TEADAL

Introducing TEADAL

Pierluigi PLEBANI

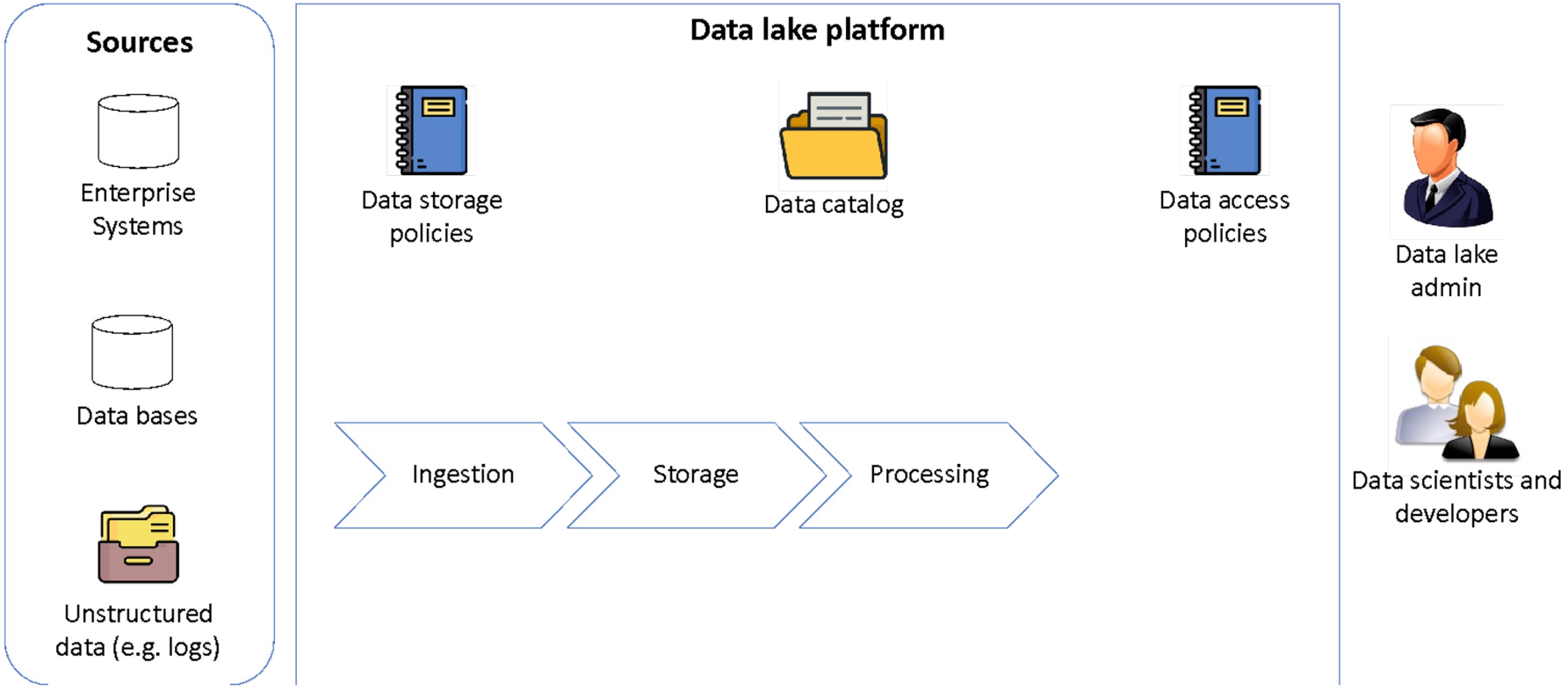
Politecnico di Milano

TEADAL 1st Workshop, Milan, Italy

14 March 2024

WWW.TEADAL.EU

Context



External data sources

External data access

EU Data strategy

EUROPEAN
COMMISSION

Brussels, 19.2.2020
COM(2020) 66 final

COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS

A European strategy for data

Availability of data: The value of data lies in its use and re-use. Currently there is not enough data available for innovative re-use, including for the development of artificial intelligence. The issues can be grouped according to who is the data holder and who is the data user, but also depend on the nature of data involved (i.e. personal data, non-personal data, or mixed data-sets combining the two[17]). Several of the issues concern the availability of data for the public good.

Data infrastructures and technologies: The digital transformation of the EU economy depends on the availability and uptake of secure, energy-efficient, affordable and high-quality data processing capacities, such as those offered by cloud infrastructures and services, both in data centres and at the edge. In this perspective, the EU needs to reduce its technological dependencies in these strategic infrastructures, at the centre of the data economy.

*New decentralised digital technologies such as **blockchain** offer a further possibility for both individuals and companies to manage data flows and usage, based on individual free choice and self-determination. Such technologies will make dynamic data portability in real time possible for individuals and companies, along with various compensation models.*

EU Data and Data governance Act

EUROPEAN
COMMISSION

Brussels, 23.2.2022
COM(2022) 68 final

2022/0047 (COD)

Proposal for a

REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

on harmonised rules on fair access to and use of data
(Data Act)

(Text with EEA relevance)

{SEC(2022) 81 final} - {SWD(2022) 34 final} - {SWD(2022) 35 final}

EUROPEAN
COMMISSION

Brussels, 25.11.2020
COM(2020) 767 final

2020/0340 (COD)

Proposal for a

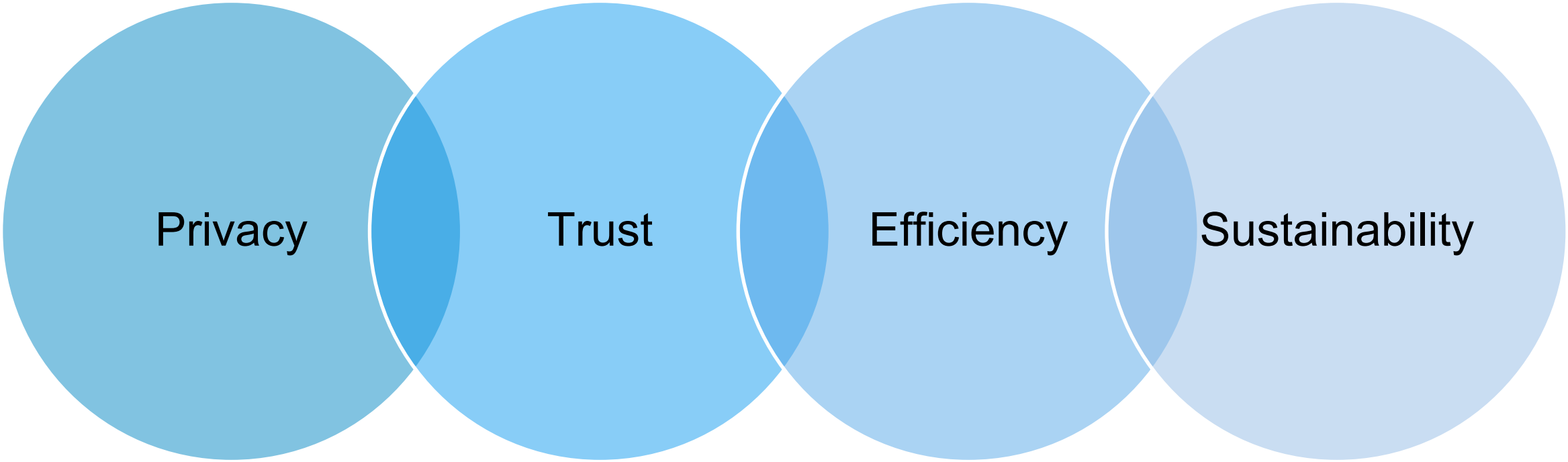
REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

on European data governance
(Data Governance Act)

(Text with EEA relevance)

{SEC(2020) 405 final} - {SWD(2020) 295 final} - {SWD(2020) 296 final}

Challenges

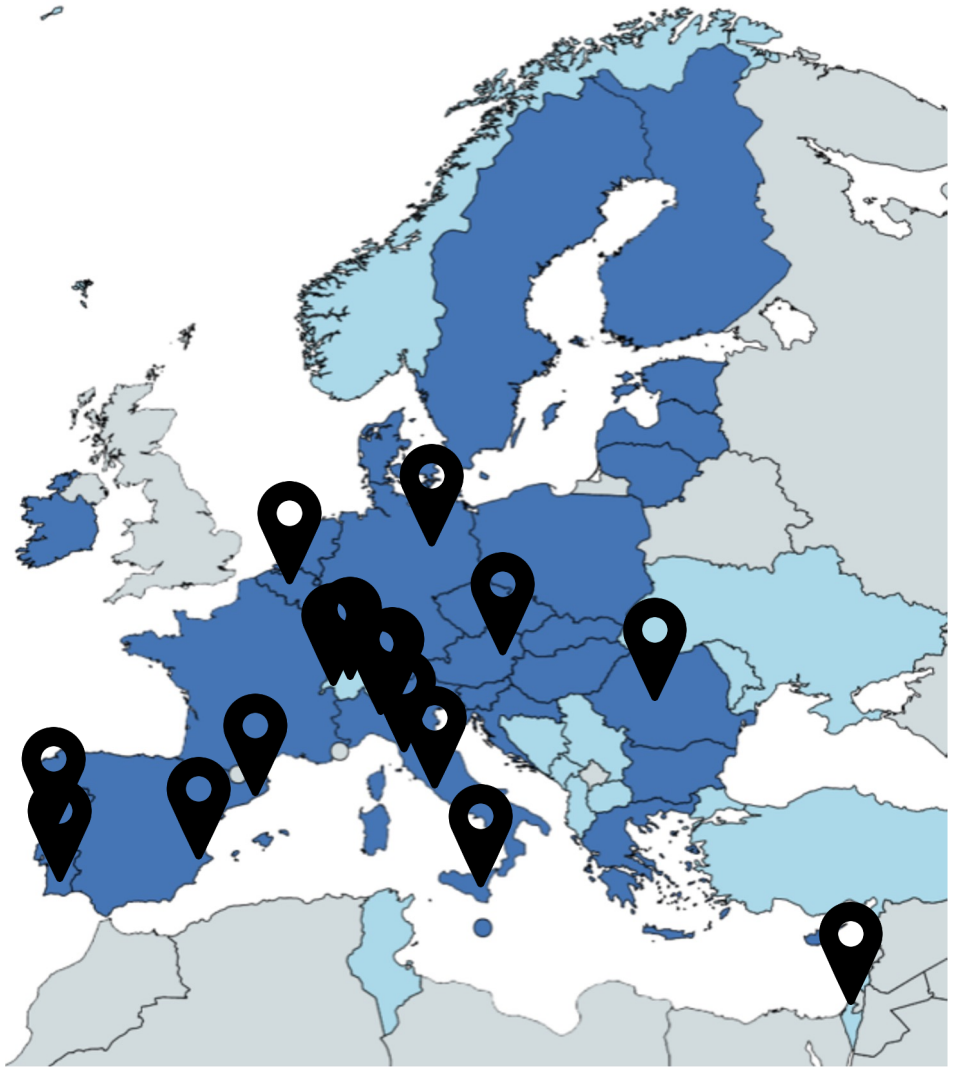


Main ambition

TEADAL

To provide key cornerstone technologies that will enable the creation of trustworthy mediatorless federations of data lakes spanning the cloud-edge continuum and, as dynamic constellations of different organizations, to improve a trusted, verifiable, and energy-efficient data sharing 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	MARINA SALUD, S.A.
8	UNION INTERNATIONALE DES TRANSPORTS PUBLICS
9	AZIENDA METROPOLITANA TRASPORTI E SOSTA CATANIA SPA
10	TECHNISCHE UNIVERSITAET WIEN
11	ALMAVIVA - THE ITALIAN INNOVATION COMPANY SPA
12	MARTEL GMBH
13	TERRAVIEW GMBH
14	ERT TÊXTIL PORTUGAL, S.A.
15	FUNDACIO PRIVADA I2CAT, INTERNET I INNOVACIO DIGITAL A CATALUNYA
16	BOX2M ENGINEERING SRL
17	REGIONE TOSCANA

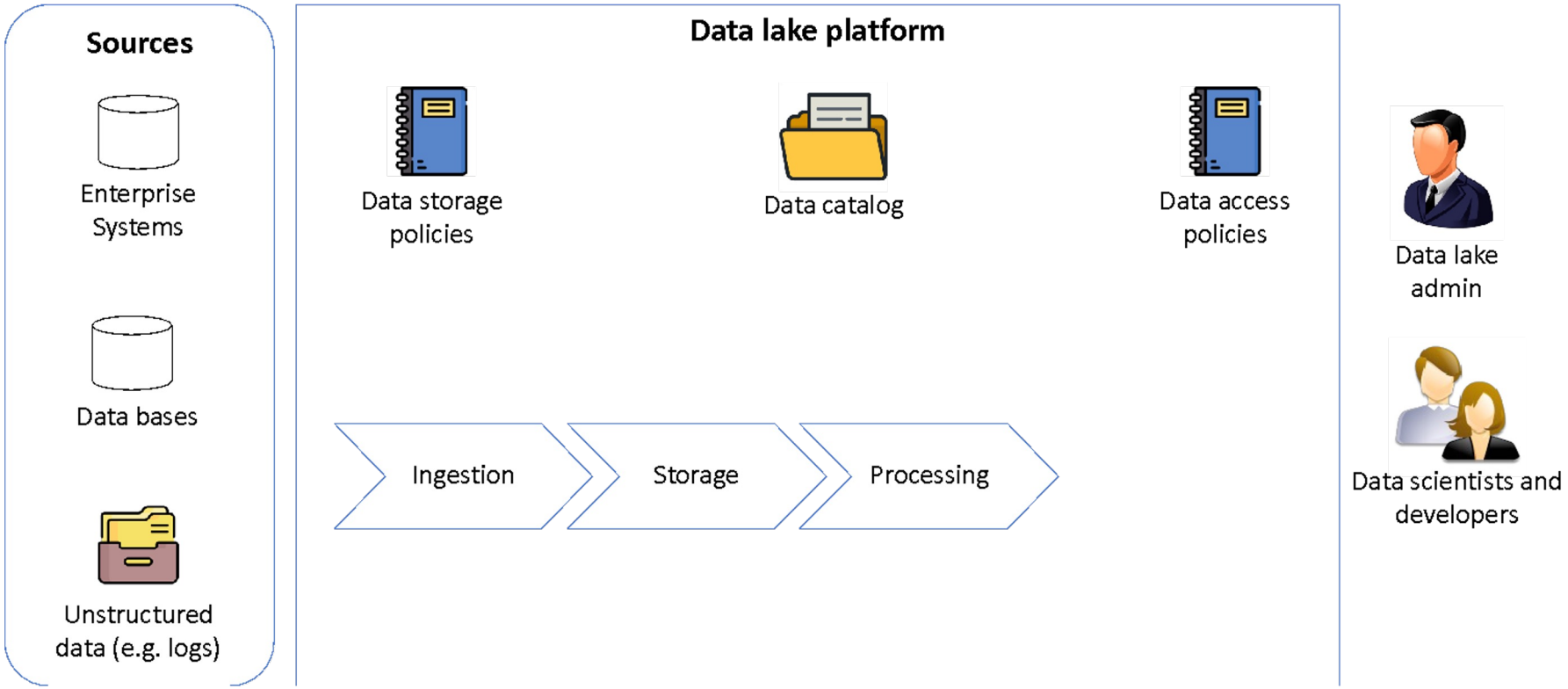
Objectives

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

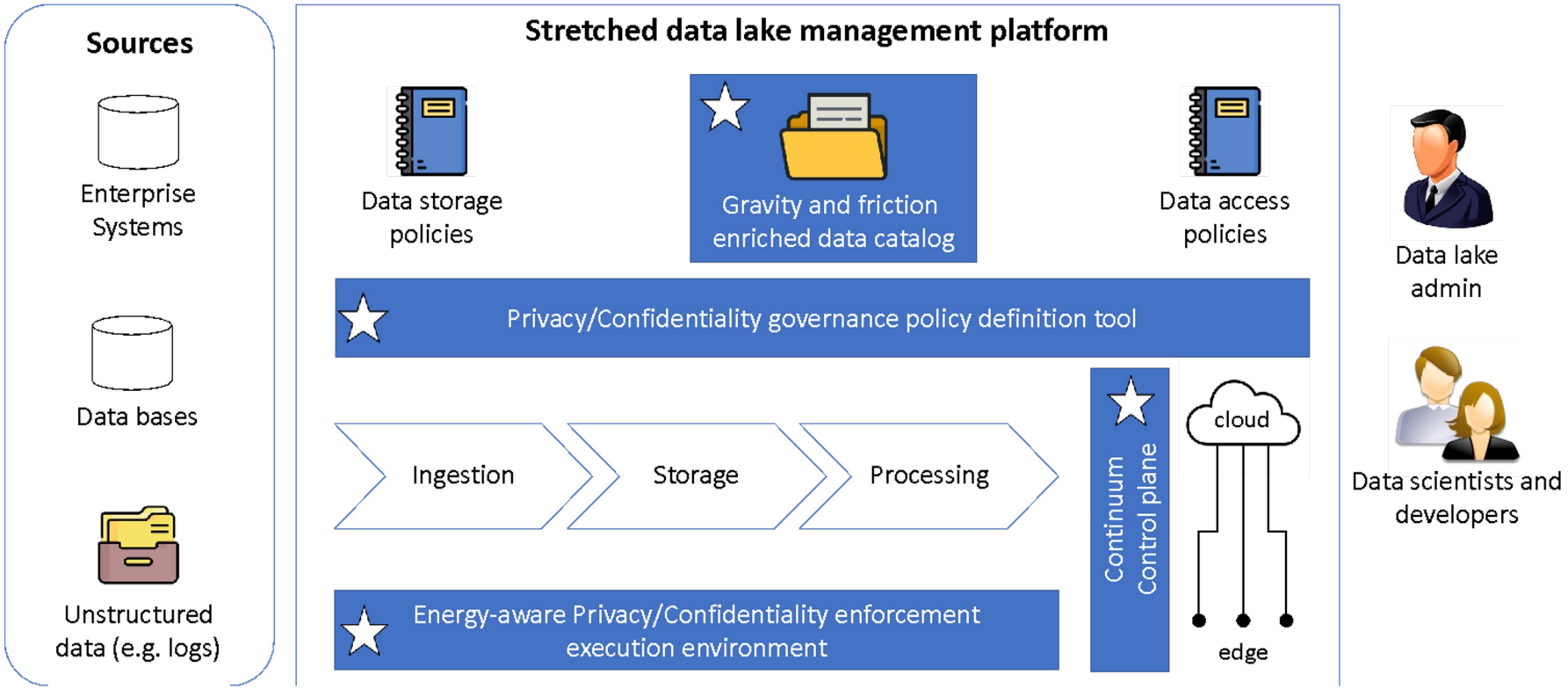
Expected results

1. A **control plane** to orchestrates and embeds performance, privacy, confidentiality capabilities into the workload data path
2. A trustworthy and mediatorless mechanisms to **federate data lakes** based on Blockchain/DLT
3. An **energy-aware data management** framework to identify possible actions to optimize data ingestion, storage, and processing
4. A **policy definition** tool in modelling privacy/confidentiality requirements
5. **Engagement of stakeholders** around concrete project outcomes

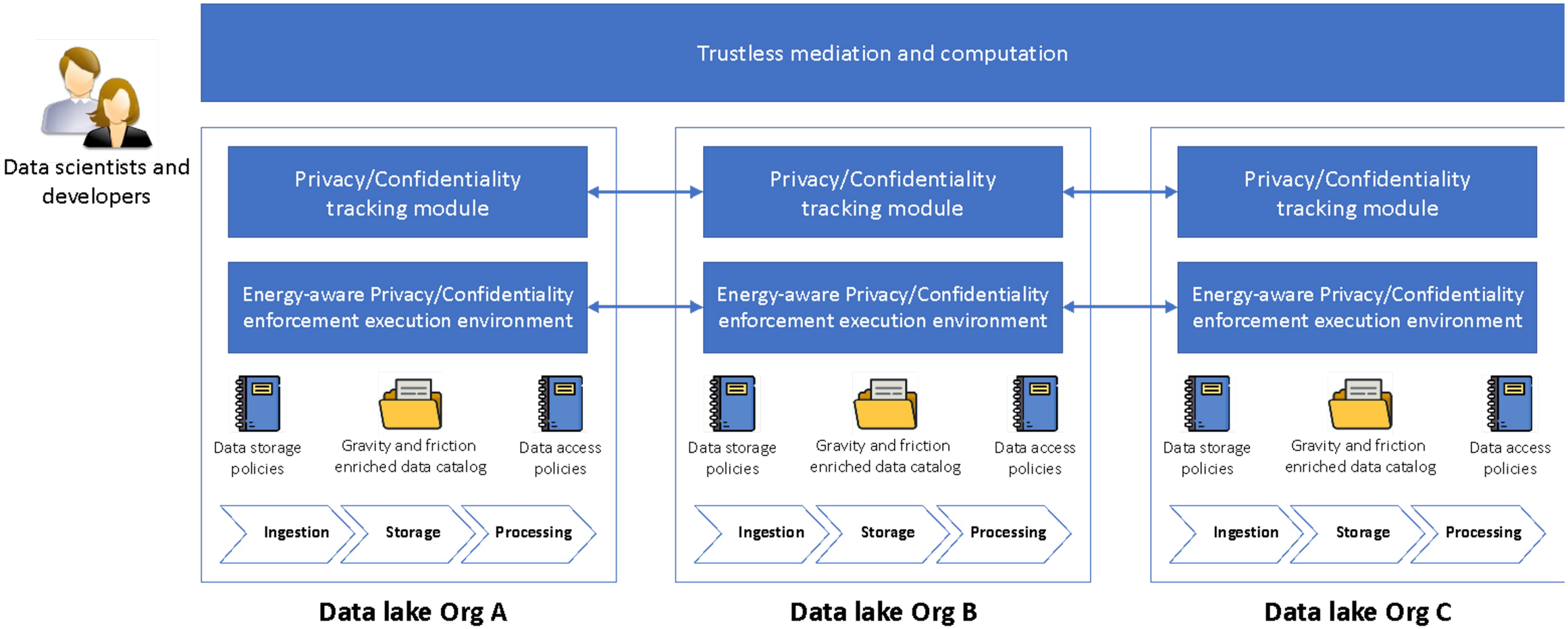
From data lake …



… to stretched data lake …



… That can be federated



S/T methodology - pillars

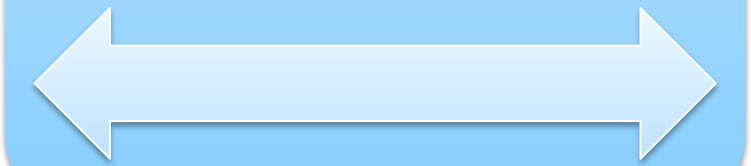
Data gravity/friction

Trustworthy data sharing

Energy-efficient use of data

S/T methodology - pillars

Data gravity/friction



Forces that will guide the data placement

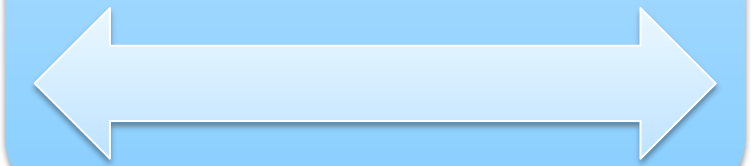
- Along the continuum
- Among the members of the federation

They will influence the development of innovative data catalogs where

- Privacy/confidentiality is related to the data location

S/T methodology - pillars

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

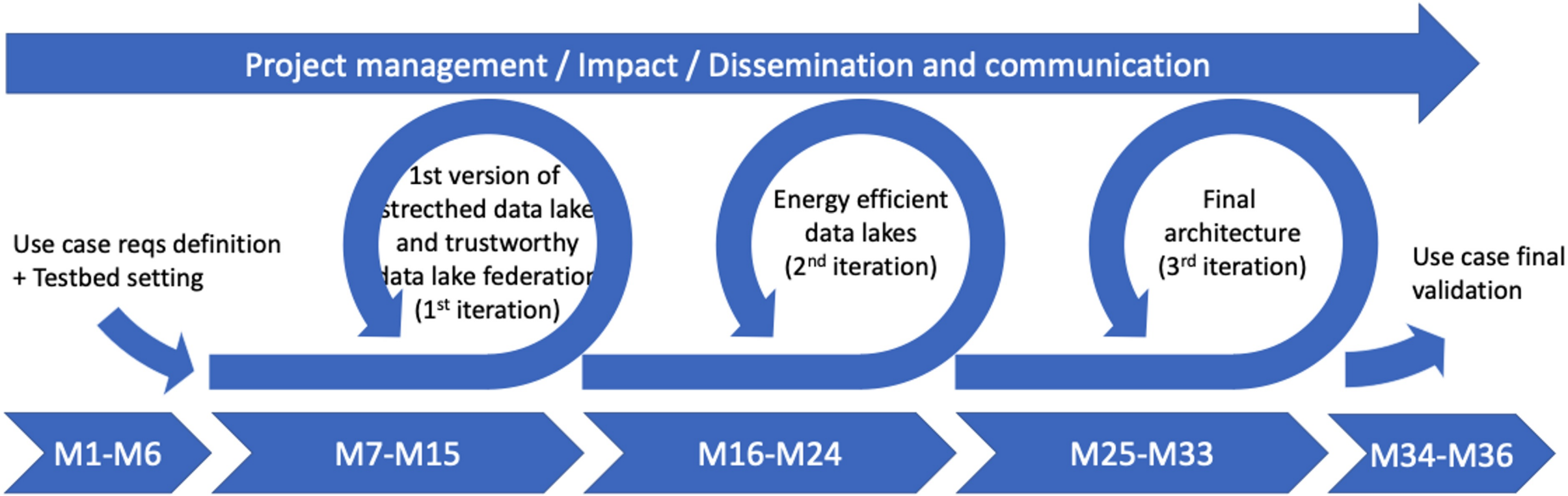
Sustainability must be a first-class requirement when developing our solutions

In all the aspects:

- Ingesting data
- Storing data
- Processing data

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

S/T methodology - phases



Milestones

Milestone number	Milestone name	Related WPs	Due date	Means of verification
MS1	Communication strategy definition	WP7	M3	D7.1 + Project visual identity defined, and project website launched
MS2	Pilot cases setting	WP2	M6	D2.1 + Relevant dataset available in the testbeds
MS3	1st version of stretched data lake and Trustworthy federation	WP4	M15	D2.2, D3.1, D4.1, D5.1, D6.1 + Federation of at least two data lakes up and running on two testbeds + Relevant code available on public repository
MS4	Energy-efficient data lakes	WP5	M24	D2.3, D3.2, D4.2, D5.2, D6.2 + Comparison of energy footprint with and without TEADAL facilities
MS5	Final architecture	WP4, WP5	M33	D2.4, D3.3, D4.3, D5.3 + All running testbeds + Final version of the code and instruction for deployment available on public repository
MS6	Validation	WP6	M36	D6.3

S/T methodology – pilot cases

Evidence-based medicine	• **Health data space – case partner: MARINA**
Mobility federated access point	• **Mobility data space – case partners: UITP, AMT**
Smart viticulture data sharing	• **Agricultural and Green Deal data spaces – case partner: TERRAVIEW**
Industry 4.0 fast KPI calculation	• **Industrial data space – case partner: ERT**
Regional planning for environmental sustainability	• **Energy/PA data spaces – case partners: RT, BOX2M**

S/T methodology - testbeds

BOX2M	• SaaS platform to collect and process data on micro-production energy plants
MARINA	• Private cloud providing serveless capabilities
POLIMI	• IaaS platform for data analystics
TERRAVIEW	• IaaS platform container-ready

Meet the TEADAL Consortium

TEADAL

ubiwhere

POLITECNICO MILANO 1863

CYBERNETICA

Cefriel

IBM

TECHNISCHE UNIVERSITÄT BERLIN

Regione Toscana

ribera | marinaSalud Departamento Salud Denia

UITP ADVANCING PUBLIC TRANSPORT

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

TU WIEN dsg DISTRIBUTED SYSTEMS GROUP

AlmavivA

MARTEL innovate

terraviewos

ERT

i2cat

BOX2M industry 4.0