GOSIM

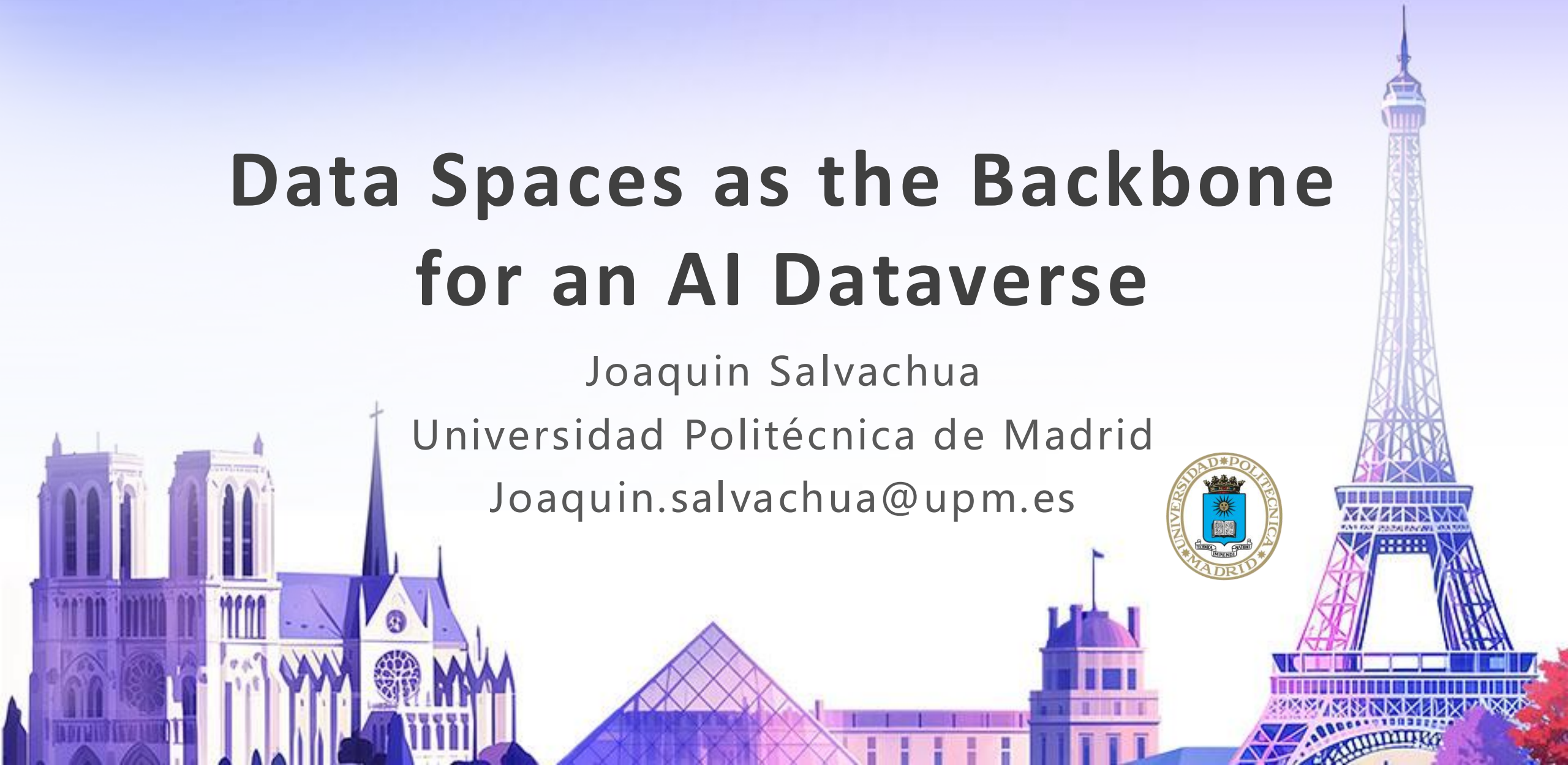
GOSIM AI Paris 2025

Data Spaces as the Backbone for an AI Dataverse

Joaquin Salvachua

Universidad Politécnica de Madrid

Joaquin.salvachua@upm.es



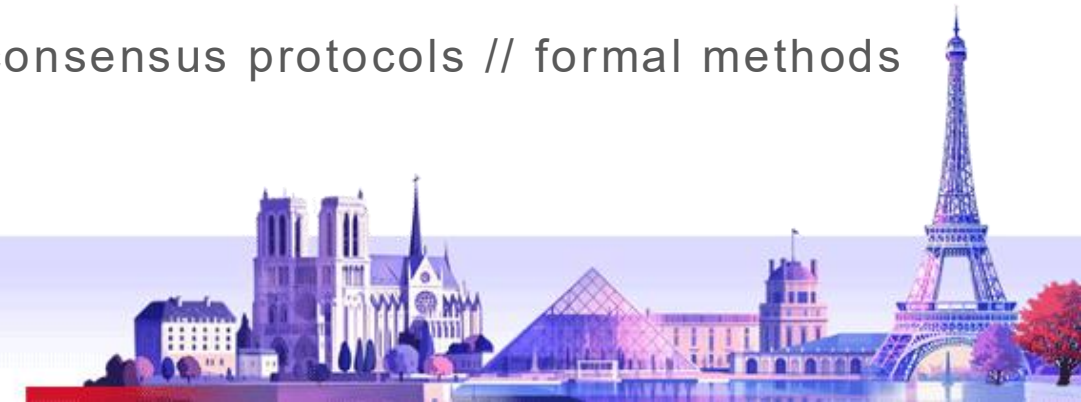
Who am I

GOSIM

- Associated Professor at Universidad Politécnica de Madrid
- IPTC (Information Processing and Telecommunications Center) // Escuela Técnica de Ingenieros de Telecomunicación.

- Researcher for Data Spaces //. Big Data-LLM infrastructure // Cloud infrastructure
- Distributed data governance and provenance. SSI infrastructure

- Working around the Gaia-X ecosystem (FIWARE, BDVA).
- Into UNE / Cen-Cenelec / W3C / IETF standarization
- Background on protocol standarization // distributed consensus protocols // formal methods (process algebra) // choreography



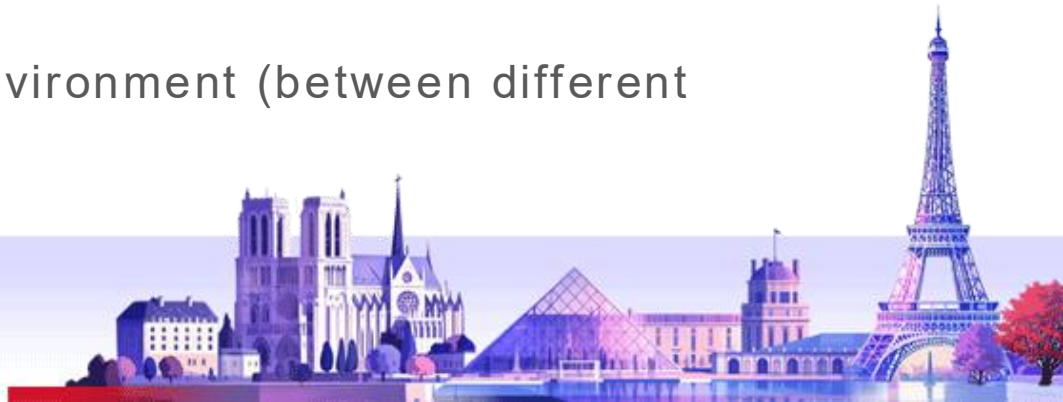
Problem statement

GOSIM

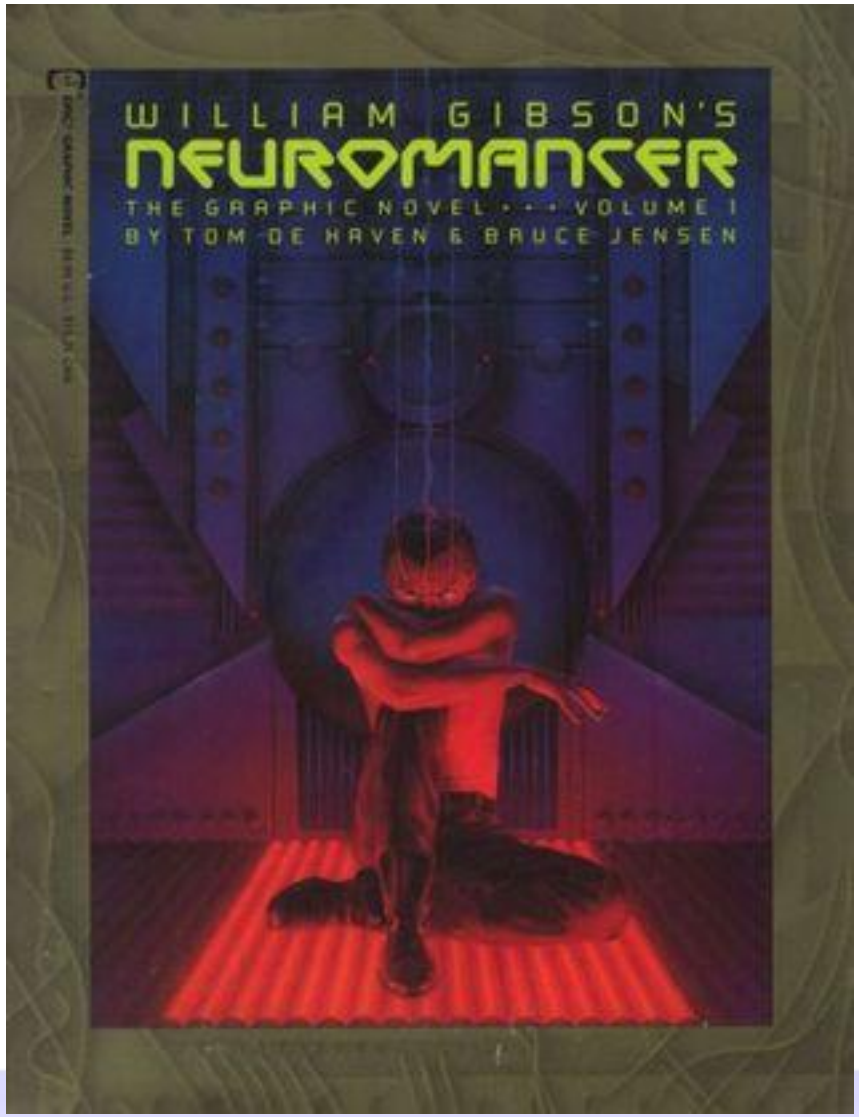
- Europe is developing the idea for data spaces : data sharing infrastructure with Trust, distributed governance, fine grained data access and usage control
- China is advancing on the Trusted Data Matrix, the same idea.
- We are developing tools for interoperability : data space definition language

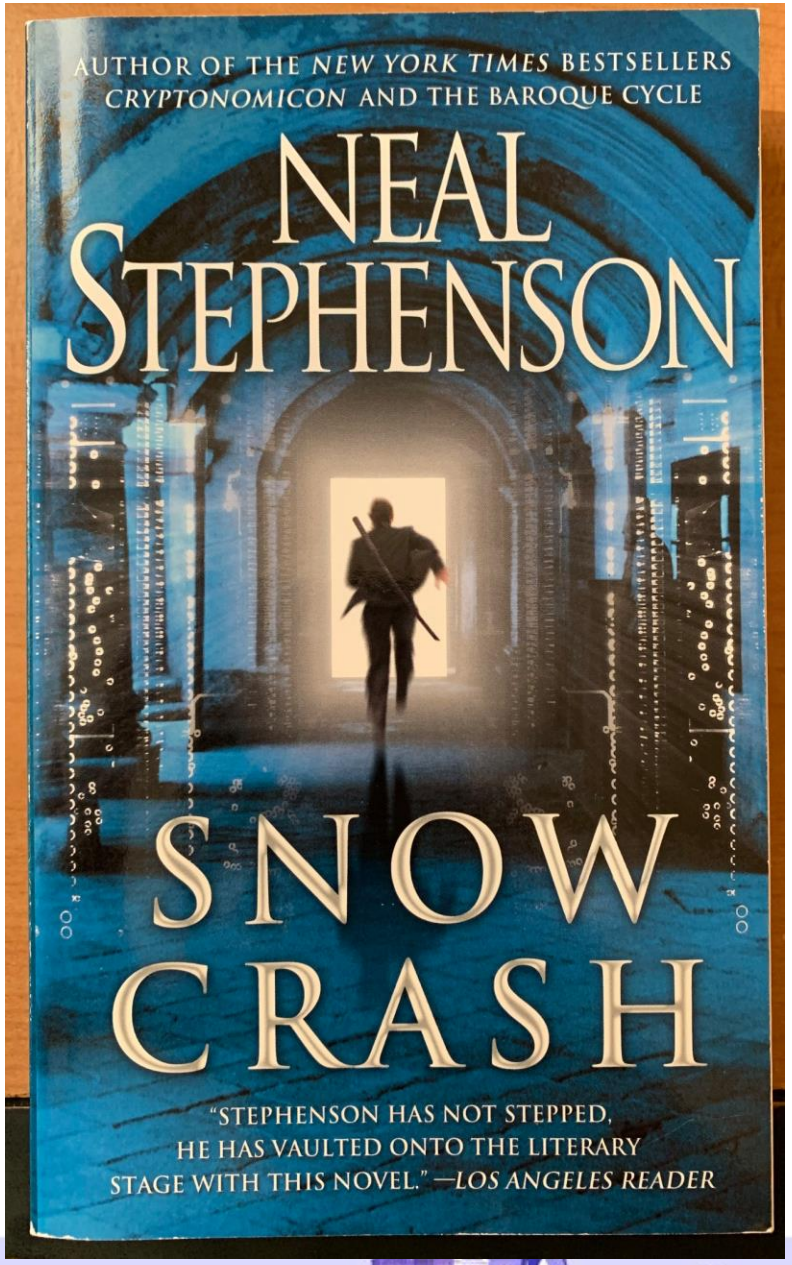
- We want to build upon this ideas different components : Dataverse
  - Digital twins and data based applications automatically build
  - Interactions with the LLMs and a true distributed

- Develop Trust models and infrastructures for data and agents to collaborate (with policy control on this interactions).
- Extend MCP and A2A protocol to allow similar Trust environment (between different partners)

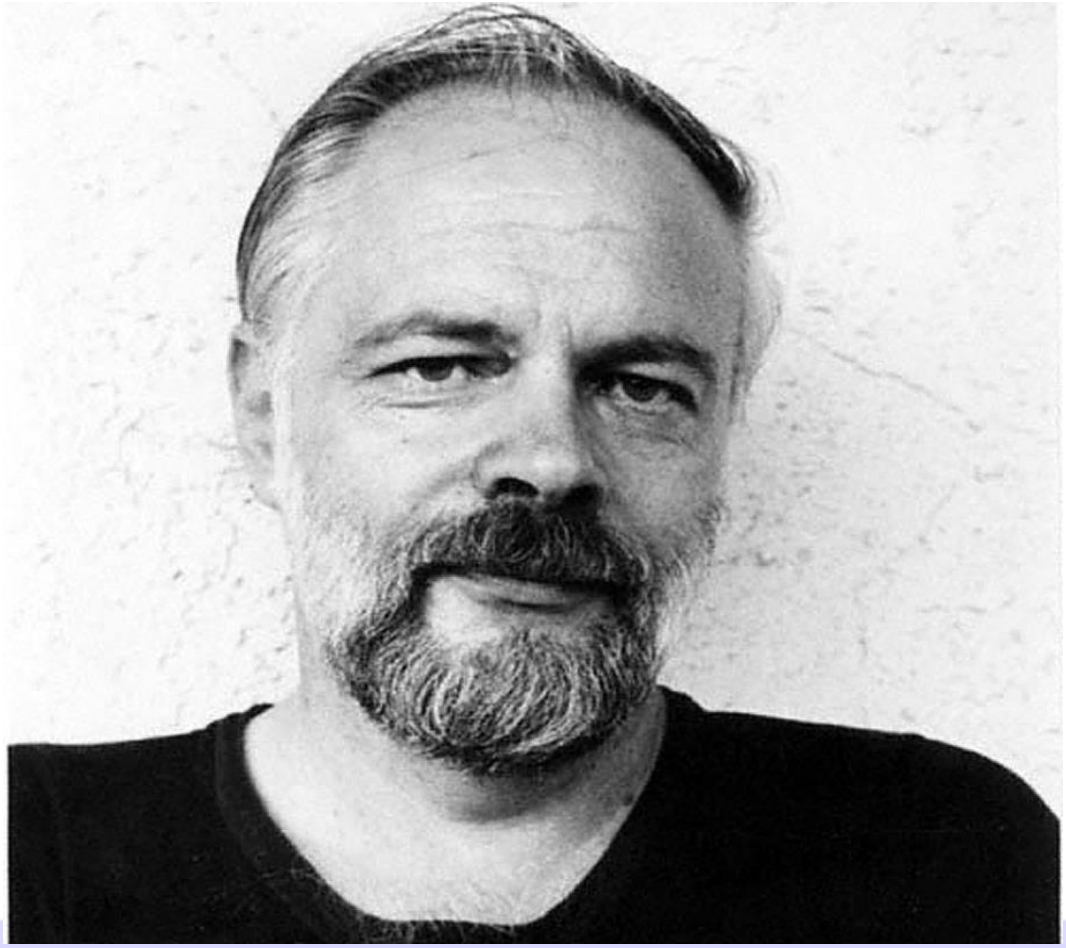


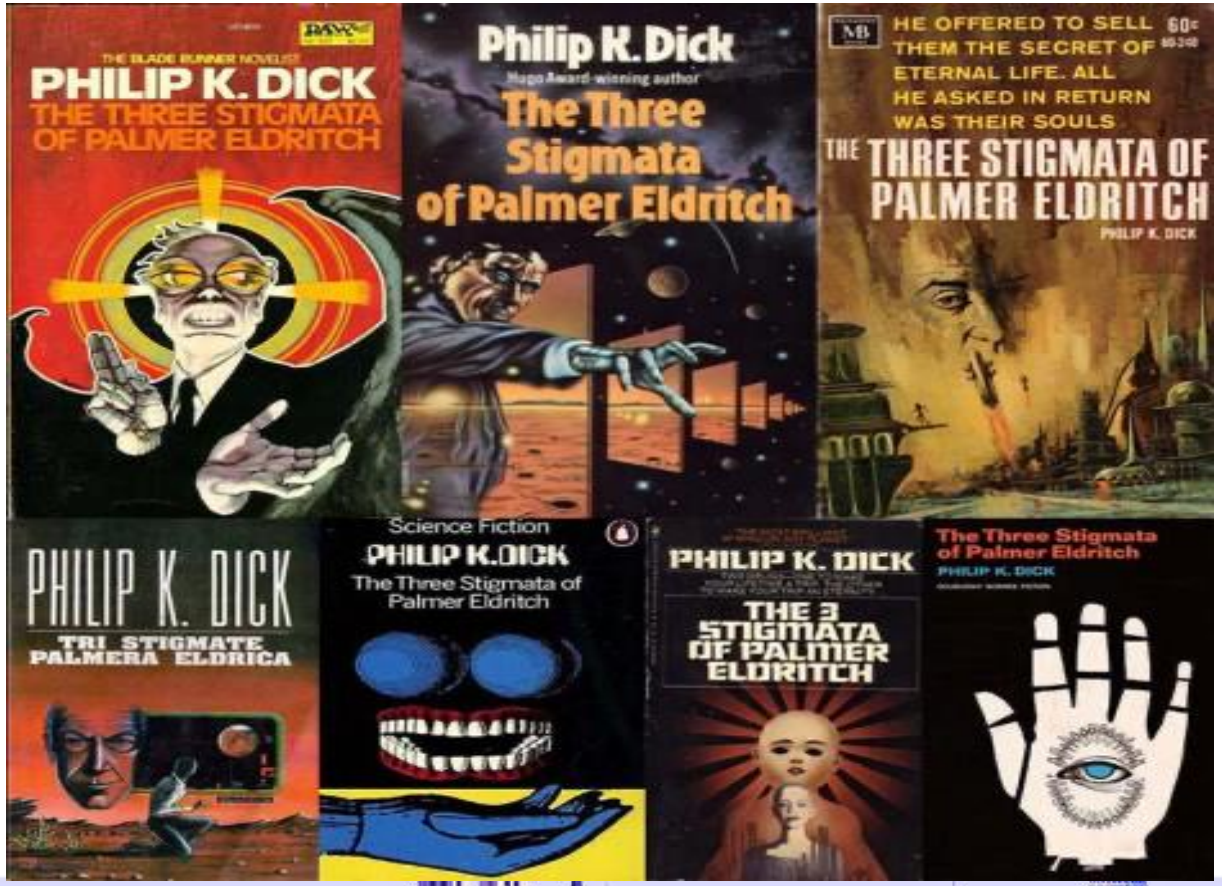
Origin for Terminology



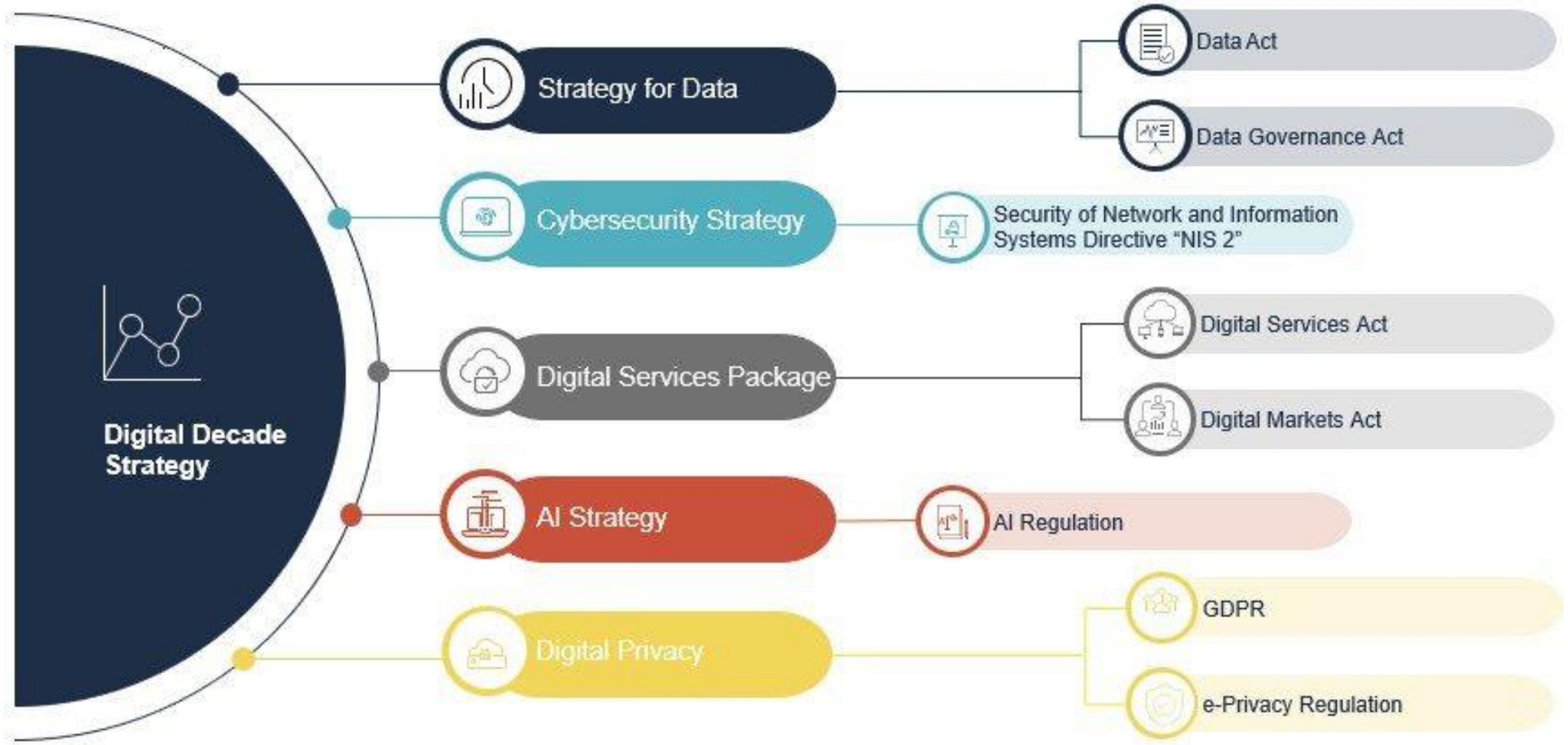


Philip K dick : Perky Pats multiverse / dataverse





Europe Acts


Digital Decade Strategy
Strategy for Data
Data Act
Data Governance Act
Cybersecurity Strategy
Security of Network and Information Systems Directive "NIS 2"
Digital Services Package
Digital Services Act
Digital Markets Act
AI Strategy
AI Regulation
Digital Privacy
GDPR
e-Privacy Regulation

The European Data strategy

GOSIM

European Alliance for Industrial Data, Edge and Cloud

IPCEI* on Next Generation Cloud
(*Important Project of Common European Interest)

EU Data Strategy

Coordination

Cloud actions:
Cloud Rulebook

Co-Investments in cloud-to-edge services, cloud federation and marketplaces.

Data actions:
New legislation (Data Act, Data Governance Act, etc.)

Co-investments in common European data Spaces

Data Spaces Support Centre
Coordination and governance

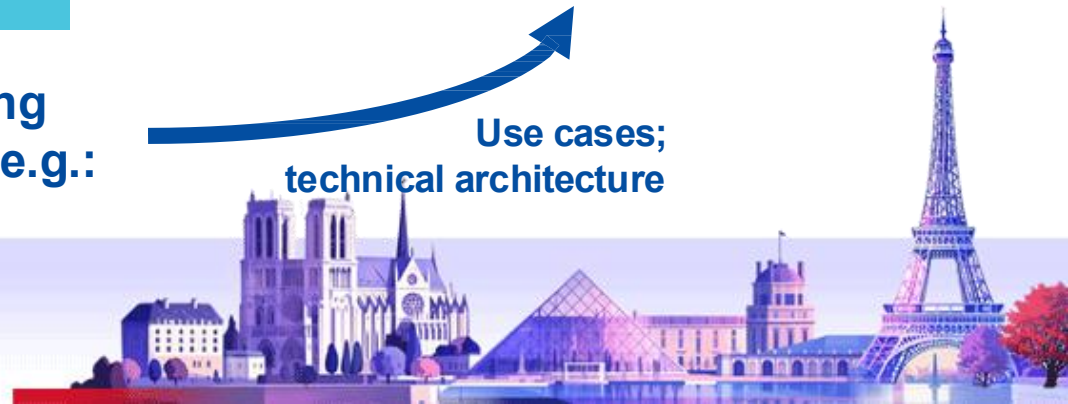
DIGITAL EUROPE PROGRAMME

Federation & interoperability standards

Complementing & integrating private and public initiatives, e.g.:

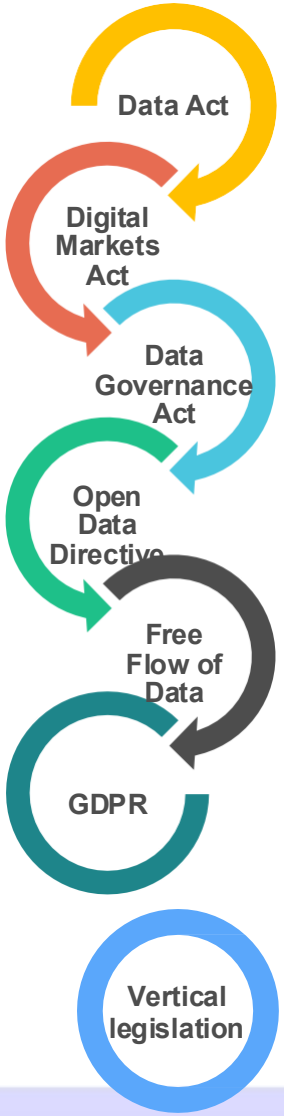
Use cases; technical architecture

gaia-x



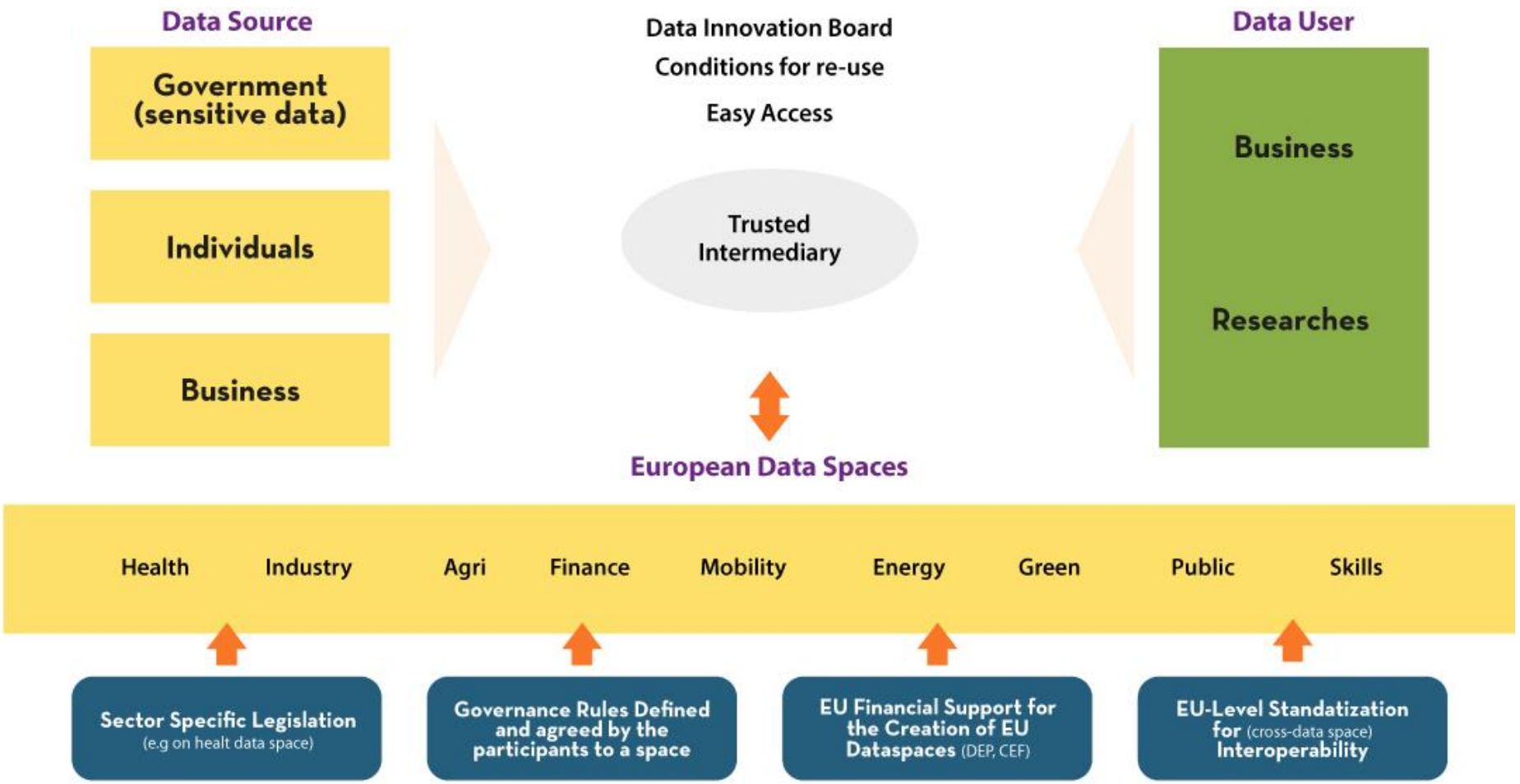
GOSIM AI Paris 2025

A comprehensive context legislative framework



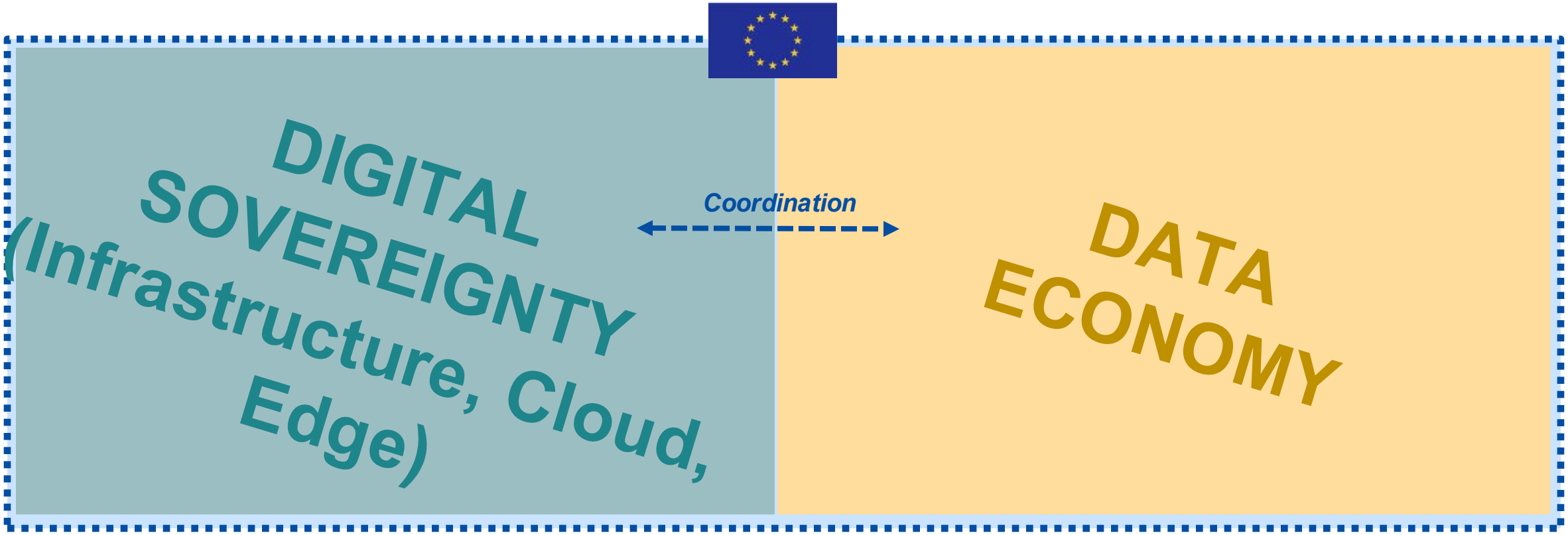
Aim	Data Covered	Regulated Actors
Ensure **FAIRNESS** in the allocation of data value among the actors of the data economy	Private sector data, personal and non-personal data, and co-generated (IoT) data	Businesses, public sector bodies, cloud and other data processing service providers
Tackle imbalances caused by the **MARKET POWER** of gatekeepers	Personal data and private sector data held by online platforms and originating from the users	Cloud and other data processing service providers, large data platforms
Ensure **TRUST** in data transactions	Public and private non-personal data, and personal data voluntarily made available by data holders	Data intermediation service providers, public sector bodies, (Recognised) Data Altruism Organisations
Promote use of **OPEN DATA**	Data in an open format that can be freely used, re-used and shared by anyone for any purpose	Public sector bodies, bodies governed by public law, public undertakings, universities
Ensure **FREE FLOW OF DATA** other than personal data within the Union	Non-personal data	Member States, competent authorities, professional users
Ensures a high-level of **DATA PROTECTION** and free flow of personal data in the Union	Personal data	Data controller, data processor, data subject, DPO, supervisory authorities, EDPB
Promote a competitive market according to **SECTOR-SPECIFIC** rules where necessary, e.g. automotive	Personal and non-personal data	Individuals and private and public sector bodies

Data Gobernance Act



Source: European commission, "Proposal for a Regulation on European Data Governance (Data Governance Act)"

The European Data Strategy



Based on: European Commission

- A secure and privacy-preserving IT infrastructure to pool, access, process, use and share data, Models and Agents.
- A data governance mechanism, comprising a set of rules of administrative and contractual nature that determine the rights to access, process, use and share data in a trustful, transparent manner and in compliance with existing legislations.
- Data holders are in control of who can have access to their data, for which purpose and under which conditions it can be used.
- Presence of vast amounts of data that are made available on a voluntary basis and can be reused against remuneration or for free, depending on the data holder's decision.
- Participation by an open number of organisations/individuals in full respect of competition rules and ensuring non-discriminatory access for all

Common European data spaces

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

Data Spaces Support Centre

- Coordinating the development of data spaces
- Assuring common standards and interoperability

Technical infrastructure for data spaces

Edge & cloud Services

Smart Middleware solutions

Marketplace

High-Performance Computing

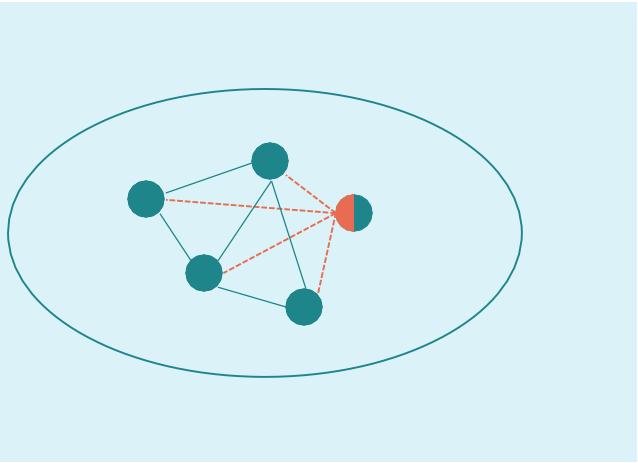
AI on demand platform

AI Testing and Experimentation Facilities

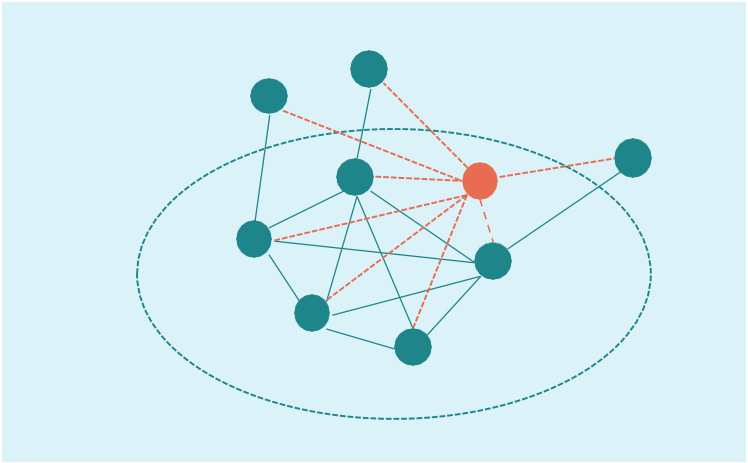
Evolution of data spaces : Trust Federations

GOSIM

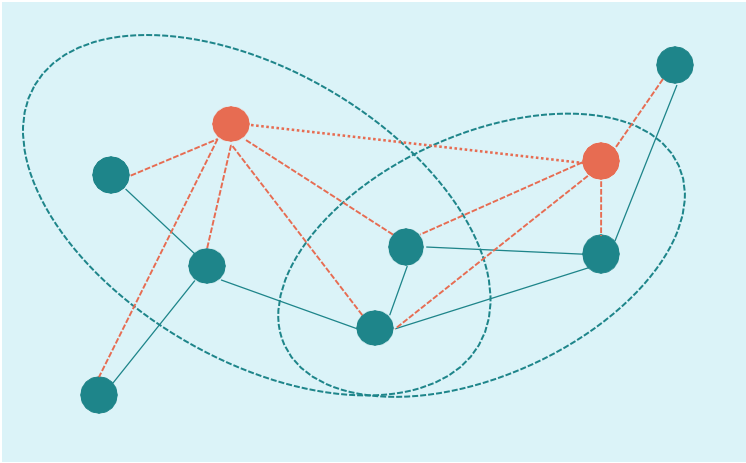
I Closed ecosystem



II Open ecosystem



III Federation of ecosystems

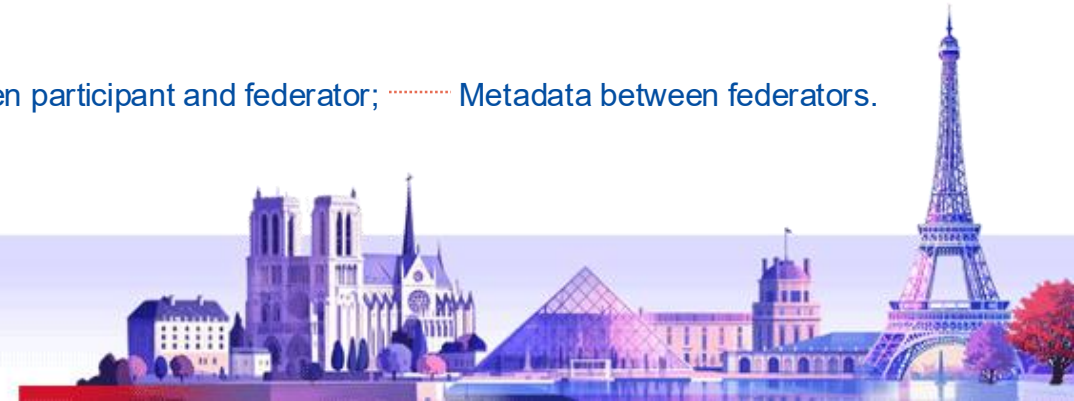


Legend:

Roles: Participant (Data Provider | Data User); Data intermediary.

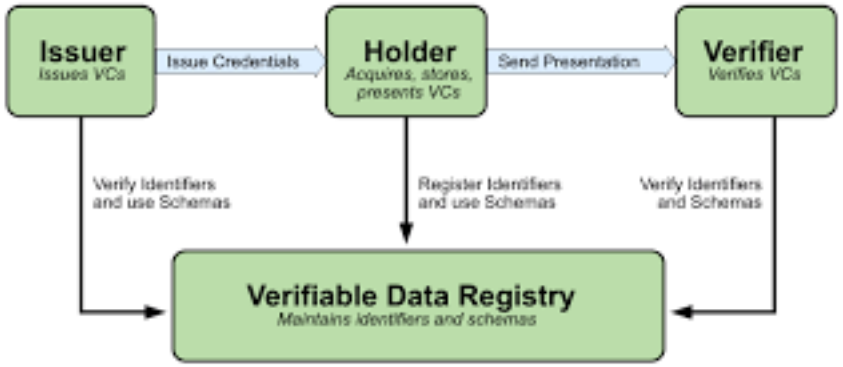
Data Exchanges: Payload data incl. metadata between participants; Metadata between participant and federator; Metadata between federators.

Ecosystems: Closed; Open.



Authentication

- Integration with Oauth / GNAP protocol.



- Based on Self Soveraign identity : Verificable credentials

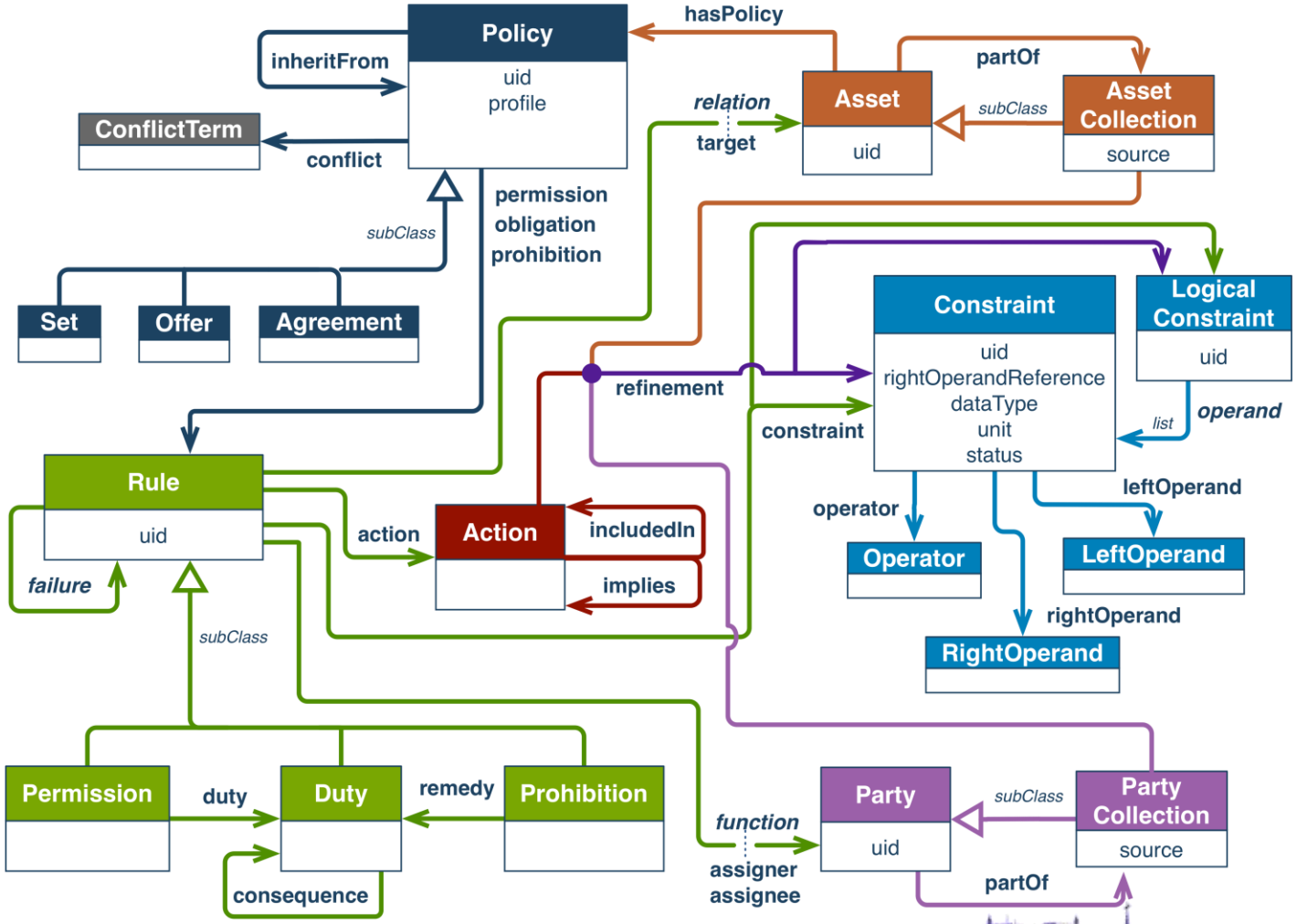
- Medium term : compatibility with EIDAS 2. Now using OpenWallet solution.
- Simplified workflow based on W3C VC API under development (now using OIDC4VP)

- Add extra control by the user about identity shared.

ODRL

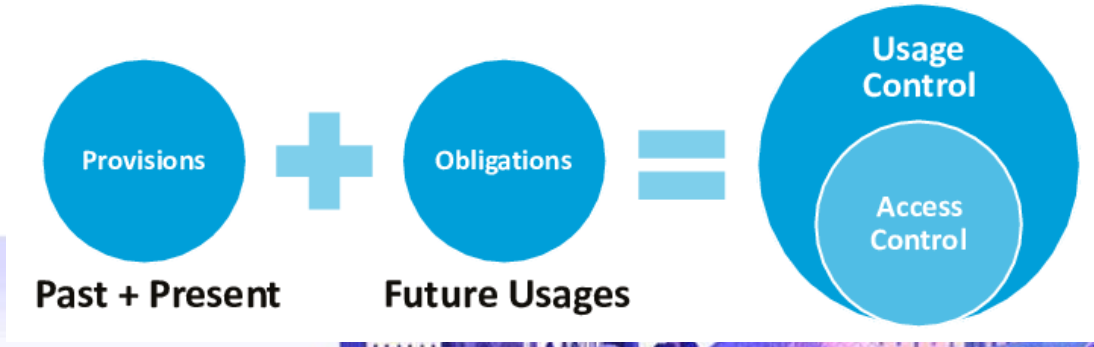
- Open Digital Rights Language (ODRL) is a W3C standard for expressing policies that define permissions, prohibitions, and obligations regarding the use of digital content and data.
- ODRL policies specify:
  - Who can access data (parties)
  - What actions are allowed or forbidden (permissions, prohibitions, obligations)
  - Which resources are targeted (assets)
  - Under what conditions (constraints, e.g., purpose, time)
- Widely used for:
  - Data access control
  - Consent management
  - Compliance with regulations (e.g., GDPR, Data governance act, etc).

ODRL 2.3 model (3.0. version ongoing)



Data **Access / Usage** Control

- Data Access Control:
  - Specify **who** can access **what** resource
  - Also the rights to access it (**actions**)
- Data Usage Control:
  - Ensures data **sovereignty**
  - Regulates what is **allowed to happen** with the data (future usage).
  - Related to **data ingestion and processing**
  - Context of intellectual property protection, privacy protection, compliance with regulations and digital rights management
  - Must ensurance governance and provenance



Access control

- ODRL Profile for Access Control (OAC) extends ODRL to manage access to personal data in decentralized environments like Solid Pods.
- Enables fine-grained, machine-readable policies such as:
  - Allowing read access to educational data *only for* academic research
  - Prohibiting write access to browsing data *unless* for commercial research
- Key benefits:
  - Users can express and enforce their data sharing preferences
  - Supports legal concepts and privacy vocabularies (e.g., GDPR, Data Privacy Vocabulary)
  - Policies are queryable and auditable for transparency and accountability
- Example ODRL Policy:
- “Permission to read educational qualification data if the purpose is a subclass of research and development”
- ODRL provides a flexible, standards-based way to define, manage, and enforce data access policies, supporting user control and regulatory compliance in modern data ecosystems

Data Usage Control

Policies definition

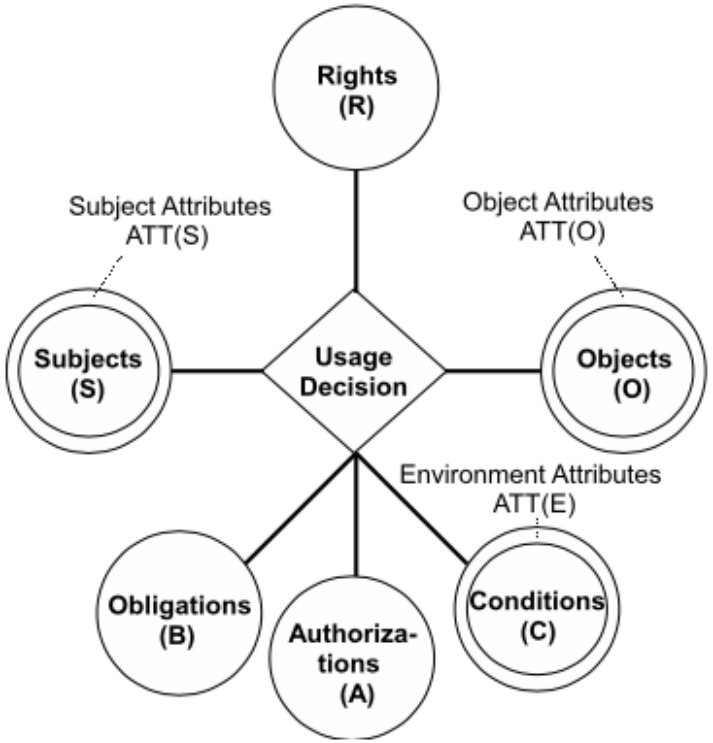
Based on the UCON specification and model.

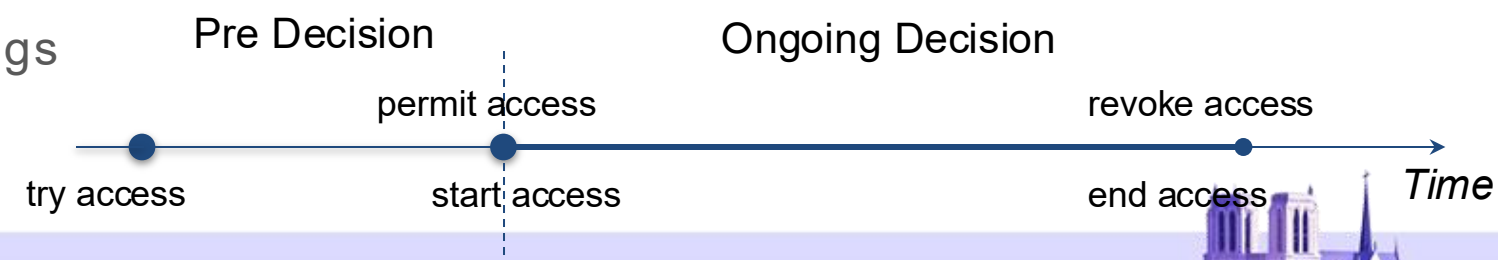
Define :

- **Obligations**
- **Authorizations**
- **Conditions**

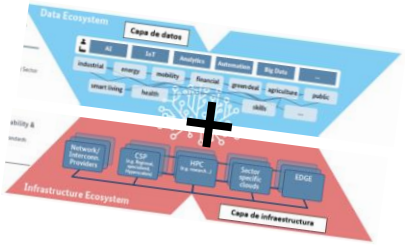
Over data and processing.

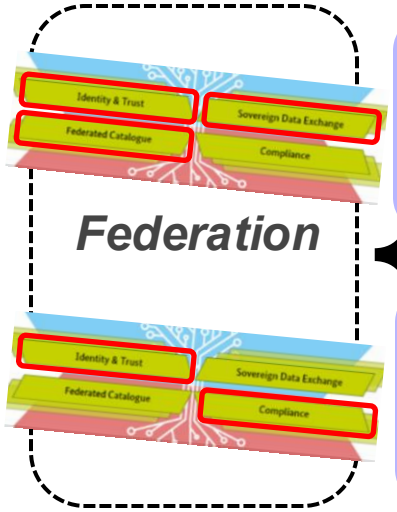
Using formal methods to generate policy agents for logs





architecture





Ecosistema 1

Ecosistema 2

Ecosistema 3

...

Ecosistem N

gaia-x

ts

Data and services

<<use>>

<<use>>

<<use>>

<<use>>

Federation

Gaia-X Trust Framework

Gaia-X Registry service

Gaia-X Compliance service

TRUST

Infraestructura de Federación y Confianza

Data spaces as enabler for data metaverse

GOSIM

- Data spaces should be the layer where to build data metaverses over it.
- Each metaverse client instance should be able to access data via dataspaces interfaces.
- Failures models are key in this definition to succed (NFS vs ZFS).

- Distributed data gobernance must be enforced over all the system :
  - We belive on formal methods for this (formal verification and validation in real time).

- Integration with 3D visualizations for agents and digital twins :
  - https://digital-strategy.ec.europa.eu/en/policies/event-web-4-governance

Actual MCP integration

- We have an Advanced Data Space Protocol. (ensurance the dataspaces workflow)
  - Policy negotiation
  - Ensure everyone agree (and have possible compensation for each step).
  - Allow control of all the interactions.
  - Allow Governance and data usage control in the future actions.
- Actual Implementation (rust based):
  - Develop a MCP connector over the data space connector.
  - Interconection with a LLM to perform operations with data over different dataspaces.

- Ongoing work :
  - Integrate trust in the MCP evolved protocol.
  - Use directly instead of the Data Spaces protocols : Hybrid one.
- Extend our Data spaces definition language (W3C ODRL group) in order to integrate LLMs
- This language is intended to specify a workflow for the different steps :
  - Interoperability with IEEE P3158. trusted data matrix
- Ongoing specification of the extended protocol for the IETF.

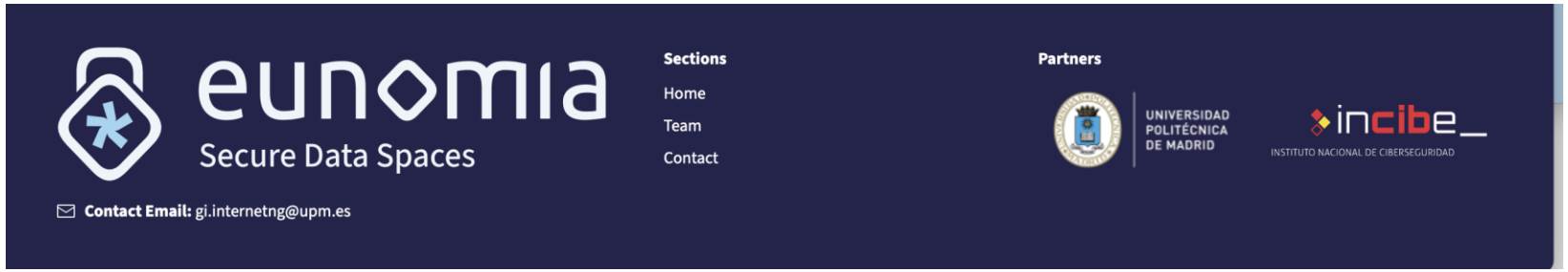
A2A integration

- Extend the Data space definition language to an Agent Federation definition language:
  - Extended model for Choreography based on process algebras extensions (similar to agents models).
  - Develop a way to deploy in a continuous cloud-edge infrastructure (Karmada based right now).
- Reuse the MCP data space integration for agents relationship between data, data spaces and LLMs and Agents.
- Extends the security model to a zero trust architecture with policy and distributed governance over actions performed : integration with a DLT (iota) for extra trust. Web4.0 evolution.

- Actual work : Extend the Agent card Json data forma to add extra authentication and authorization model (adding ODRL payload). Make it compatible with data spaces Trust.
- Extend skills in order to have a more formal model (also based on ODRL obligations model).
- Ongoing prototype using the Data space components.
- Future work Agent discovery protocols in data services marketplaces

Prototype ongoing

- Developed using the EUNOMIA project components for data spaces



Prototype after the Summer

Standardize the protocols on IETF (Advanced data space protocol) and Data spaced definition language (compatibility with MCP and A2A) and IEEE trusted data matrix.

Standarize ODRL profiles and VC profiles in w3C

Contribution in UNE and CEN-Cenelec working groups

Financiado por la Unión Europea
NextGenerationEU
GOBIERNO DE ESPAÑA
MINISTERIO PARA LA TRANSFORMACIÓN DIGITAL Y DE LA FUNCIÓN PÚBLICA
SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL
Plan de Recuperación, Transformación y Resiliencia
incibe_
INSTITUTO NACIONAL DE CIBERSEGURIDAD

GOSIM

GOSIM AI Paris 2025

THANK YOU