Data Considerations

Gyu Myoung Lee

Liverpool John Moores University

g.m.lee@ljmu.ac.uk

LIVERPOOL JOHN MOORES UNIVERSITY

Contents

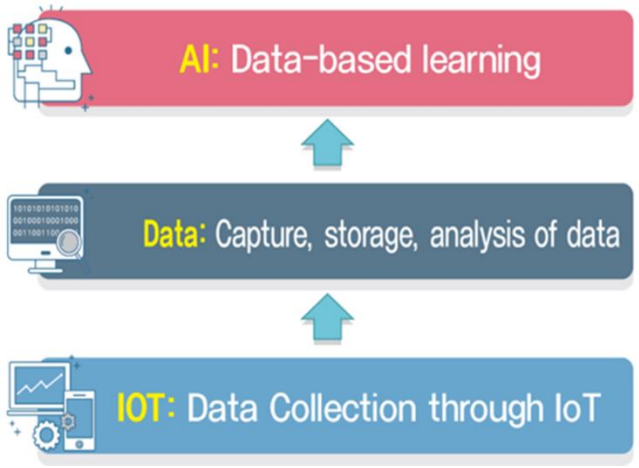
Introduction

Introduction

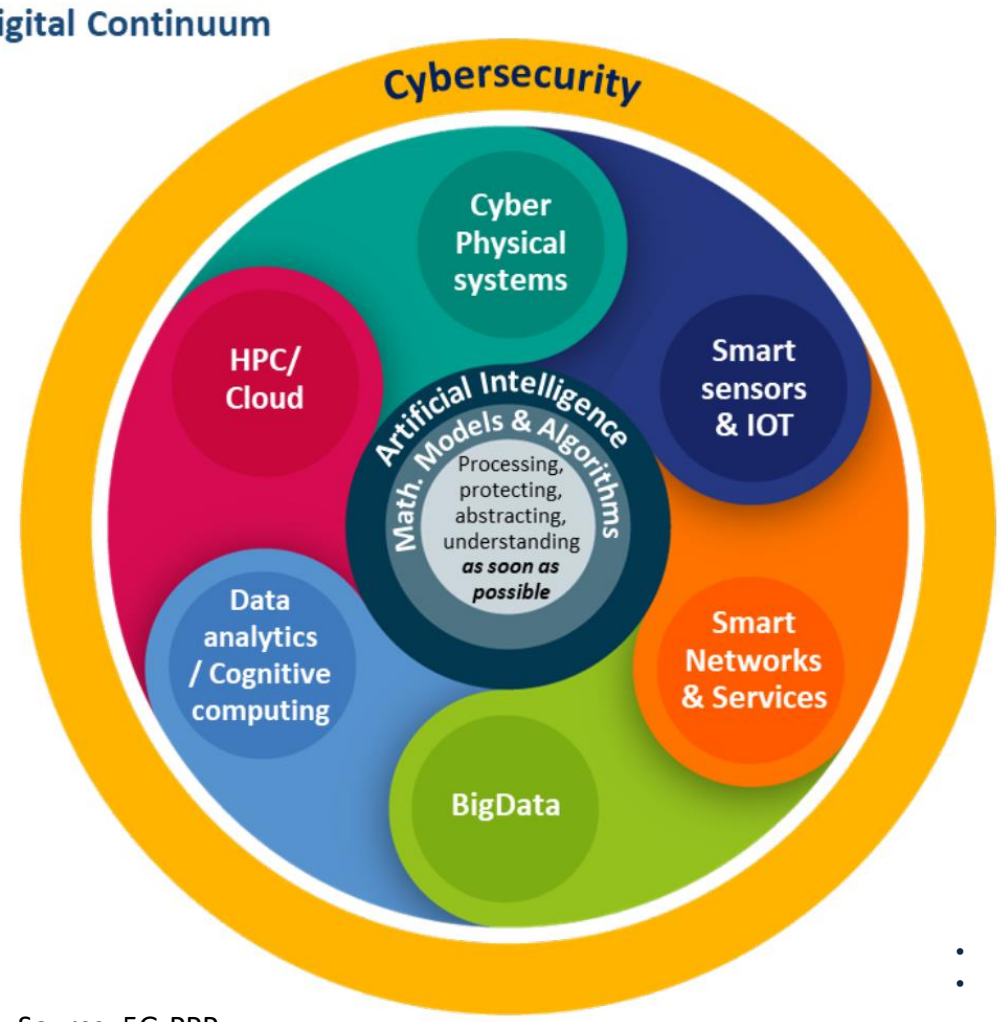
- **Data**
  - From data to actionable knowledge for creating value
- **Connected Intelligence**
  - From Cloud Native to AI Native
  - Decentralized intelligence
- **Fully automated Infrastructure**
  - AI for networks and Networks for AI

Trustworthy Ecosystem of Ecosystems

Towards the digital continuum from AI of Things



- **AI for Digital Infrastructure**
- **Digital Infrastructure for AI**
- **AI for Science, Industry and Societal Challenges**



Source: 5G-PPP

A continuous dynamic workflow

Between
Smart Sensors and IOT devices at the edge
and
HPC / cloud centers
over
Smart Networks and Services
executing
Simulation & Modelling, Big Data Analytics, ML*
based on
Math. Methods & Algorithms incl. MSODE**
pervasively augmented by
Artificial Intelligence
protected and secured by
Cybersecurity
back to
Cyber-Physical Systems

- ML: Machine Learning
- MSODE: Modelling, Simulation and Optimization in Data-rich Environment

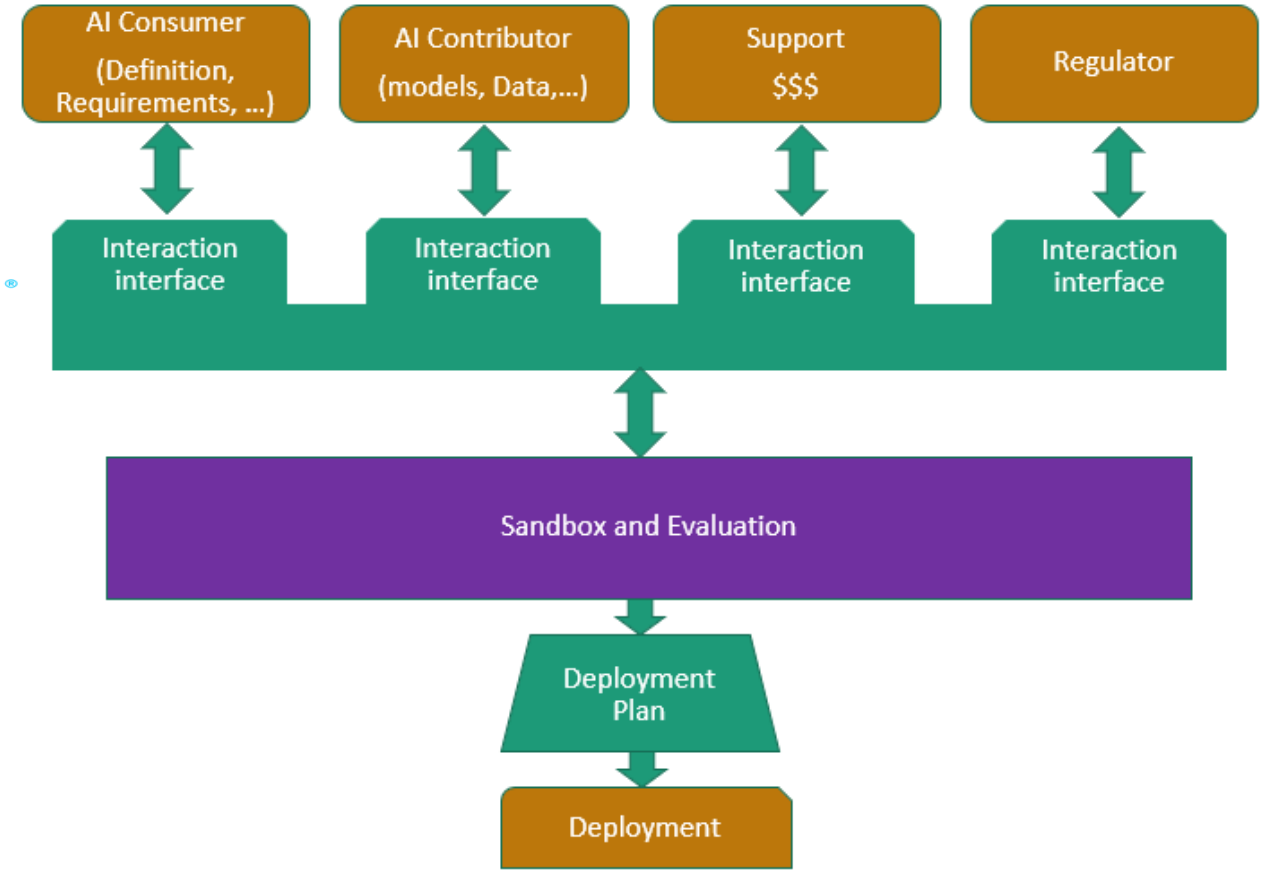
Project Resilience

AI and Data Commons

Project Resilience

AI and Data Commons
Global Initiative

XPRIZE
Cognizant
IEEE Advancing Technology for Humanity
Blavatnik School of Government
University of Oxford
intel
The University of Texas at Austin
City of Los Angeles Founded 1781
ITU
aws



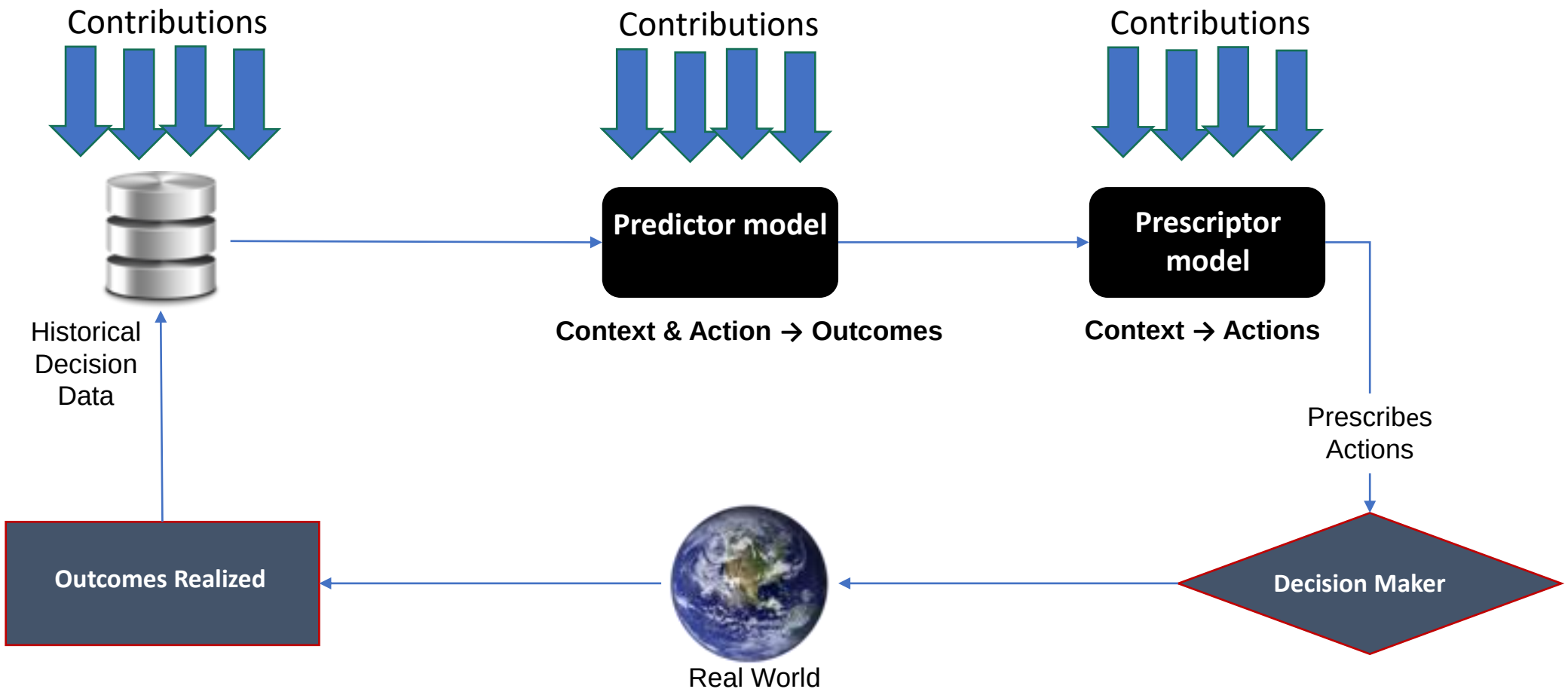
- **WG1 (Platform MVP)**: Babak Hodjat, Risto Miikkulaninen
- **WG2 (Data Contributions)**: Gyu Myoung Lee, Toby Philips
- **WG3 (Product Experience)**: Mohanty Sharada, Sean McGregor

AI for Good

Webinars - https://aiforgood.itu.int/event/project-resilience-2/

MVP AI Contribution Points

AI and Data Commons
Global Initiative



Data WG: Data sharing in a standardized way with interoperable interfaces

UN - data and digital transformation strategies

- License regime for derived data products and ML models using multiple sources of input data
- Provisions around shared IP for digital products and services developed in partnership with digital companies
- Design and publication of **Application Programming Interface (APIs) standards**
  - UNEP is building API standards for environmental data
- Digital public good standard
  - **The Digital Public Goods Alliance** has issued the digital public good standard for digital products and services being published as digital public goods.

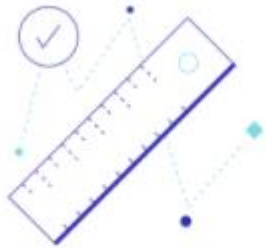
Digital Public Goods Standard

- a set of specifications and guidelines designed to maximise consensus about whether a digital solution conforms to the definition of a digital public good
  - open-source software, open data, open AI models, open standards, and open content that adhere to privacy and other applicable best practices

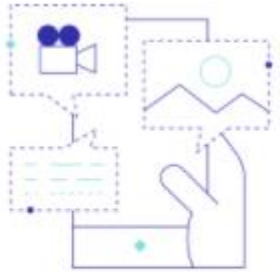
Open Software

Open Data

Open AI Models



Open Standards



Open Content

Key Principles for Data Spaces

12 Principles for Data Spaces

	Principles	Challenges
1	Data spaces are **ecosystems of systems**	Structuring and operating an ecosystem of ecosystems
2	**Data usage** require provisioning from connecting devices	Creation of value associated with usage control
3	Data spaces support **data lifecycle**	Characterizing and managing data lifecycle
4	Data interoperability enabled by a **common language**	Common language for semantic interoperability
5	Data usage enabled by **common data models**	Common data models for behavioral interoperability
6	**Data curation**	Organisation, description, cleaning enhancing and preserving for public use
7	**Trust** in data sharing	Trustworthiness and risk management
8	**Governance** for ethical usage of data	Governance and ethics
9	**Decentralisation**	Decentralisation
10	**Integrated data management**	Data fabric
11	**Extensible data spaces**	Scaling-up data spaces
12	**User-centricity**	Business roles and interactions

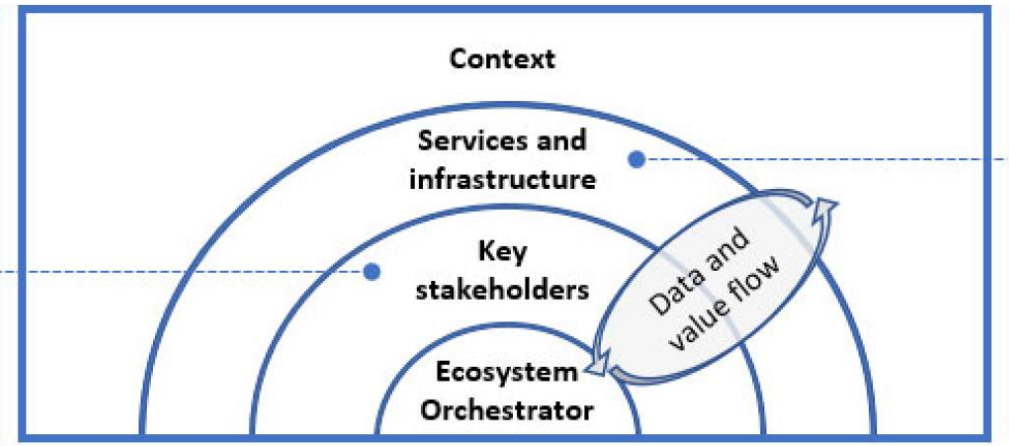
Gyu Myoung Lee, AIOTI WG-ULA Technical Report, HSN workshop 2022, IoT Week 2022, etc.

Key findings and challenges – 1

- **The Ecosystem of ecosystems**
  - Technology ecosystems (e.g., 5G, Clouds, IoT, Big Data & AI, etc.)
  - Vertical domain specific ecosystems (e.g., industrial, health, energy, etc.)

Data spaces as ecosystem

Key stakeholders and their roles
Who are the key stakeholders? What are their roles in the data ecosystem?
Who is the Data Ecosystem Orchestrator?
Who are the Data Providers?
Who are the Data Consumers?

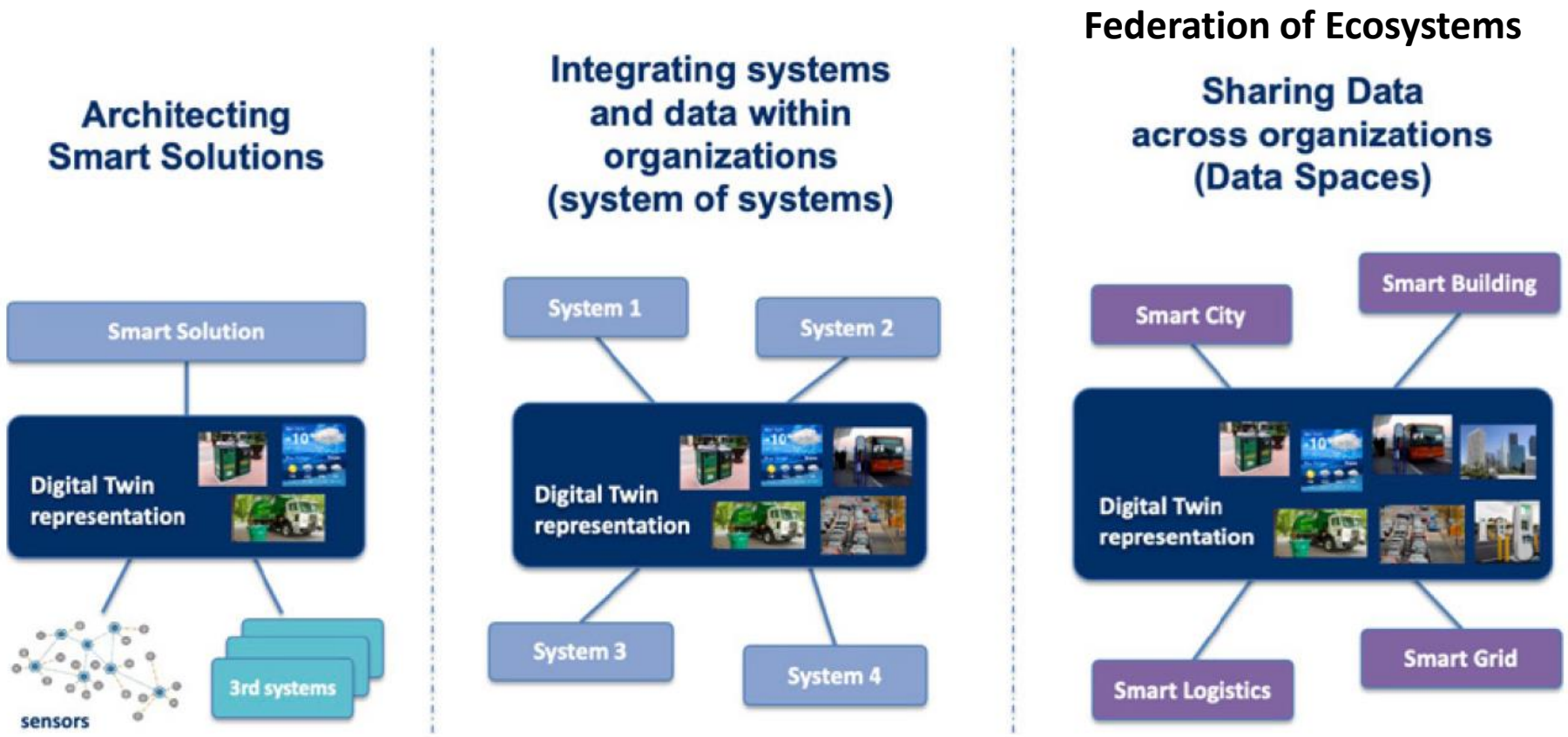


Services and infrastructure
What value adding services are needed in the ecosystem?
What infrastructure is required?
Who are the trusted intermediaries?

Key findings and challenges – 2

- **Scaling up – large scale virtual continuum (space-time)**

FIWARE
Digital Twin

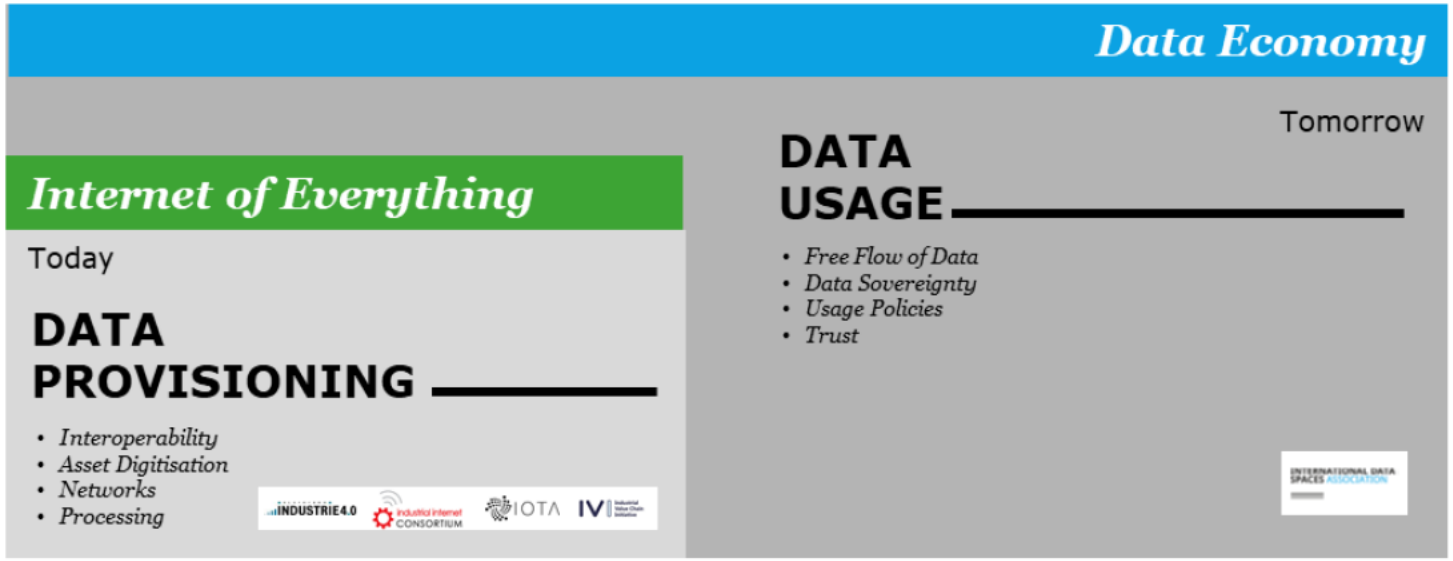


cdbb
Centre for Digital Built Britain

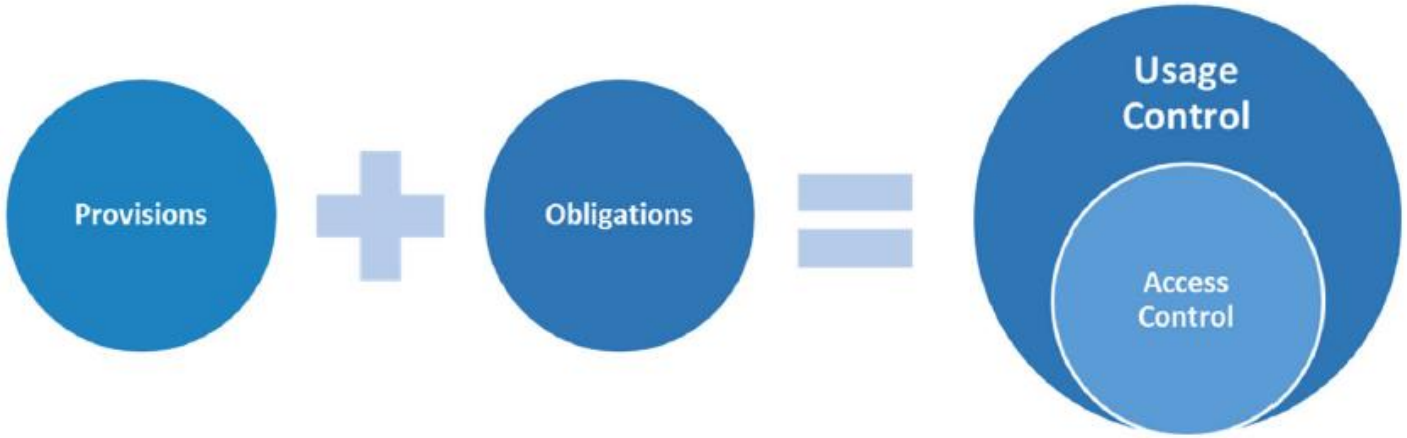
Key findings and challenges – 3

- **From data provisioning to data usage**

From Connecting Devices to Creating Value

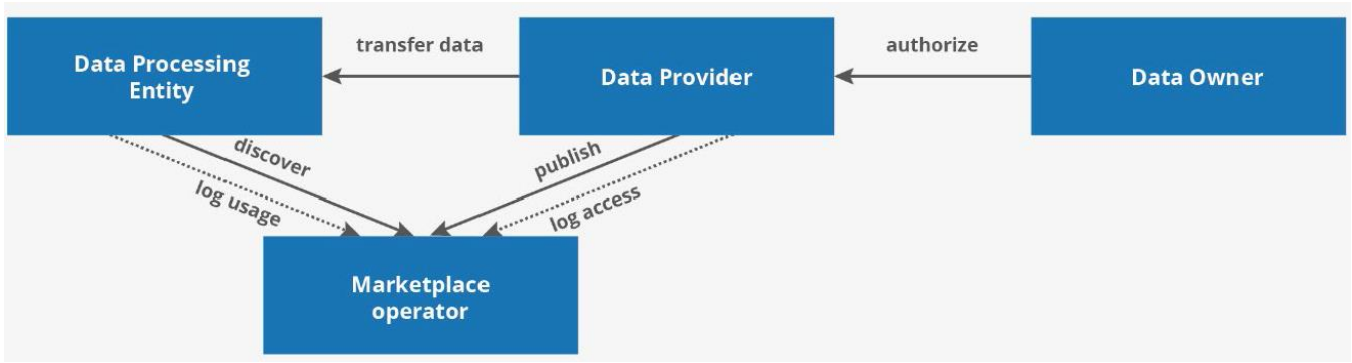


- **Usage control**

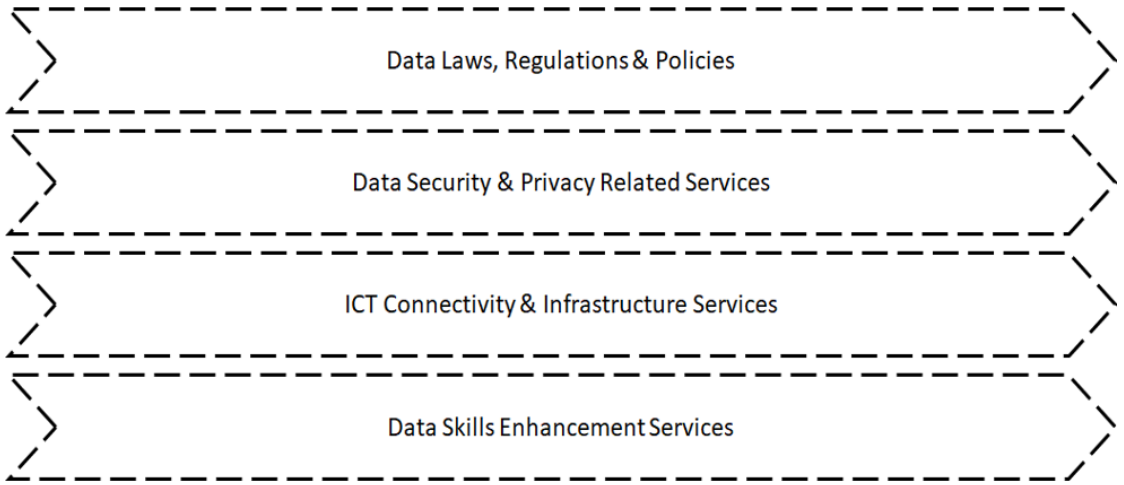


Key findings and challenges – 4

- **Business roles and interactions (business layer)**
  - User-driven approach
    - A user-friendly ecosystem
  - Ownership
  - Stakeholder management
  - Different functionalities
    - Consumer functionalities
    - Business functionalities
    - Public functionalities

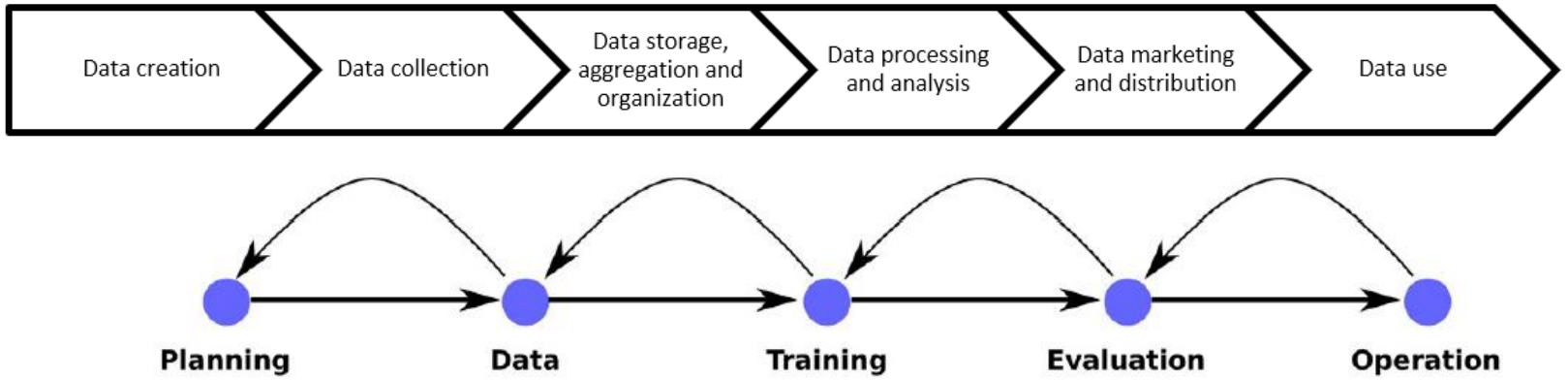


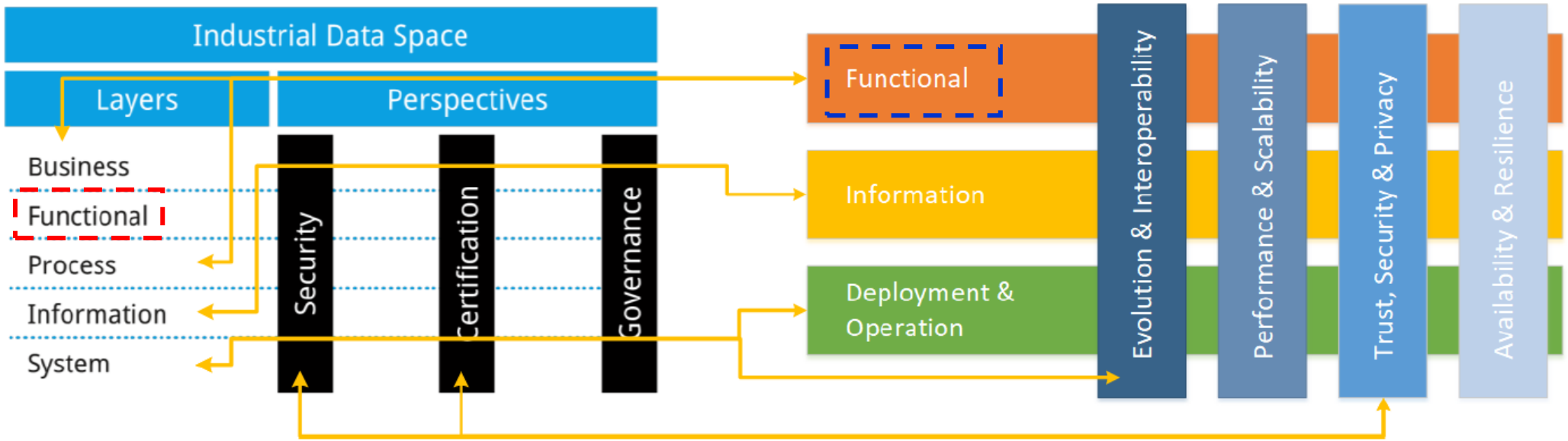
Data support activities



Key findings and challenges – 5

- **Data lifecycle**
  - Data and value flow
- **Operations (OT)**





IDS and EU IoT-A reference model

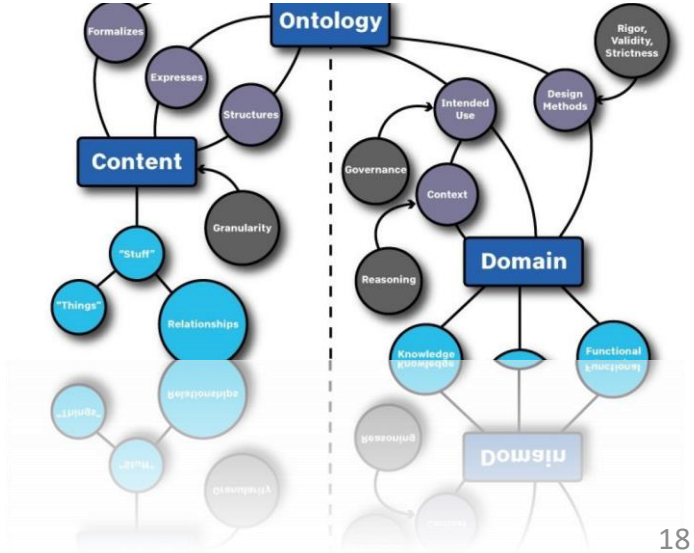
Key findings and challenges – 6

- A **common language** for Data Interoperability and Intelligence
  - Metadata as meaning and vocabulary package
  - Ontology as the foundation and capability of machine interpretation, inference, and logic
  - Semantics for better understanding
- Key roles in **knowledge discovery and data federation for shared meaning**

GOUI: “Global Observatory for Urban Intelligence”

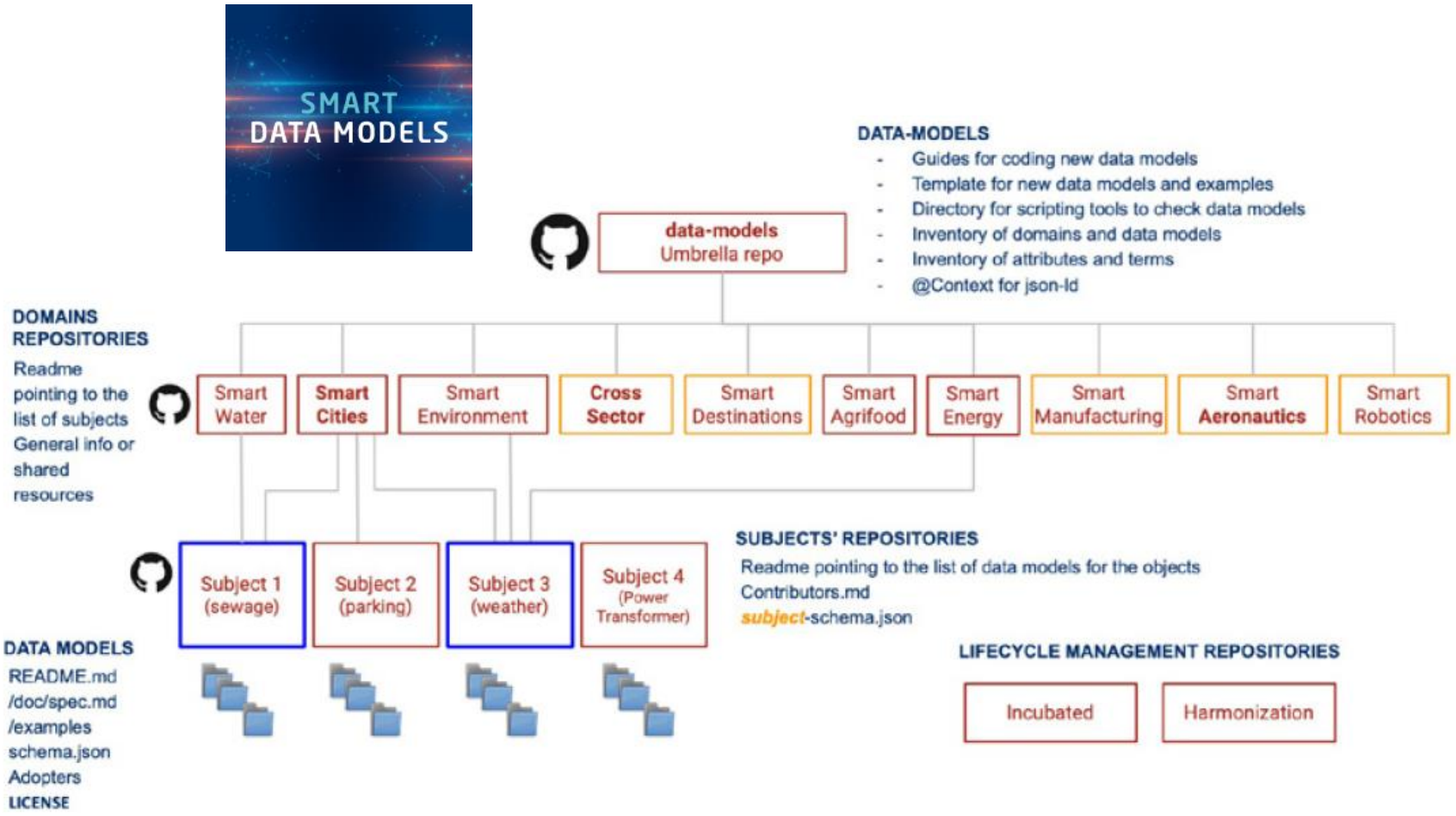
In June 2021, IEEE and ITU initiated a joint-collaboration to develop GOUI

- Create a NEW Smart Cities **Ontology** as a common language
- Correlations via **semantics**
- **Digital Twins** - model cities to better understand them



Key findings and challenges – 7

- **Common data models**
  - Domain-agnostic
  - Represented in formats compatible with the API

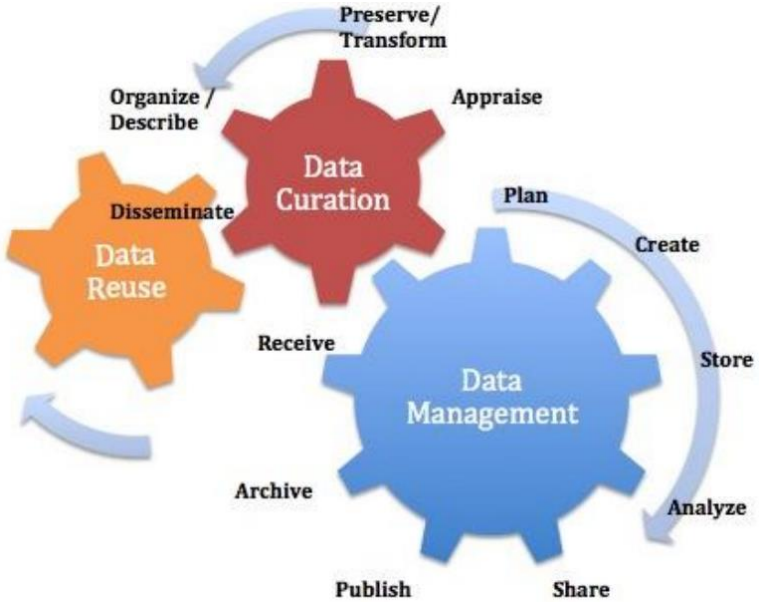


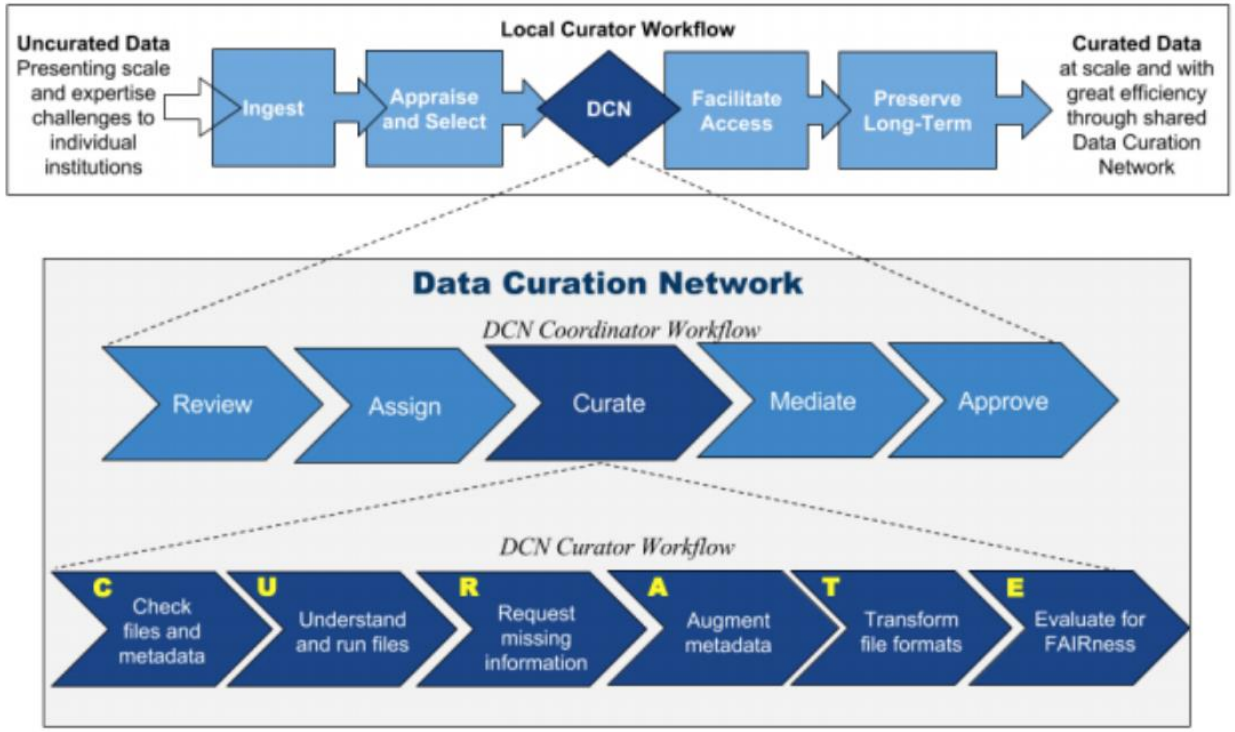
(source) https://github.com/smart-data-models (A program led by FIWARE, IUDX, TM Forum, OASC and others)

Key findings and challenges – 8

Findable Accessible Interoperable Reusable

- **Data curation** for maintaining the value of data
  - Data are organized, described, cleaned, enhanced and preserved for public use





The need for explanations
(Human + AI)

Data Curation Network DRYAD

Key findings and challenges – 9

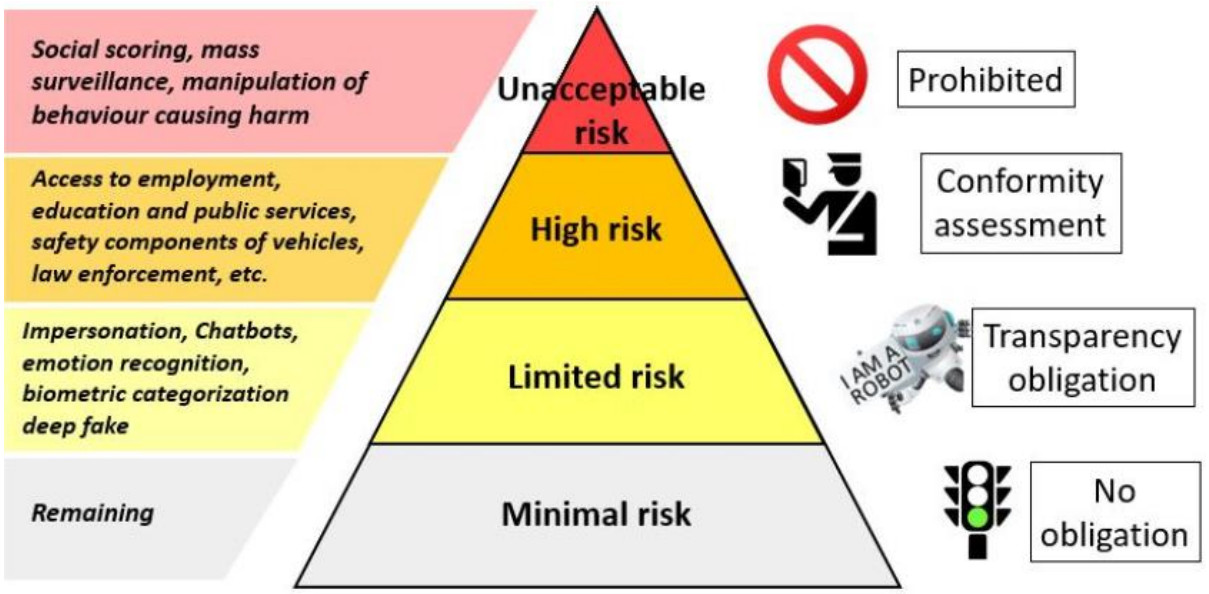
- **Trust in data sharing**
  - Consent to share
  - Control of personal data
  - Privacy (GDPR Compliance)
  - Transparency
  - Accessibility
  - Fairness
  - Accountability
  - Security and data integrity

Utility vs. Privacy

COMMONS

Digital Public Goods Alliance

- **Risk management**
  - Federated security management
  - Federated privacy management
  - Federated assurance management

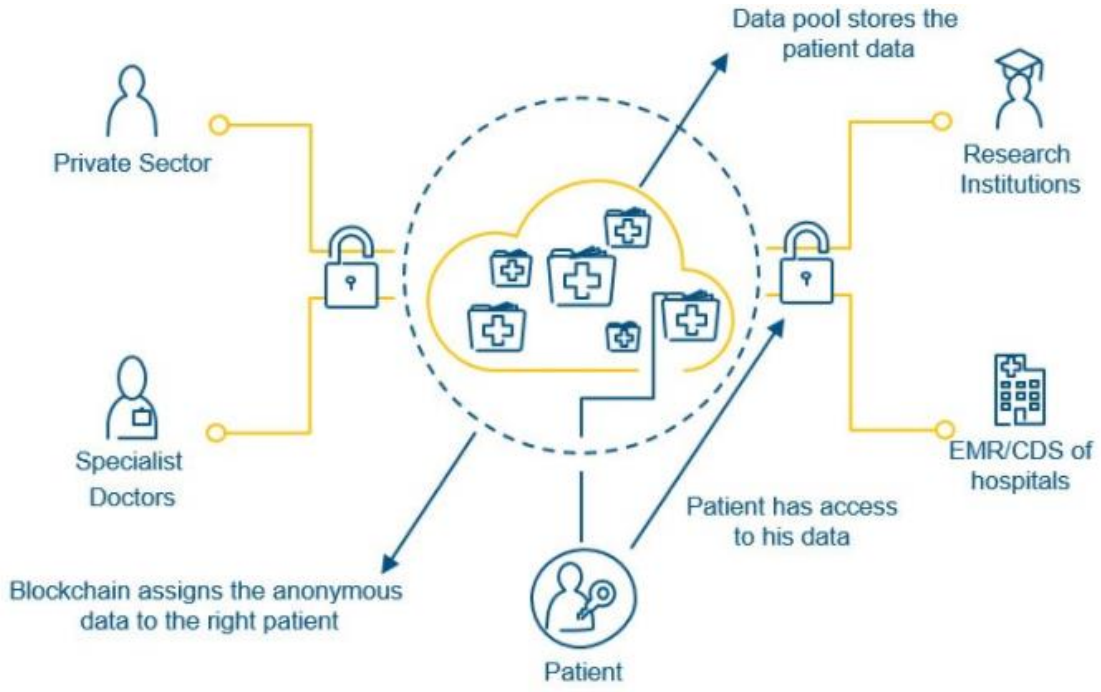


(Source: Telefónica)

Key findings and challenges – 10

- **Decentralization**
  - A decentralised architecture agreed upon by all relevant stakeholder groups with Blockchain
  - Blockchain enabled value creation

Internet of Value

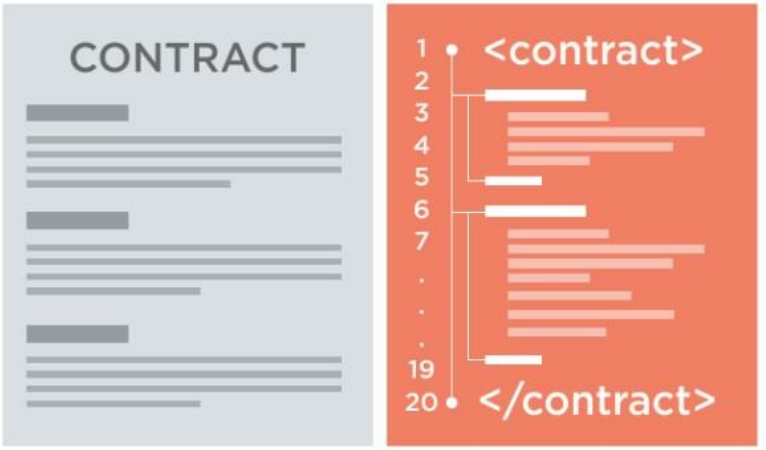


Framework of medical records in Europe

Key findings and challenges – 11

- **Governance**
  - Rights and Responsibilities
    - What actions can be taken
    - By whom
    - With what data
  - Compliance
- **Ethics**
- Key performance indicators (KPI)

Blockchain and smart contracts



MultiChain Governance

The blockchain as the "perfect code of law"

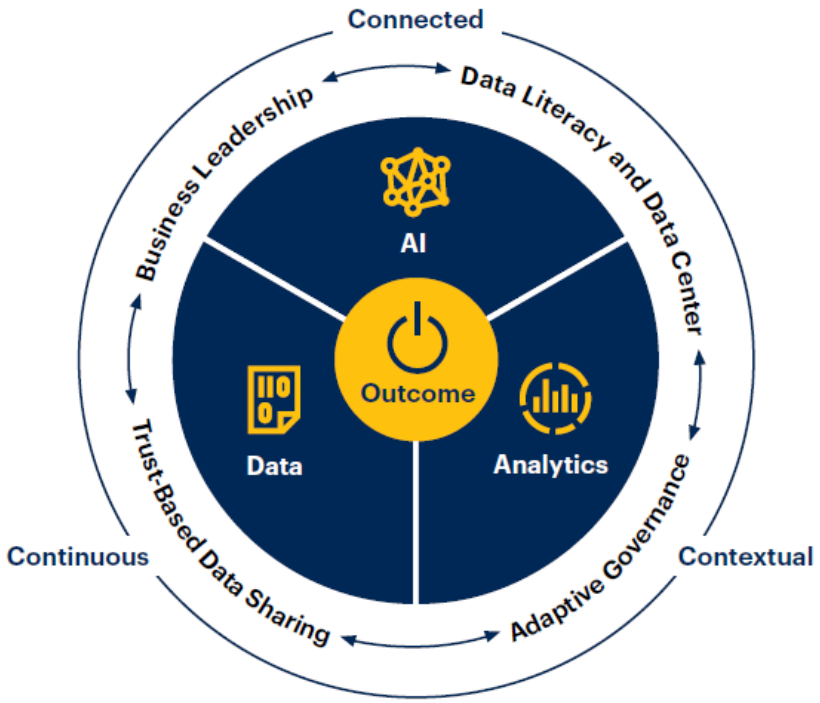
DIN SPEC 4997 Privacy by Blockchain Design

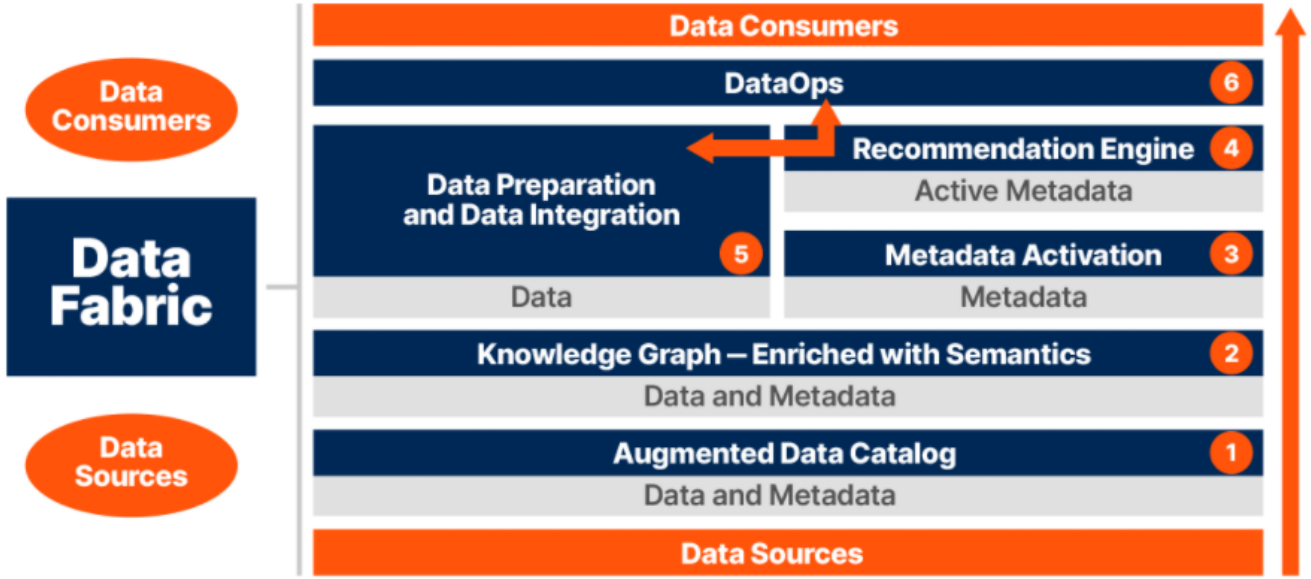
decode

DEcentralised Citizens Owned Data Ecosystem

Key findings and challenges – 12

- **Data Fabric**
  - **An integrated data management platform** that enables the full breadth of integrated data management capabilities including discovery, federated governance, curation, and orchestration.





Gartner Data Fabric: Top Strategic Technology Trend for 2022

Data-centric Ecosystems

GAIA-X – a European Data Ecosystem

"Project GAIA-X" as the cradle of an open, digital ecosystem where data can be made available, securely collated and shared while enjoying the trust of its users.

Principles based on European values

1. European data protection
2. Openness and transparency
3. Authenticity and trust
4. Digital sovereignty and self-determination
5. Free market access and European value creation
6. Modularity and interoperability
7. User-friendliness

Trend

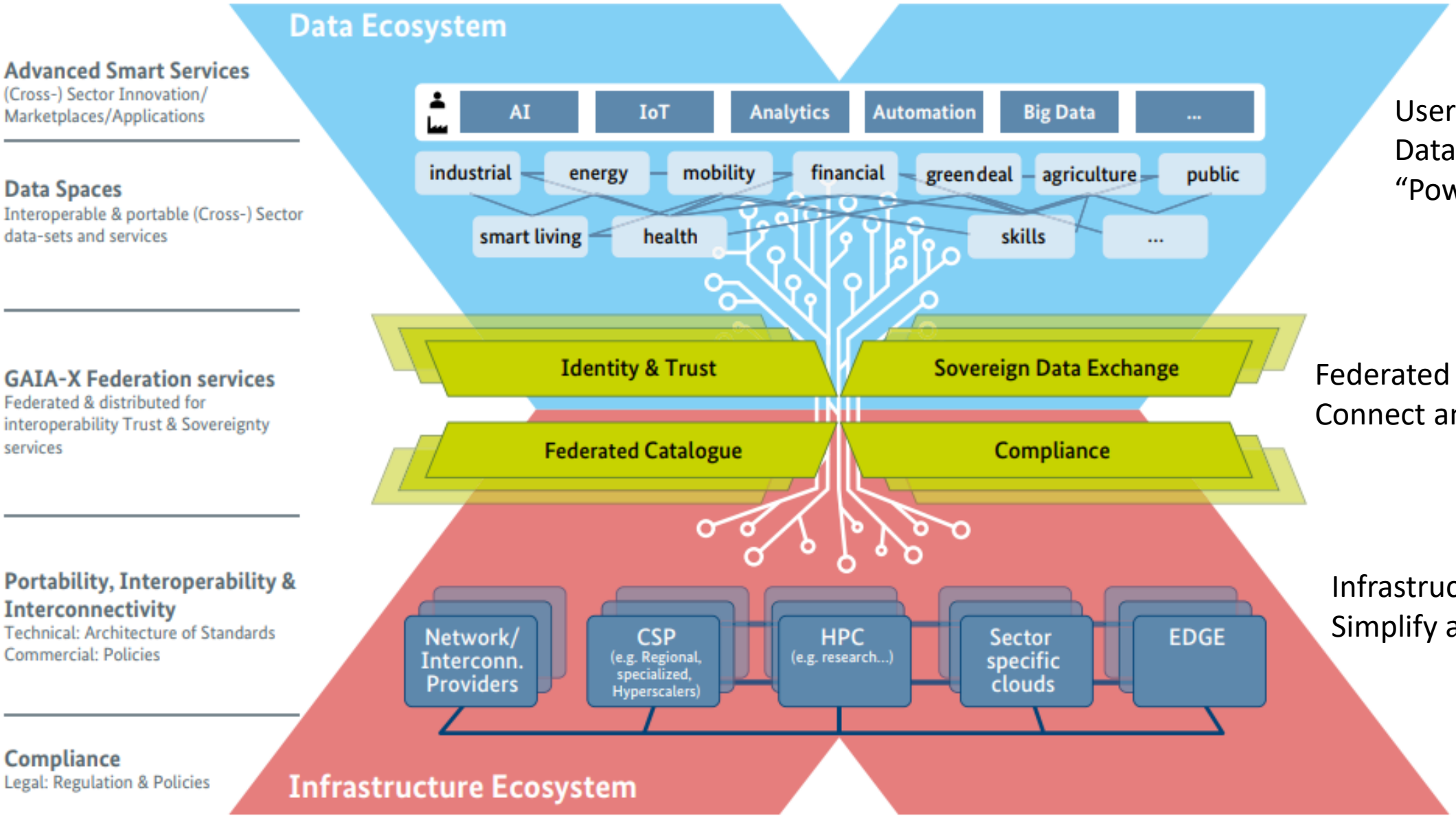
- Cloud Computing as the offer, use and charge of IT services
- Edge as a decentralized data architecture principle

Goals

1. We are striving for data sovereignty
2. We want to reduce dependencies (lock-in effects)
3. We want to make cloud services attractive on a broad basis
4. We are creating an ecosystem for innovation

(source) https://www.gaia-x.eu/

Architectural concept with GAIA-X federated services



User Domain
Data Ecosystems –
"Powered by GAIA-X"

Federated Services –
Connect and govern

Infrastructure Ecosystem –
Simplify and accelerate

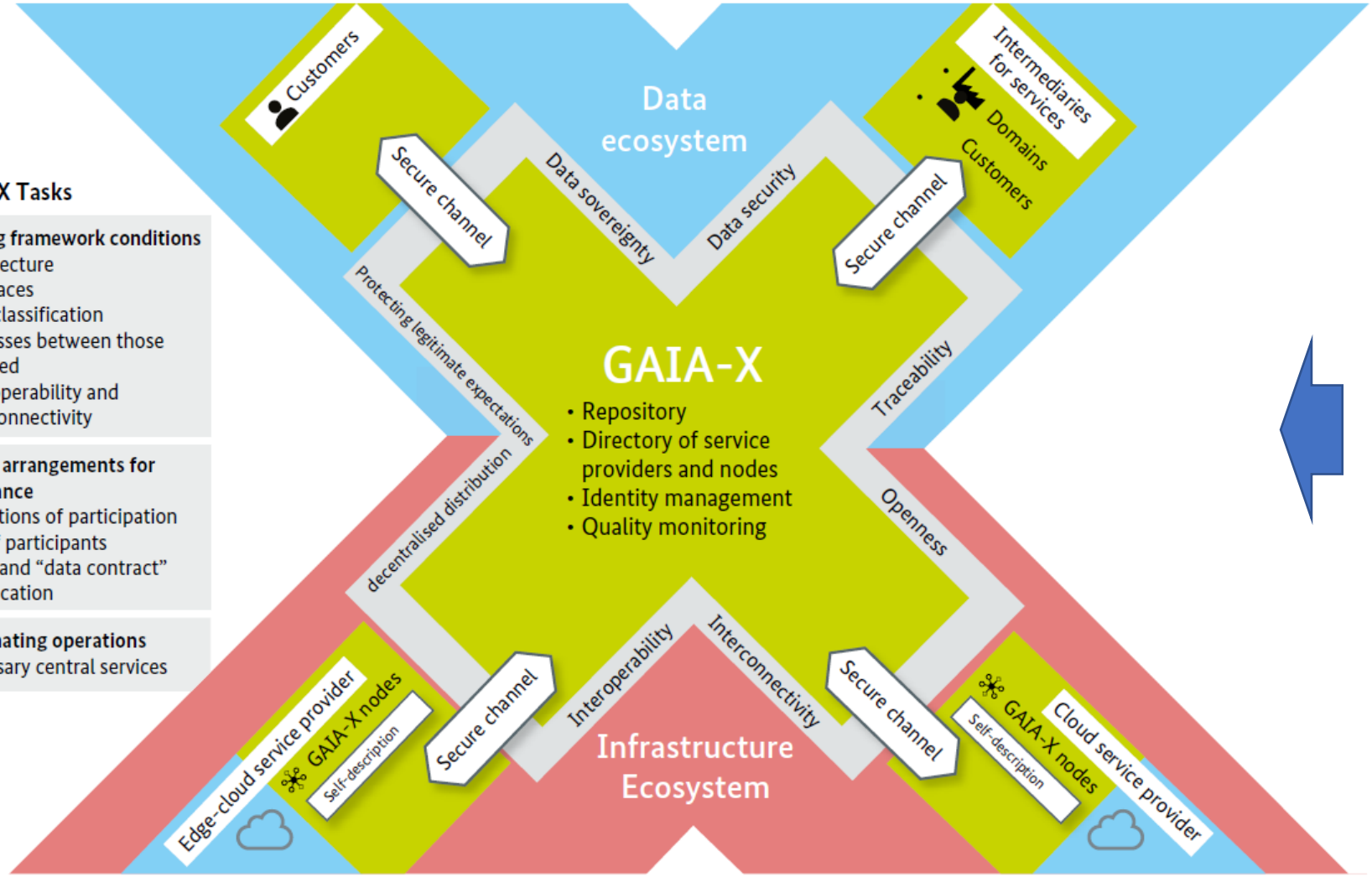
Overall picture of data infrastructure and ecosystem

GAIA-X Tasks

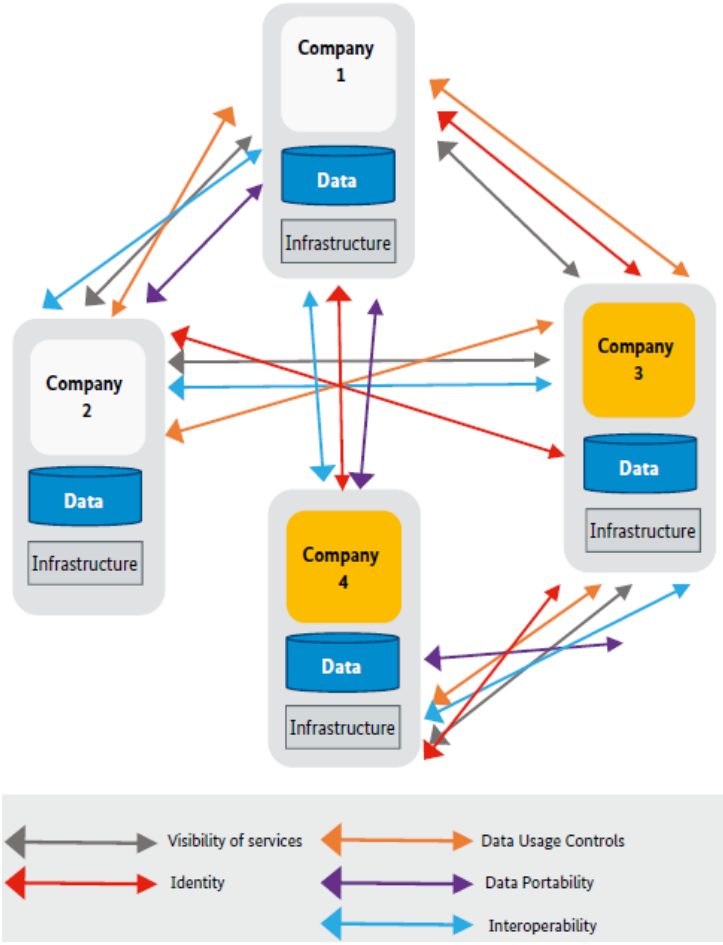
Creating framework conditions
- Architecture
- Interfaces
- Data classification
- Processes between those involved
- Interoperability and interconnectivity

Making arrangements for governance
- Conditions of participation
- List of participants
- Rules and "data contract"
- Certification

Coordinating operations
- Necessary central services



Collaboration today



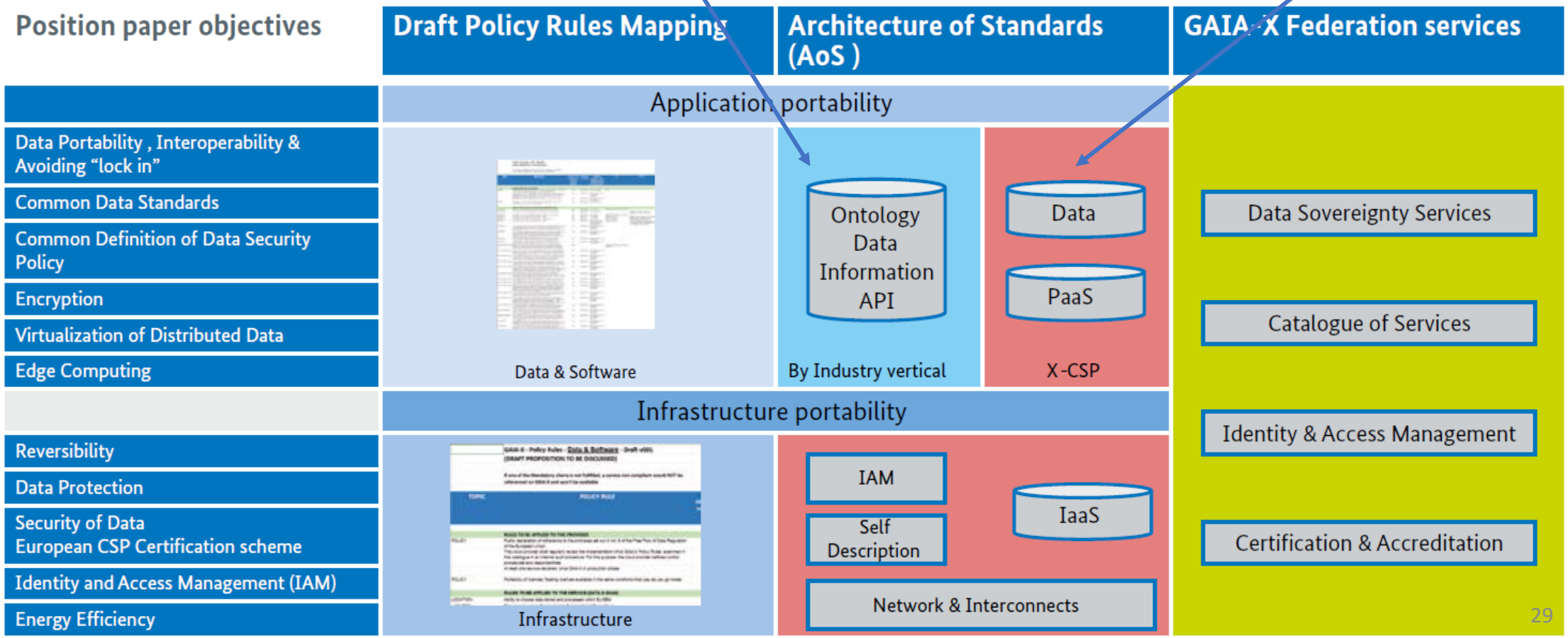
Standards by industry vertical

Definition of Ontologies, specific APIs (and their semantic), required technology standards and compliance defined by (existing) **Industry Associations**

Standards across CSP

Alignment of technical standards and set of technologies/products across Cloud Service providers provide
- Interoperability
- Portability

Mapping of Policy Rules and AoS



Data Spaces

INTERNATIONAL DATA SPACES ASSOCIATION

A data space is defined as a decentralized infrastructure for trustworthy data sharing and exchange in data ecosystems based on commonly agreed principles

- Data platforms
  - providing support for effective data sharing and exchange as well as for engineering and deployment of data exchange and processing capabilities
- Data marketplaces
  - where data providers can offer and data consumers can request data, as well as data processing applications
- Data sovereignty
  - i.e. the ability for each stakeholder to control their data by making decisions as to how digital processes, infrastructures, and flows of data are structured, built and managed, based on an appropriate governance scheme enabling specification of terms and conditions.

(source) International Data Spaces Association, https://internationaldataspaces.org/

Design principles

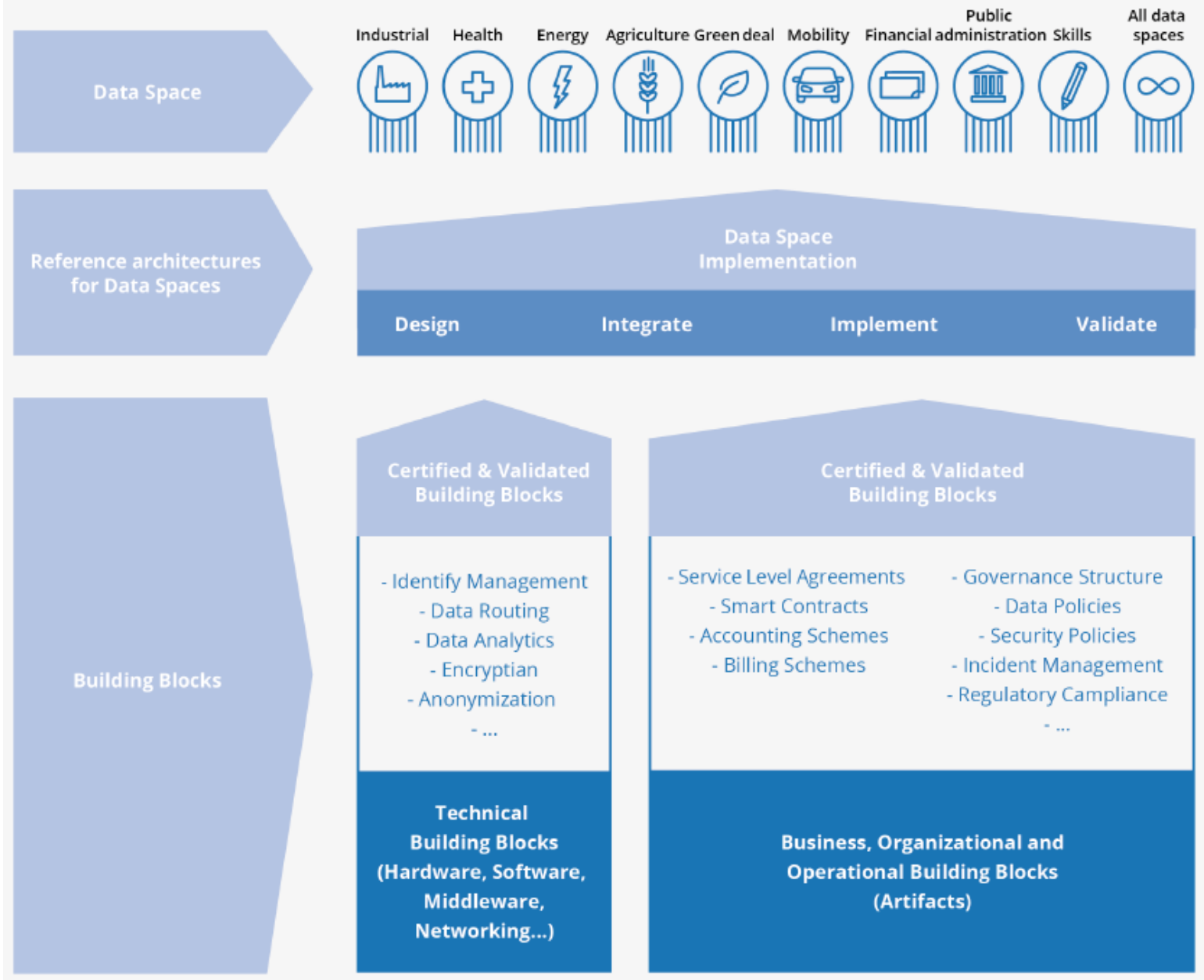
1. Data sovereignty
2. Data level playing field
3. Decentralised soft infastructure
4. Public-private governance

Architecture requirements

Data Sharing

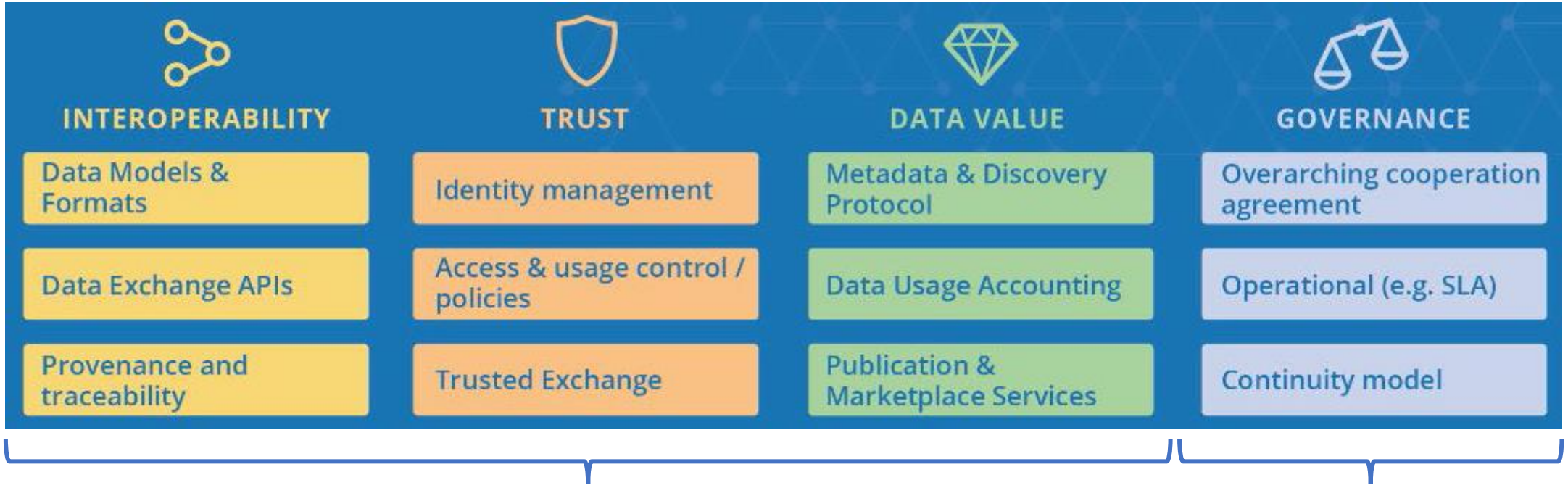
- Data-sharing empowerment
- Data-sharing trustworthiness
- Data-sharing publication
- Data-sharing economy
- Data-sharing interoperability
- Data space engineering flexibility
- Data space community

Data space solution based on the synthesis of building blocks

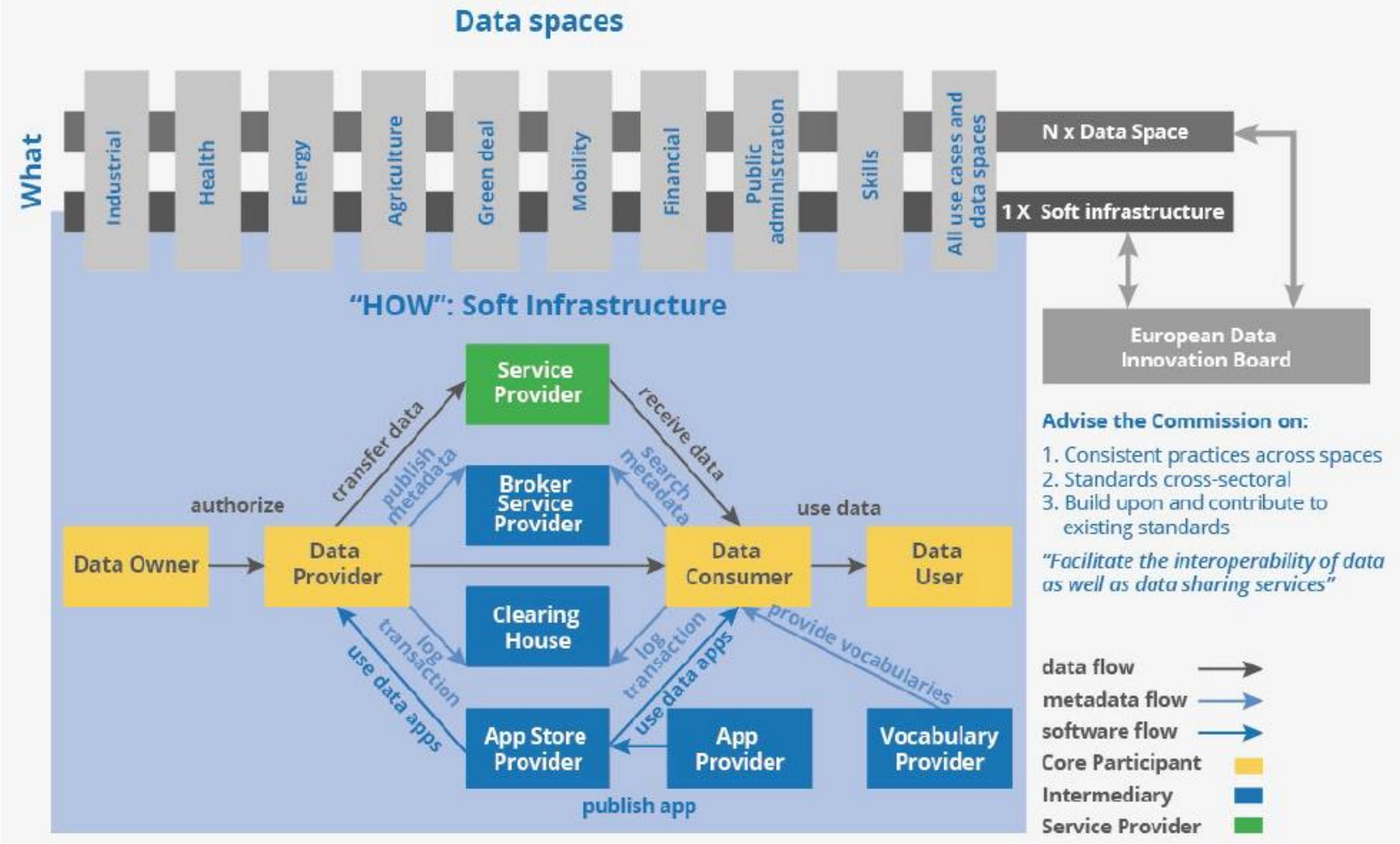


Data spaces building blocks

- Technical building blocks + Governance building blocks



Data governance act



International Data Spaces Association reference architecture

Activities per area for governance

Framework management and innovation

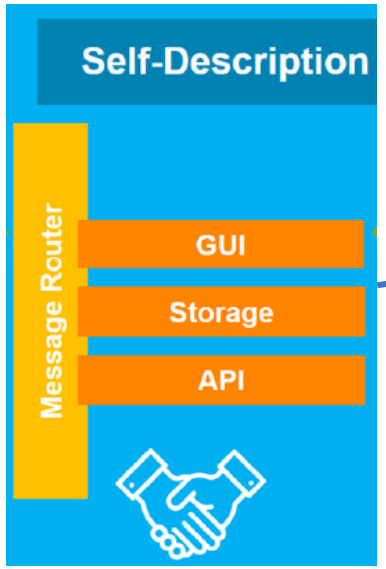
1. Maintenance and innovation
2. Accession and certification

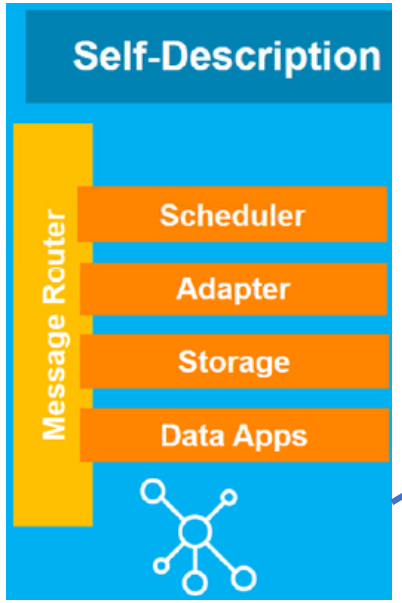
Adoption: Implementation, support, communication

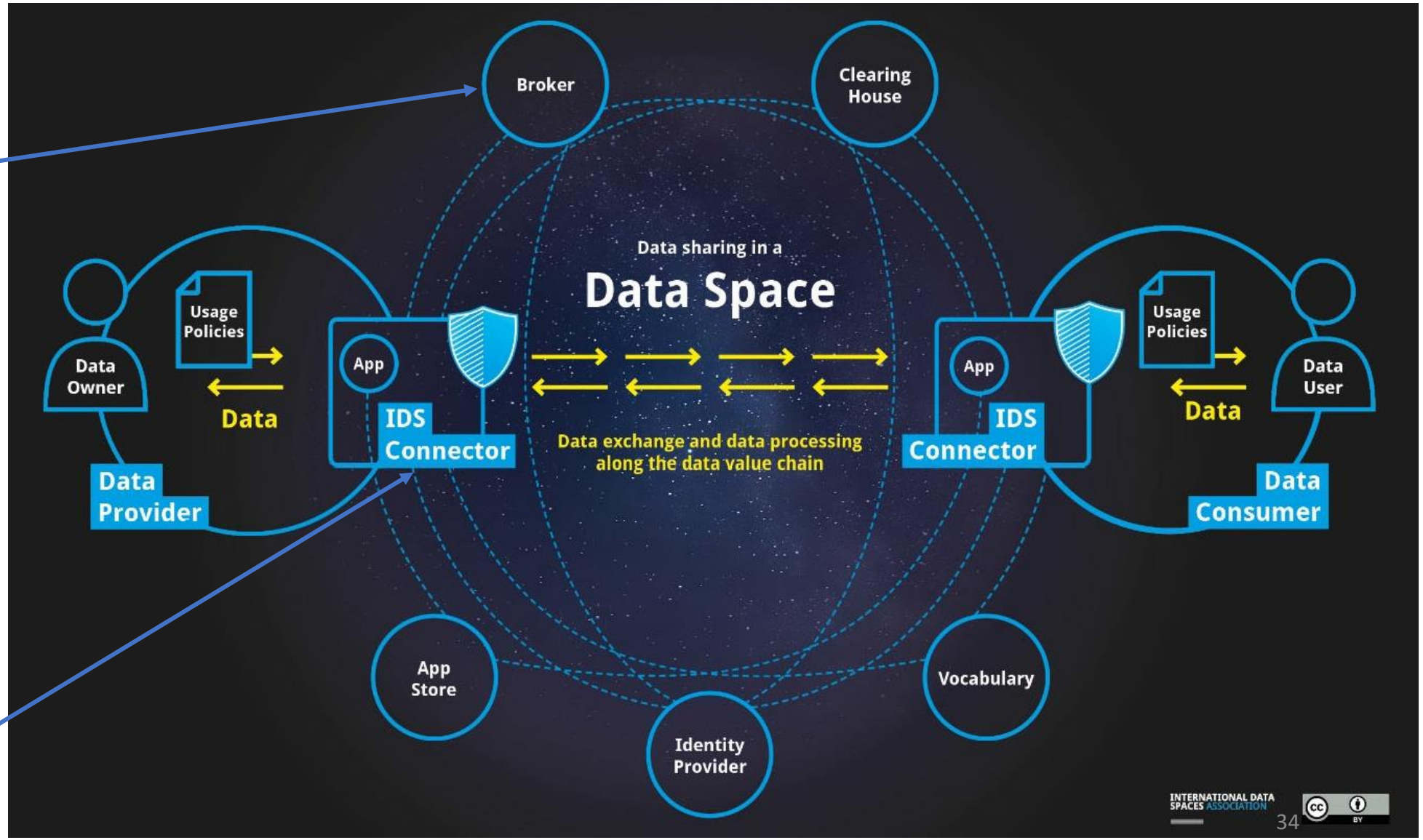
3. Technical and implementation support
4. Communication and education

Governance: Business, operational and organizational agreements among data space participants

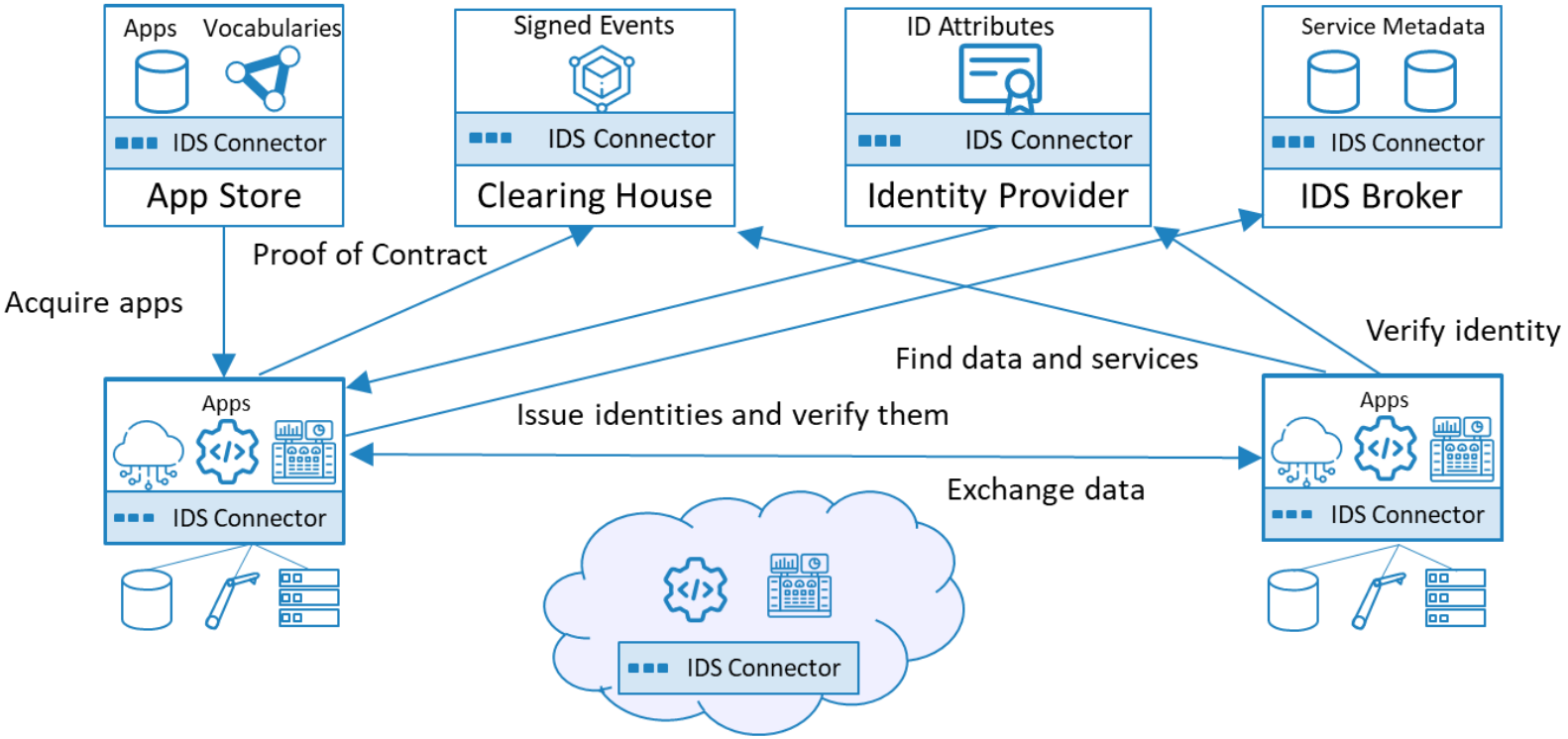
Basic architectural concepts of the International Data Spaces (IDS)

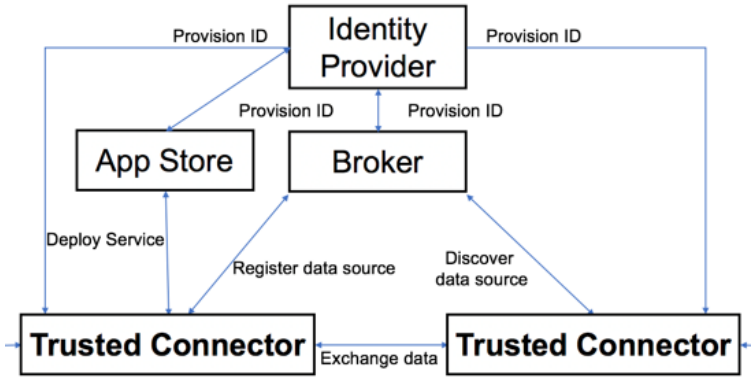


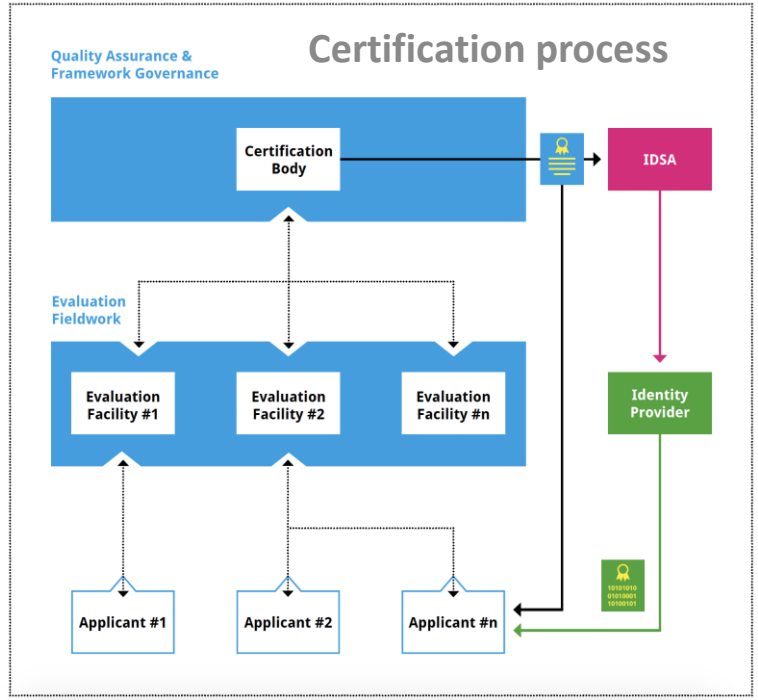




Target vision: International Dataspaces







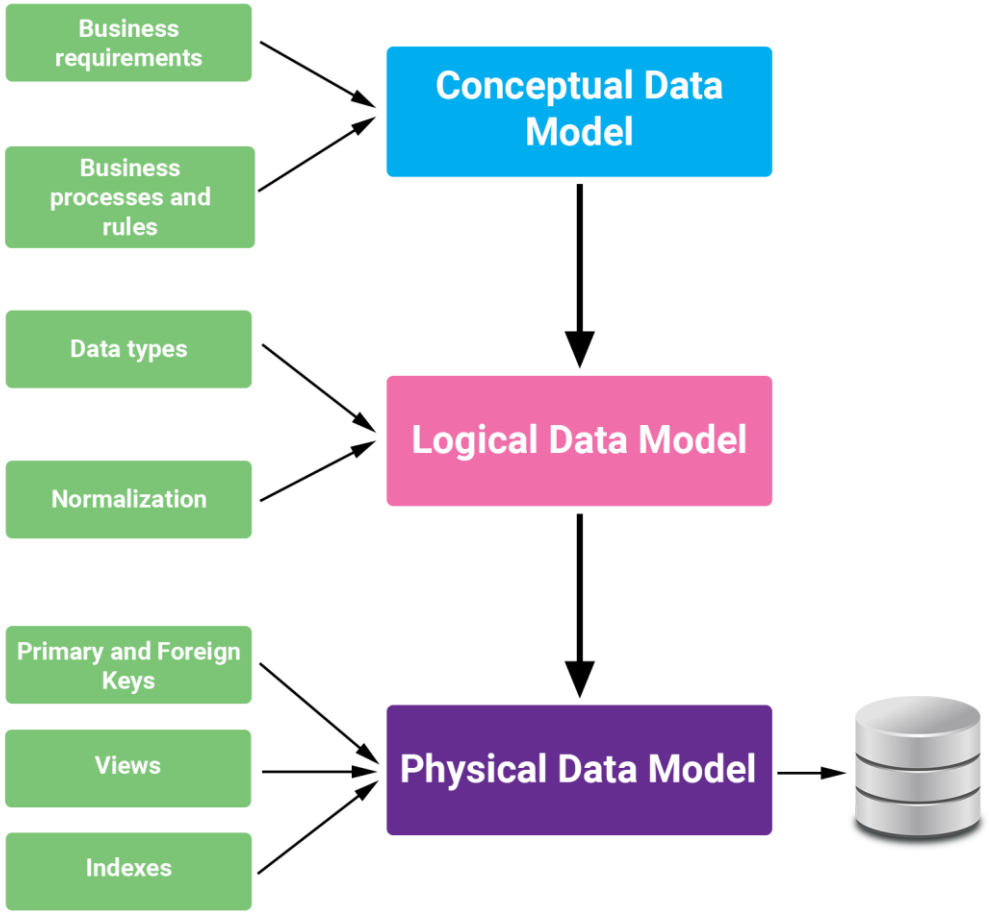
INTERNATIONAL DATA SPACES

Fraunhofer AISEC

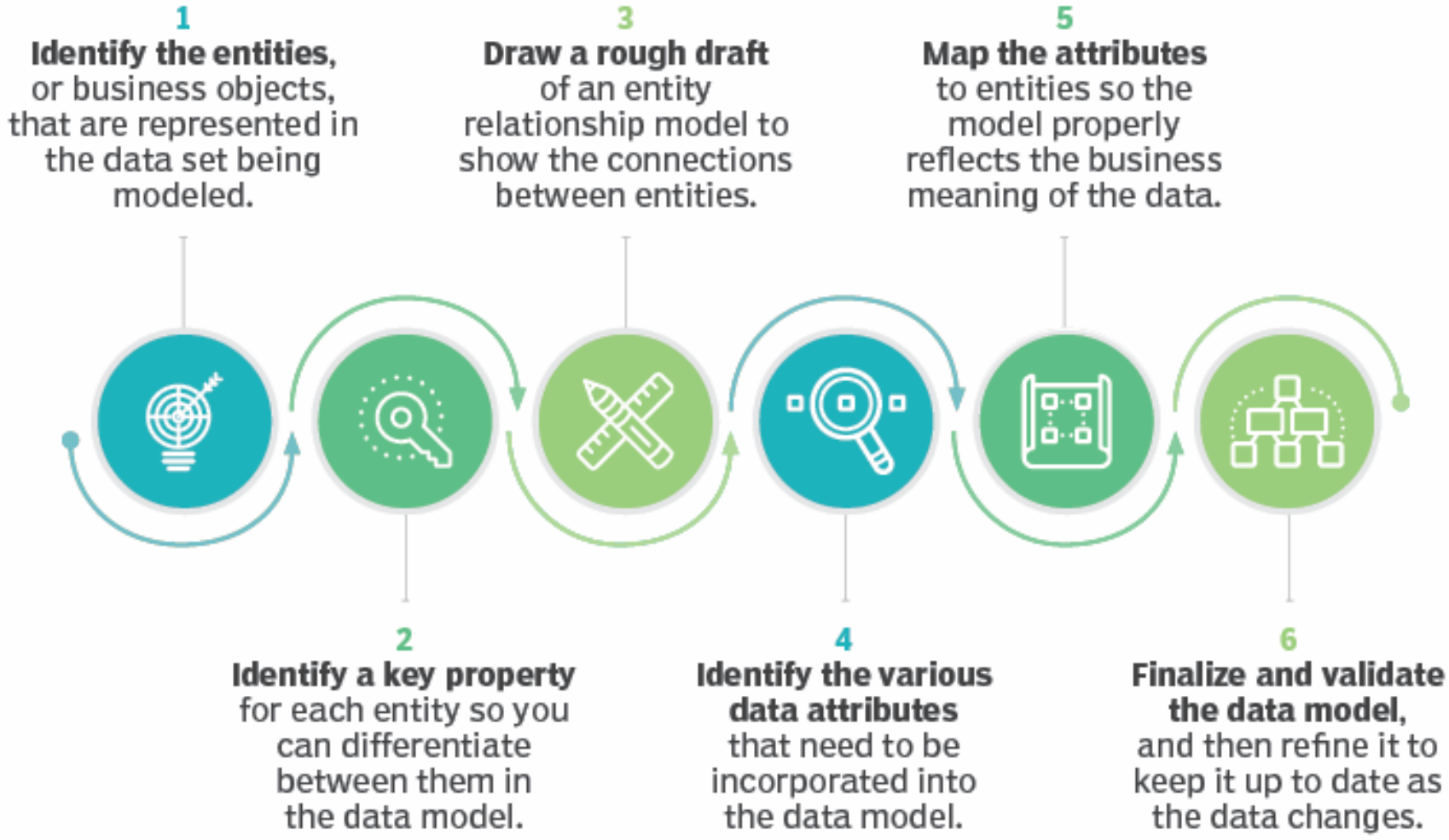
Data Modeling Technology

Types of data models

- Conceptual data models (domain models)
  - offer a big-picture view of what the system will contain, how it will be organized, and which business rules are involved.
- Logical data models
  - less abstract and provide greater detail about the concepts and relationships in the domain under consideration.
- Physical data models
  - provide a schema for how the data will be physically stored within a database.



Data modeling process

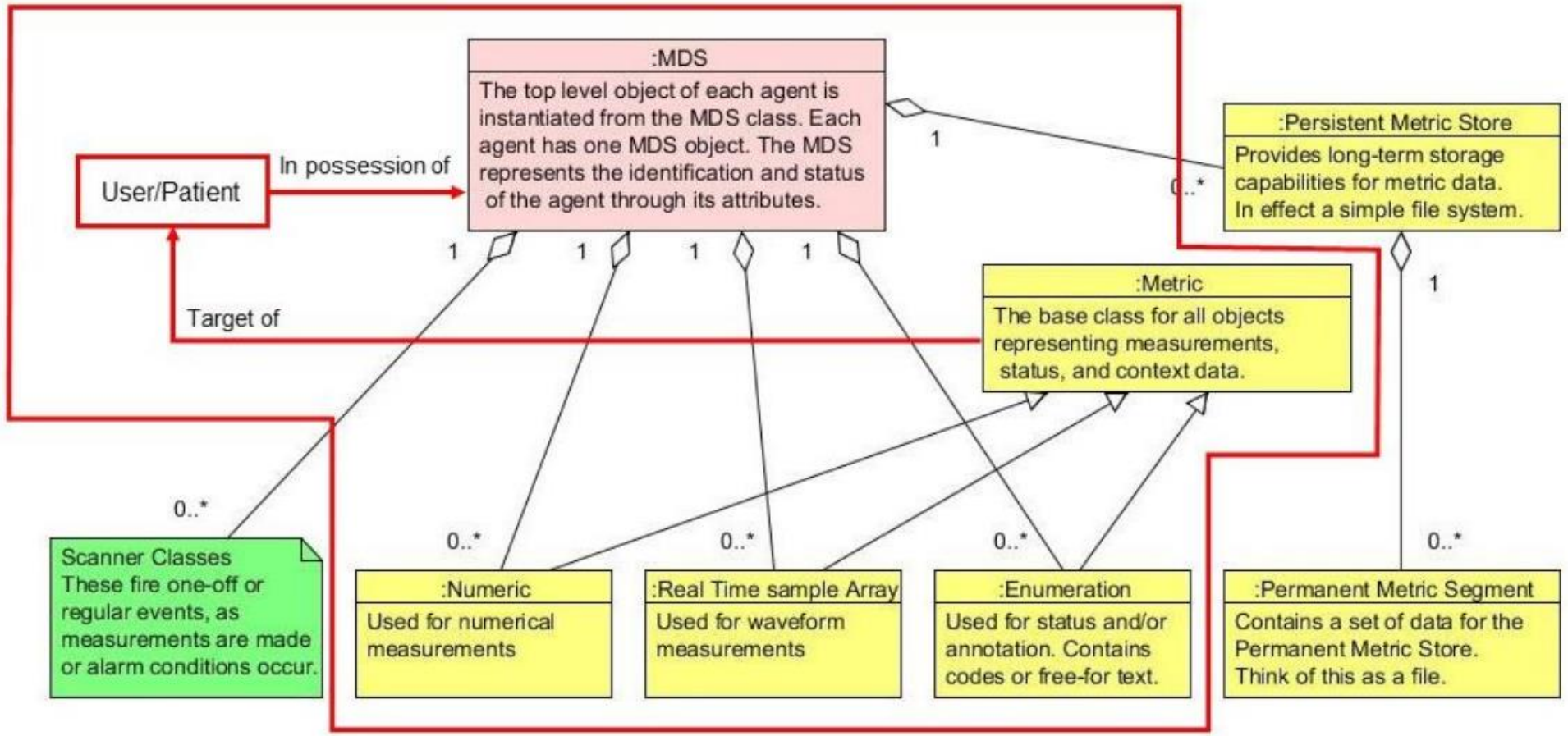


Types of data modeling

- Hierarchical data models
- Relational data models

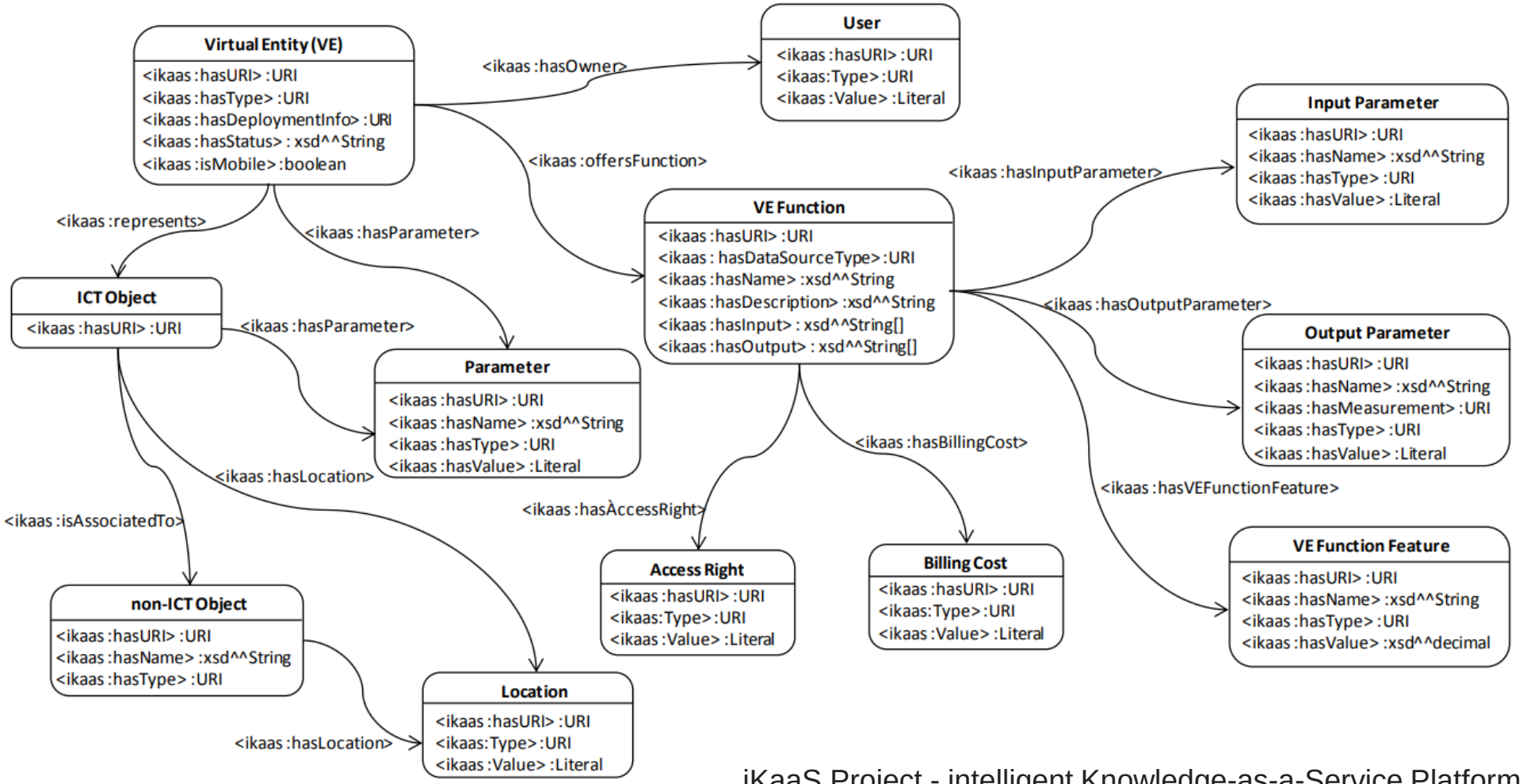
- Entity-relationship (ER) data models
  - represent the relationships between entities in a database
- Object-oriented data models
  - The “objects” involved are abstractions of real-world entities. Objects are grouped in class hierarchies, and have associated features.
- Dimensional data models
  - optimize data retrieval speeds for analytic purposes in a data warehouse.

Example – Heath data model



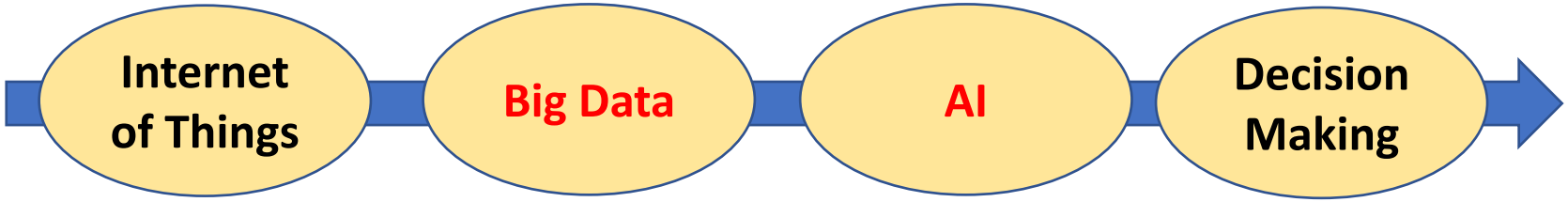
IEEE Std 11073-20601™-2014: Health informatics - Personal health device communication - Part 20601: Application profile - Optimized Exchange Protocol

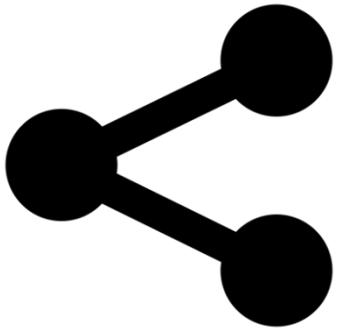
Example – Virtual entity data model



Related Standardization

Why Standards?





Data **Share** **Reuse**

Data and AI Ecosystem (Stakeholders, Value-chain)

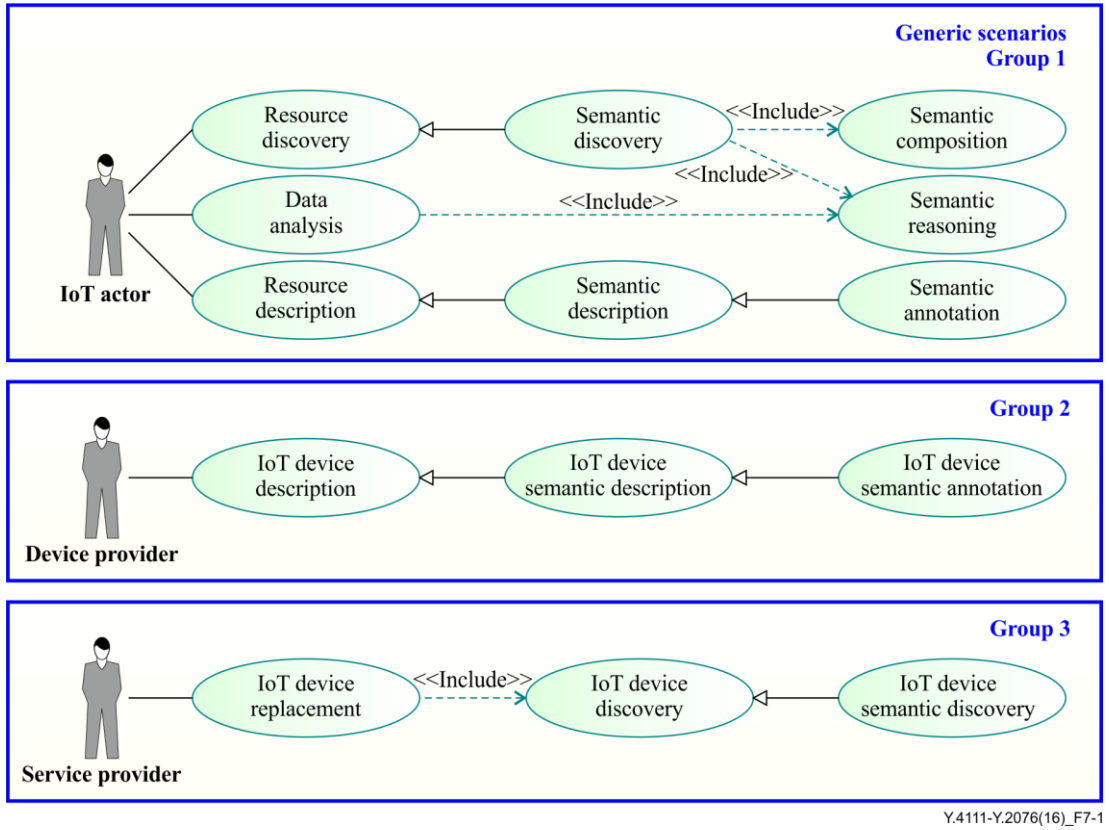
Work Items on data model related studies in ITU-T

- Y.4111 (02/2016) Semantics based requirements and framework of the Internet of things
- Y.4203 (02/2019) Requirements of things description in the Internet of Things
- Y.Suppl.69 (05/2021) Web based data model for IoT and smart city systems and services

Y.4111 - Semantics based requirements and framework of the Internet of things

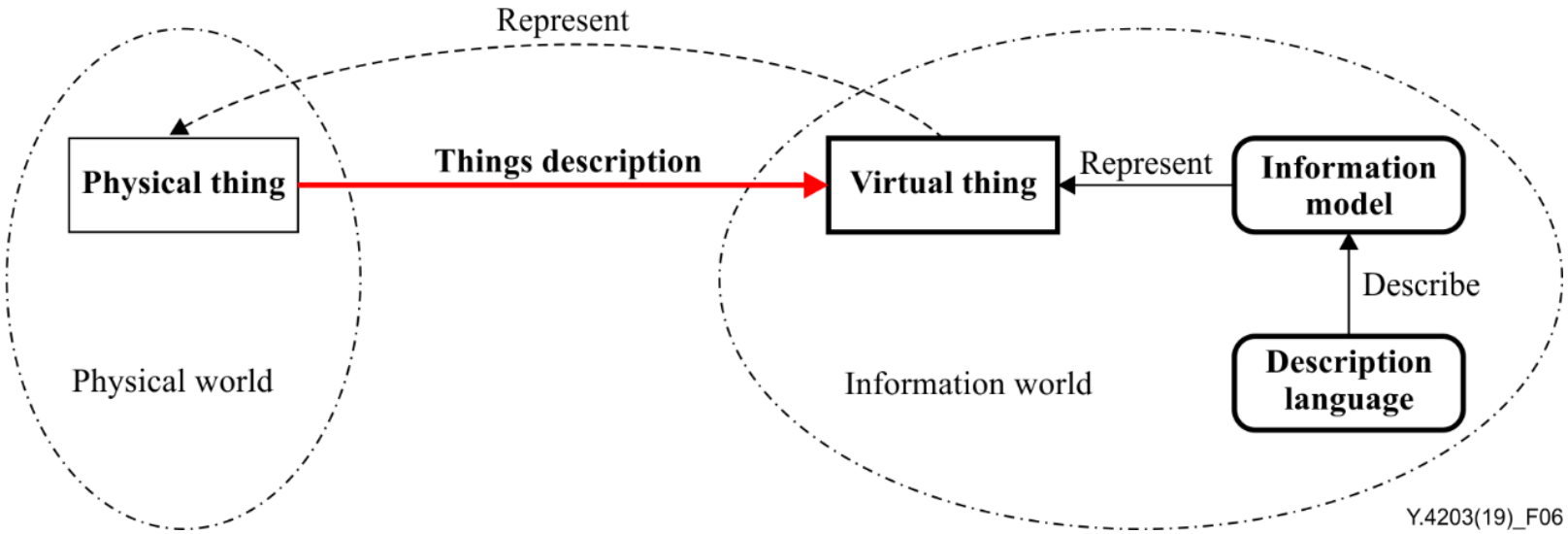
- Essential requirements of the IoT infrastructure for data and services:
  - interoperability, scalability, discovery, consistency, reusability and composability, analytics and reasoning for actionable intelligence, automatic operations
- Semantics based requirements for IoT
  - IoT ontology, Semantic annotation, Semantic interoperability, Semantic discovery, Semantic reasoning, Semantic composition

Semantics based use cases for IoT actors



Y.4203 - Requirements of things description in the Internet of Things

- Things description specifies a general way for how to map physical things in the physical world to virtual things in the information world in order to enable things of the IoT to be effectively discovered, interpreted and used.



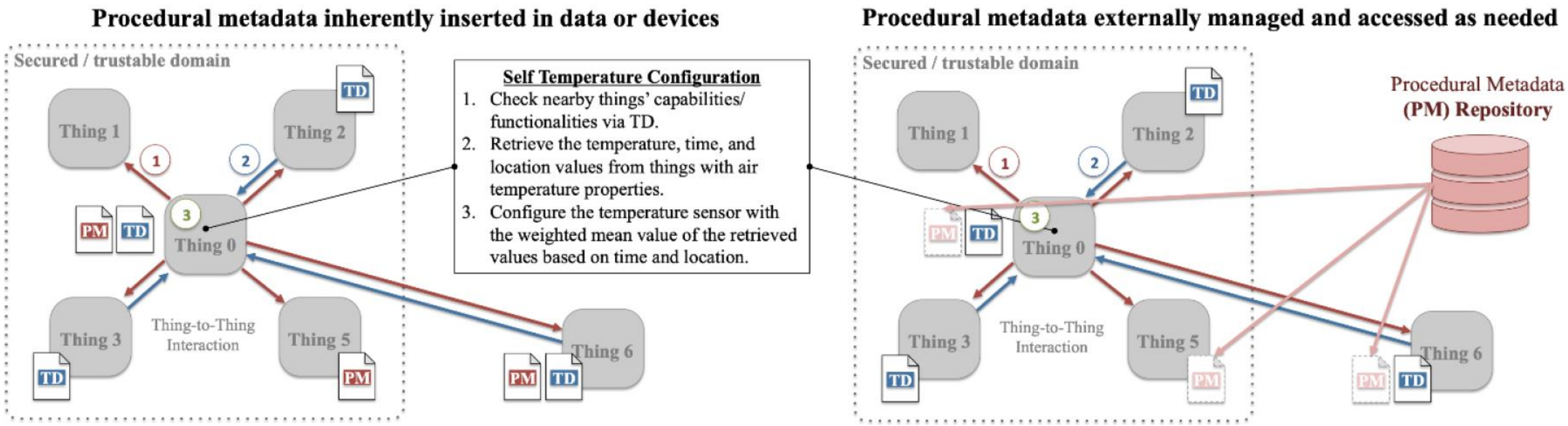
Y.Suppl.69 - Web based data model for IoT and smart city systems and services (1)

- Metadata: common descriptions of devices and data
  - Descriptive metadata*: information for finding or understanding a resource
  - Administrative metadata*: information to help manage the data resource
  - Structural metadata*: information on how the components are organized, including relationships
- Procedural metadata
  - to provide the common descriptions on composable procedures of not only individual devices but also smart systems as a whole based on existing data models and ontologies

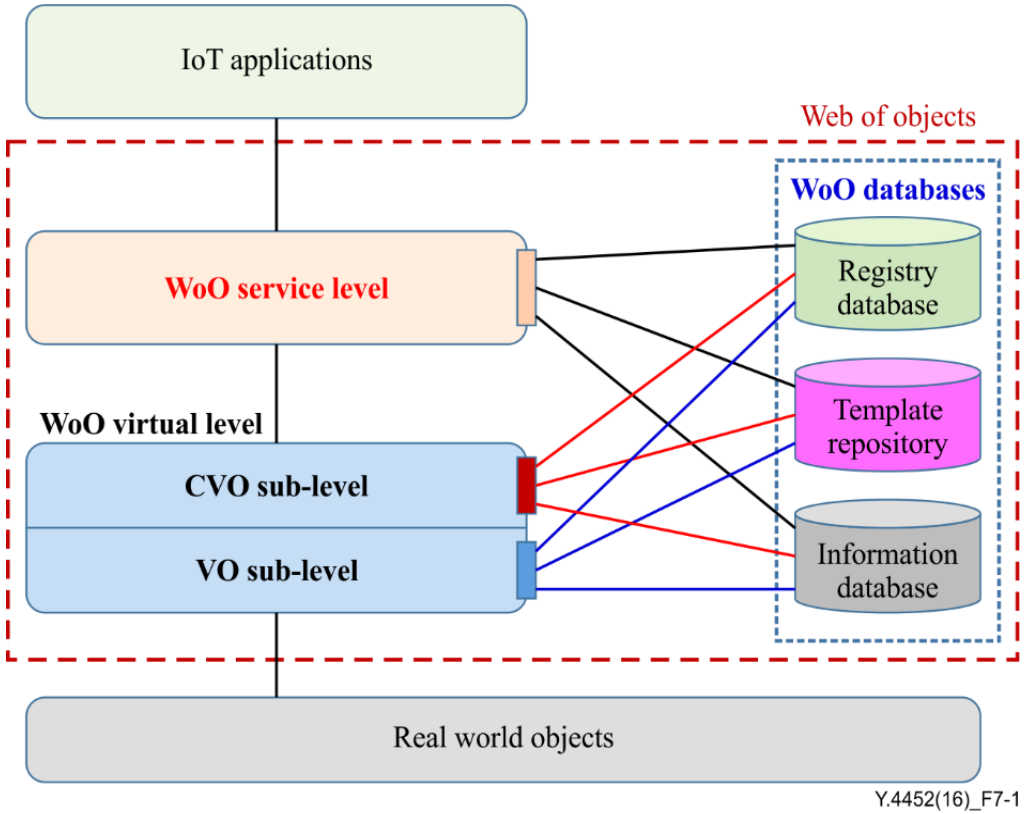
* J. Riley (2017). *Understanding Metadata: What is Metadata, and What is it For?: A Primier*. National Information Standards Organization (NISO). Jan. 2017.

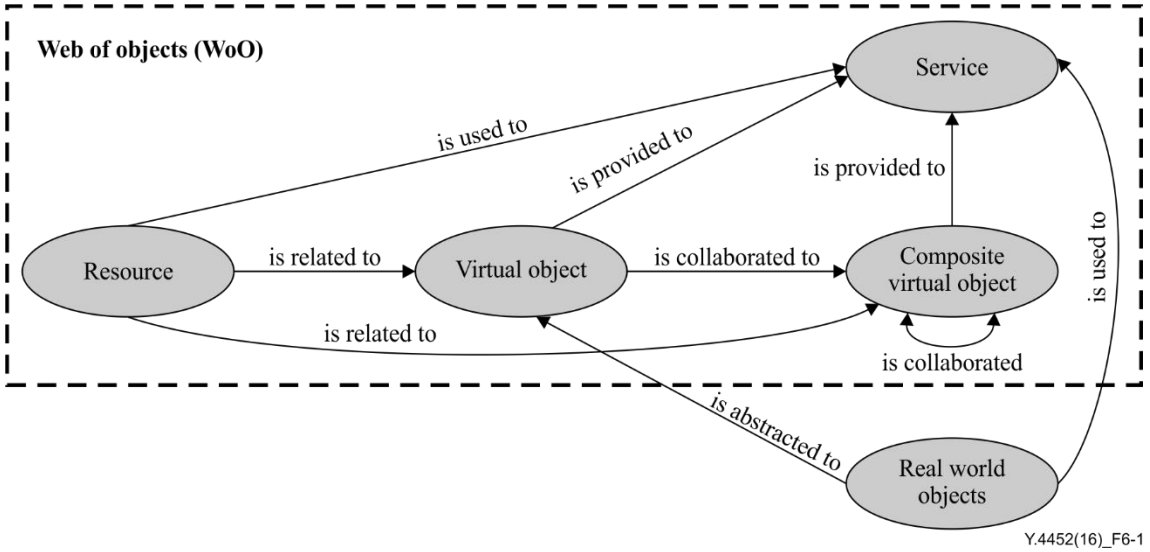
Y.Suppl.69 - Web based data model for IoT and smart city systems and services (2)

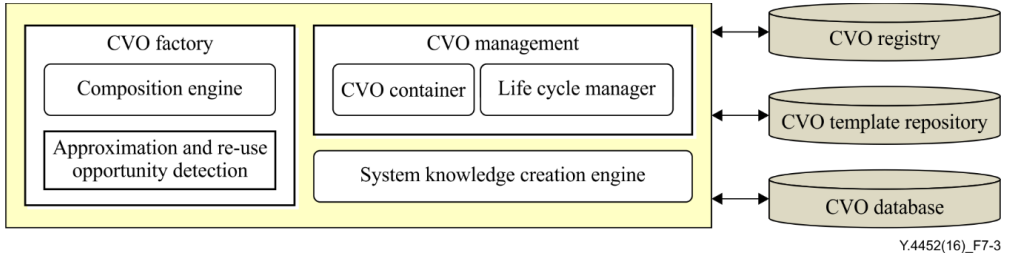
- Procedural metadata: Example

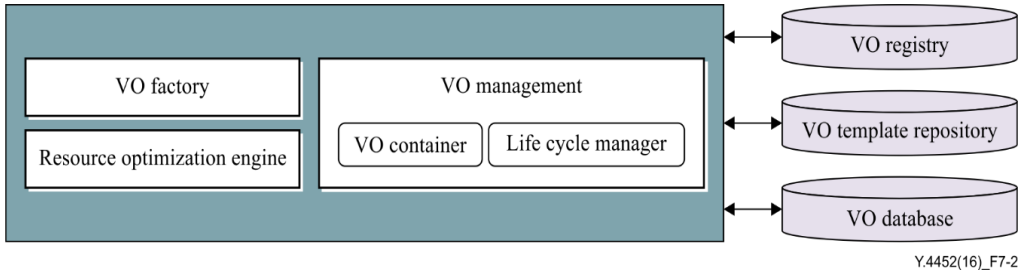


Y.4452 - Functional framework of web of objects

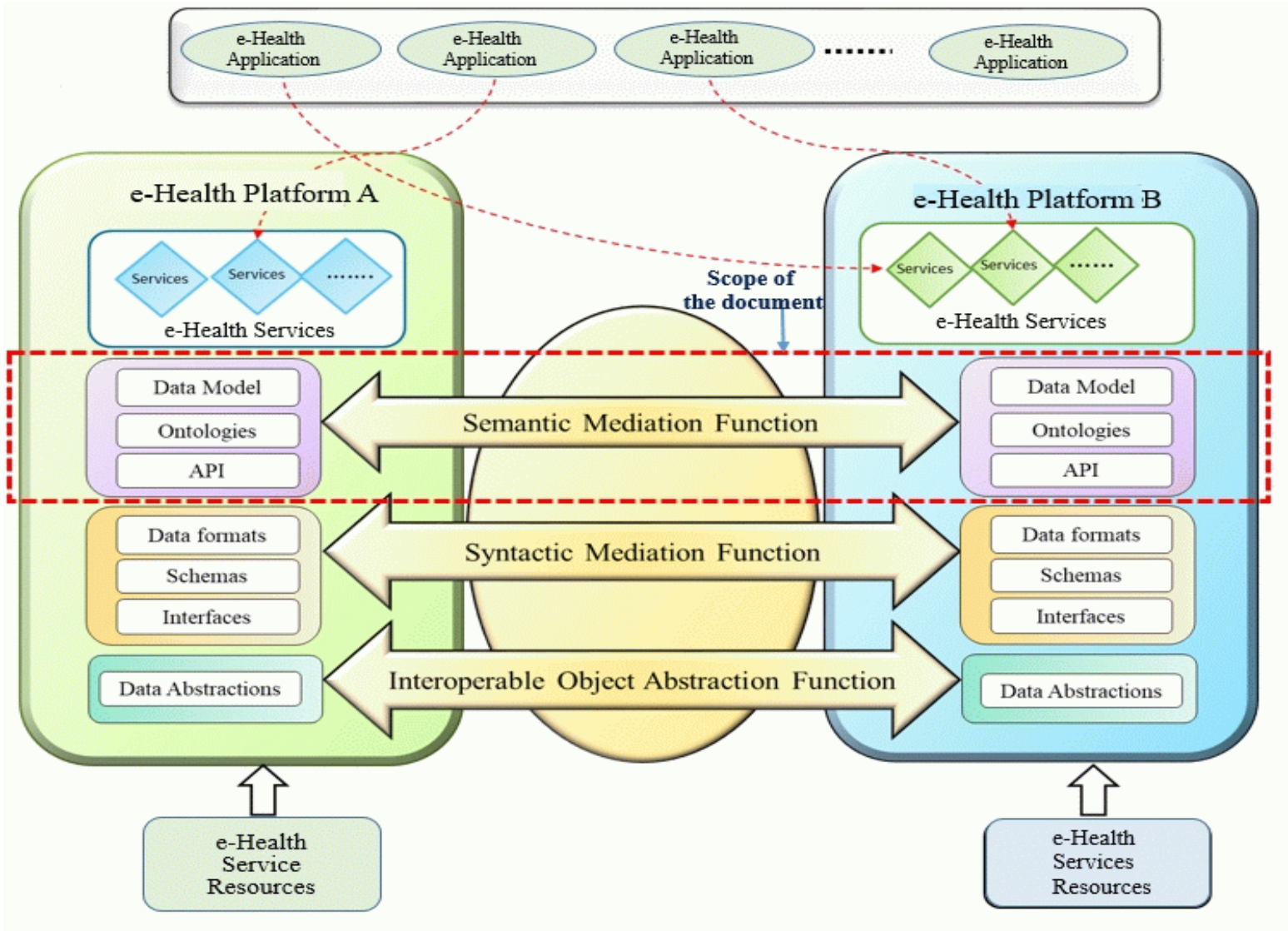








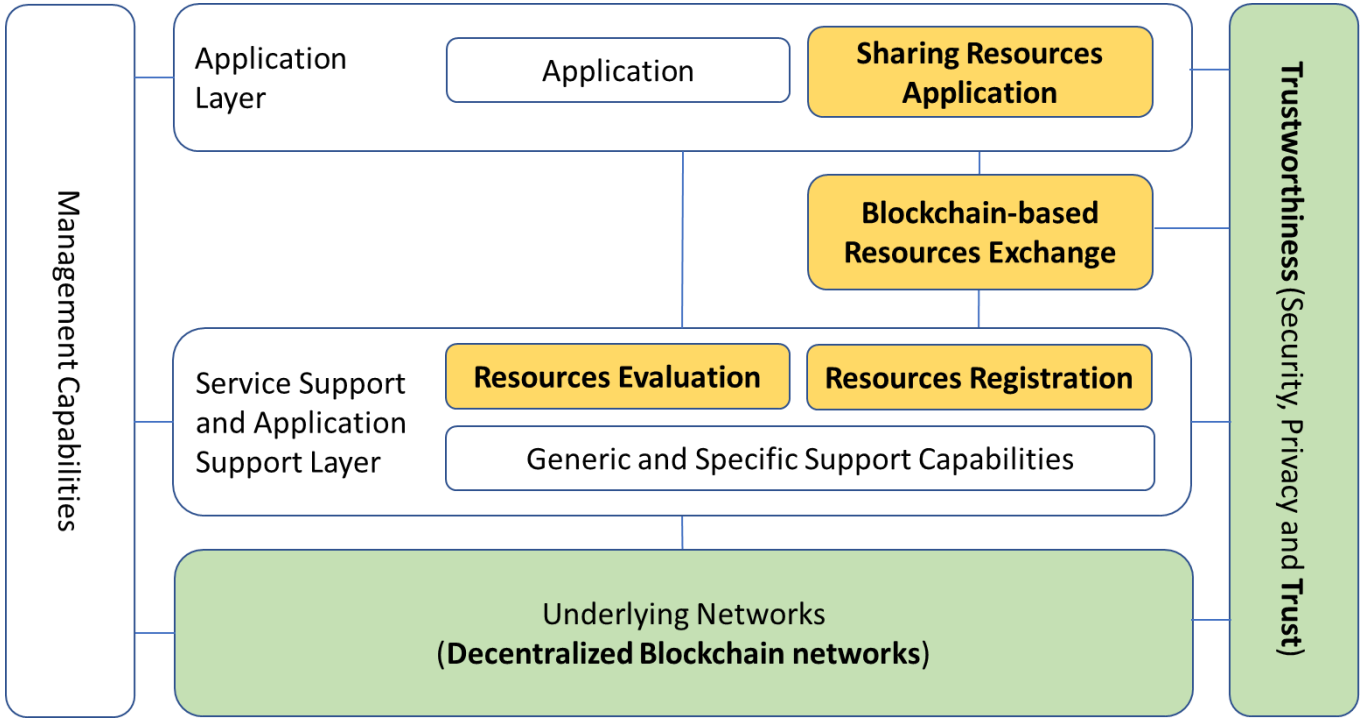
Y.eHealth-Semantic - Architecture of web of objects based semantic mediation model in eHealth service



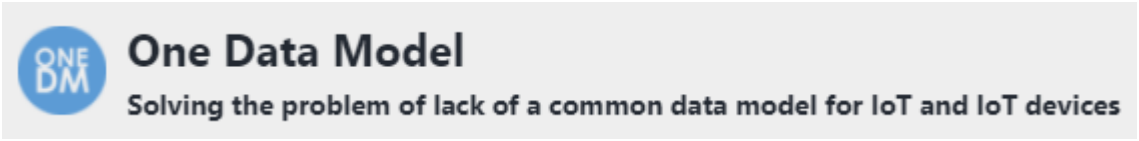
Blockchain-based resources management

- Blockchain-based data sharing (ITU-T FG-DPM, SG20)
  - **Y.Suppl.62 (07/2020)** - *Overview of blockchain for supporting Internet of things and smart cities and communities in data processing and management aspects*
  - **Y.4560 (08/2020)** - *Blockchain-based data exchange and sharing for supporting Internet of things and smart cities and communities*
  - **Y.4561 (08/2020)** - *Blockchain-based data management for supporting Internet of things and smart cities and communities*

Trust-enabled blockchain-based sharing networks



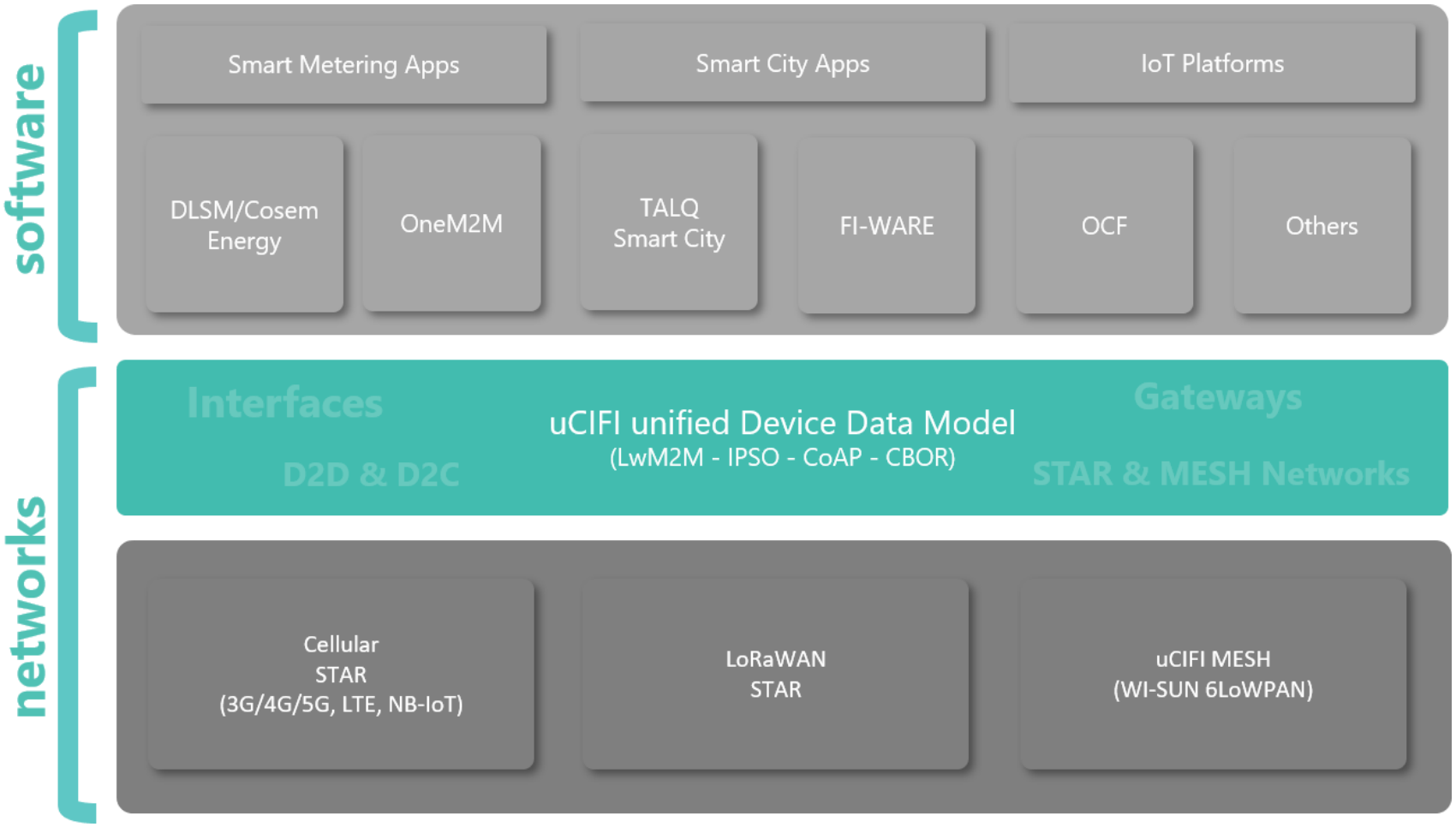
One Data Model



- Each IoT standards organization, and many IoT platform vendors, have created their own version of an IoT data model framework, each with a bespoke meta-model and representation language.
- To start alignment of existing data models, the models should be described in the same format. Hence, the Semantic Definition Format (SDF) has been created. SDF and related tools and frameworks can manage the contribution of data models from diverse sources, and collection of these data models in a single place.

https://onedm.org/

uCIFI Alliance



Smart data models

A GLOBAL PROGRAM LED BY

FIWARE IUDX OPEN & AGILE SMART CITIES tmforum

- The availability of widely adopted (de-facto standard) information models is key for creating a global digital single market of interoperable and replicable (portable) smart solutions in multiple domains (smart cities, smart agrifood, smart utilities, smart industry, …).
- All data models are **public and royalty-free** nature of specifications.
- General Principles
  - Driven-by-implementation approach
  - Open-closed (Breaking changes to already approved specs is not allowed)
  - Open contribution

Open Digital Framework & Open Digital Architecture

tmforum

Open Digital Framework

Transformation Tools

- 'As-Is' OSS/BSS Application Map (TAM)
- Transformation Toolkits

Maturity Tools

- Metrics
- Maturity Models

Data

- Benchmark Data
- AI Training Data

Open Digital Architecture

Governance

- Concepts & Principles
- Design Guides
- API Governance
- AI Governance
- Data Governance
- Security Governance

Business

- Business Capability Repository
- Process Framework (eTOM)

Information Systems

- Functional Framework & Architecture
- Information Framework (SID)
- Integration Framework

Agile lifecycle management

Deployment & Runtime

- Canvas
- Operation frameworks
- Reference Implementation

Implementation

- Technical Architecture & Components
- Open APIs
- TM Forum Data Model

Urban deluge of data

stakeholders, technologies and levels of legacy, spoken

Silos-based

stakeholders, technologies and levels of legacy, spoken/written

Cities are highly heterogenous

(internally and nationally/regionally): multidiscipline, departments

= there is a critical

- **need for a common language**
- **need for correlation**
- **need to better understand cities**

GOUI: The Objectives

I. **Create a NEW Smart Cities Ontology** as a common language

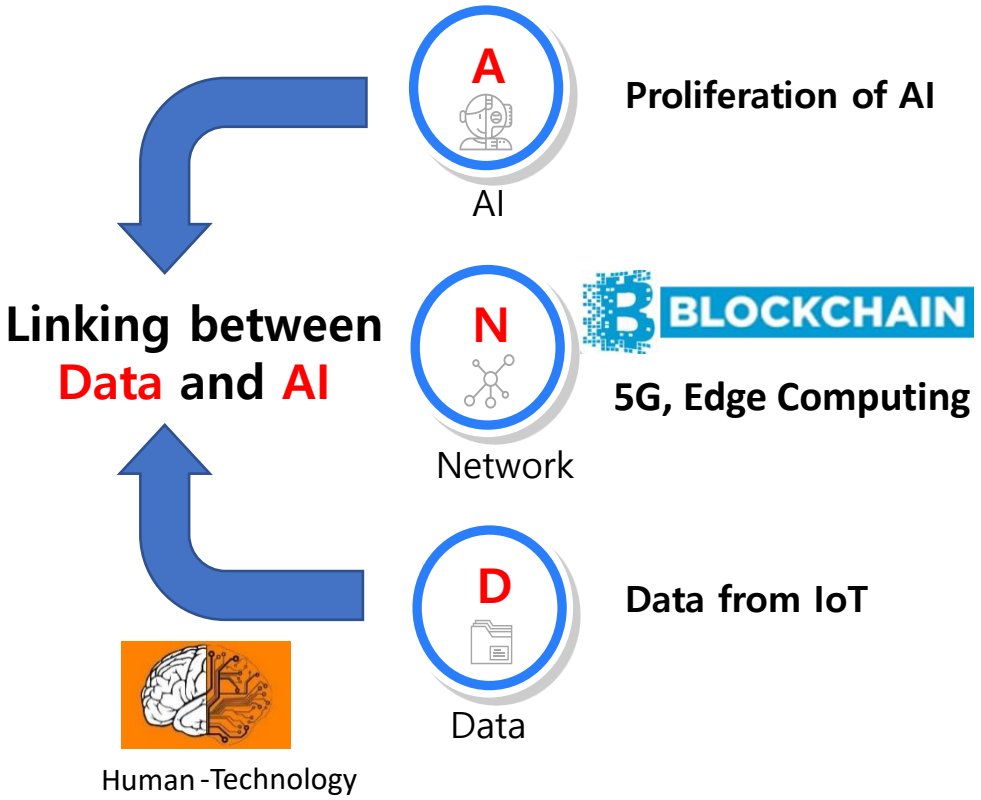
II. **Correlations** via semantics

III. **Digital Twins** - model cities to better understand them

IEEE ITU

Wrap up

Towards Digital Economy with Decentralization



Trustworhiness with Blockchain

- Transparency
- Data protection
- Privacy preserving
- Policy and regulatory issues
- Ethics

Concluding remark

- Towards Data-centric approach
  - **Trustworthy Decentralized Data Ecosystem with AI**

Standardization of Data-Driven ICT

- Common features, but unlimited number of solutions

- Fregmentization
- Assembling


Q
&
A