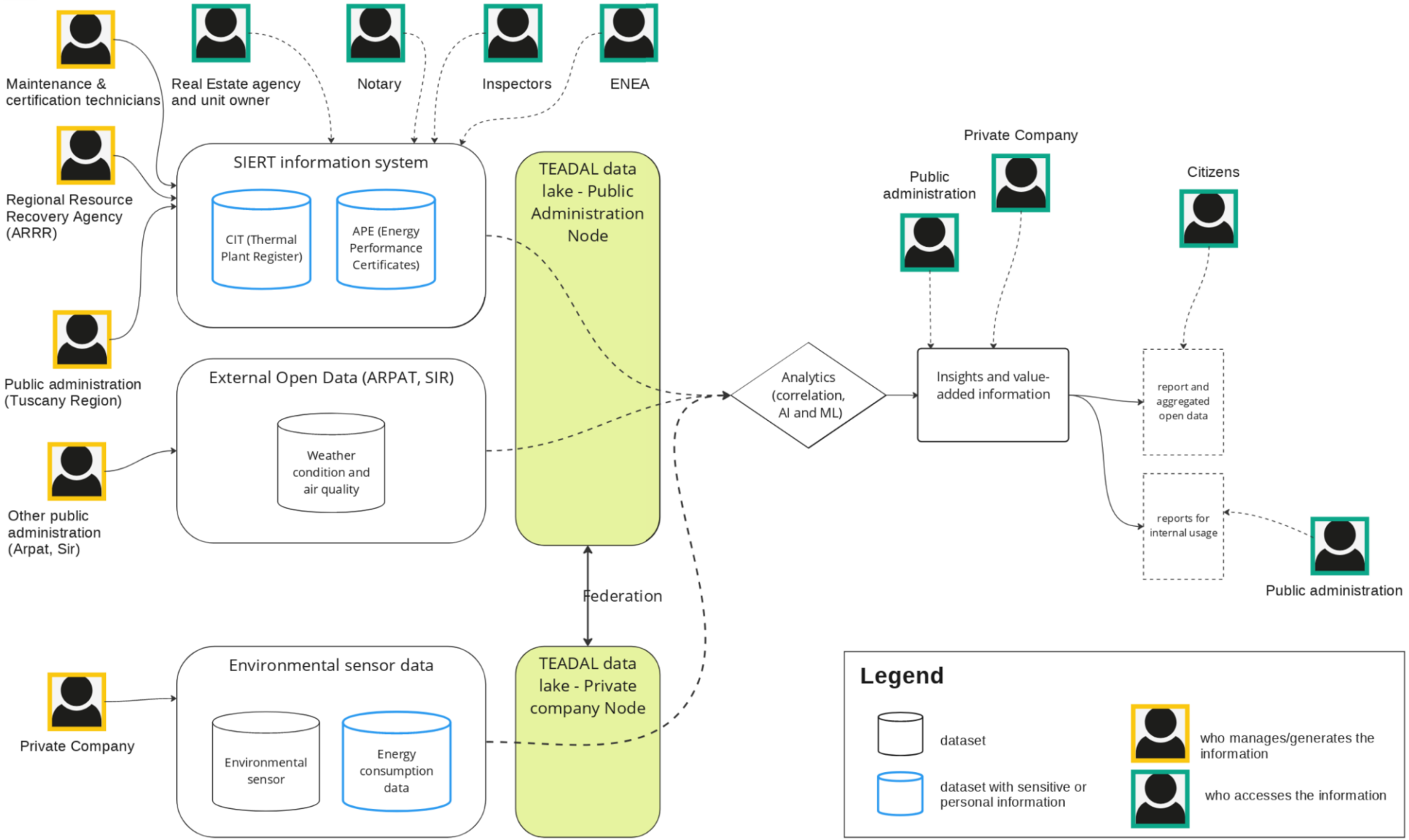
Why is it hard?

- 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

TEADAL

ubiwhere

MARTEL innovate

POLITECNICO MILANO 1863

TECHNISCHE UNIVERSITÄT BERLIN

i2cat

ERT

Cefriel

CYBERNETICA

TU WIEN dsg DISTRIBUTED SYSTEMS GROUP

AlmavivA

terraviewos

BOX2M industry 4.0

Regione Toscana

AMTS Azienda Metropolitana Trasporti e Sosta Catania S.p.A.

UITP ADVANCING PUBLIC TRANSPORT

ribera | marinaSalud Departamento Salud Denia

TEADAL

THANKS

TEADAL.EU

@TEADAL_eu

@TEADAL

TEADAL project is funded by the EU's Horizon Europe programme under Grant Agreement number 101070186