

Unlocking the Power of Data Spaces through Legal, Ethical, Social and Environmental Compliant Data and AI Pipelines

Prof. Dessislava Petrova-Antonova
GATE Institute, Sofia University

15 January 2025, Sofia

Big Data for Smart Society (GATE) Institute

From 2019, Develop Big Data and AI applied research, innovation and education in collaboration with government, industry and entrepreneurs

СОФИЙСКИ УНИВЕРСИТЕТ
„СВ. КЛИМЕНТ ОХРИДСКИ“



CHALMERS
UNIVERSITY OF TECHNOLOGY

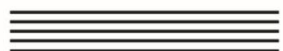


CHALMERS
INDUSTRITEKNIK

GATE


GATE IDSA Hub

INTERNATIONAL DATA SPACES ASSOCIATION







Bulgaria: Big Data for Smart Society Institute (GATE)

The IDSA Hub in Bulgaria is facilitated by the Big Data for Smart Society Institute (GATE) in Sofia. GATE is the first Centre of Excellence in Bulgaria to work on scientific excellence and innovation in big data and artificial intelligence. It was established in 2019 as an autonomous structure of Sofia University "St. Kliment Ohridski". The team around Director Prof. Sylvia Ilieva is focusing on knowledge transfer on data sovereignty and IDS components into the Bulgarian industry and research community.



**Data Spaces
Discovery Day –
Business value of
sovereign data sharing**

Sofia | November 9, 2022

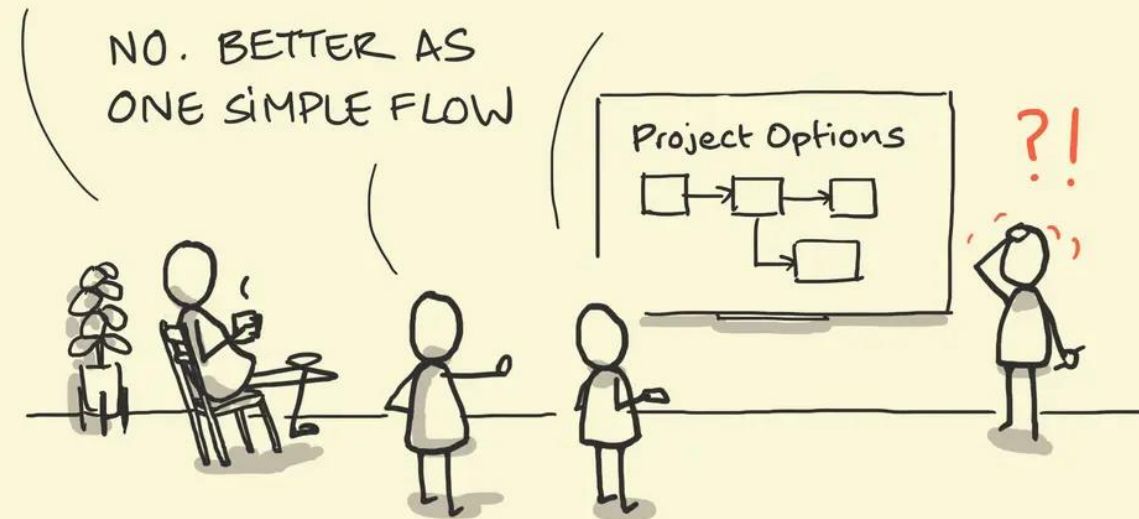


Data Value

Data has only a value when it becomes information, and information becomes knowledge giving us wisdom to make decisions!

JUST SHOW ALL THE
OPTIONS UPFRONT

IT'S BORING!
IT NEEDS PIZZAZZ



**WITHOUT DATA YOU'RE JUST
ANOTHER PERSON WITH AN OPINION**

— W. EDWARDS DEMING

European Data Spaces



Unique Selling Proposition

Ecological

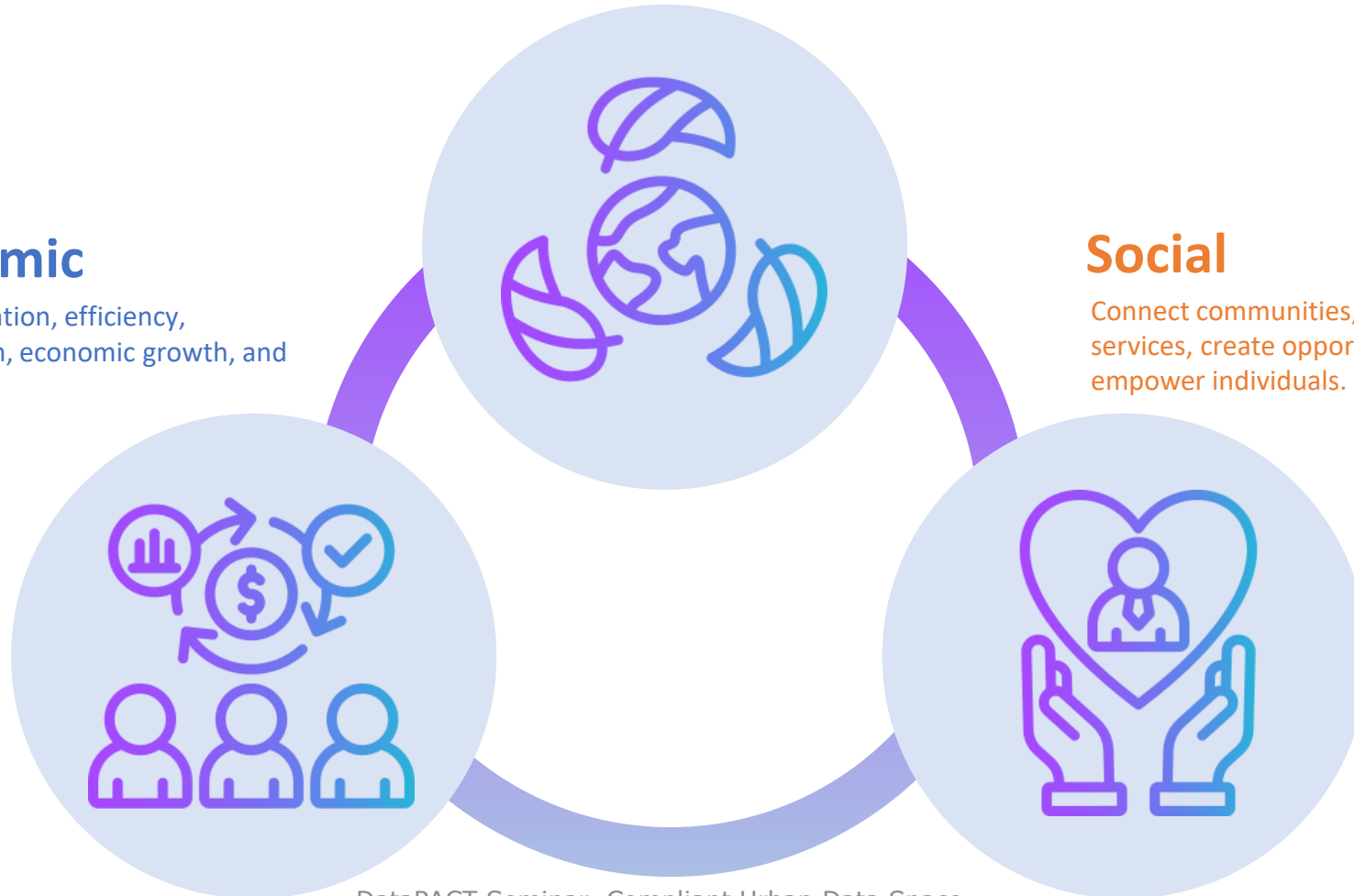
Enable resource optimization, reduced waste, sustainability, and environmental insights.

Economic

Boost innovation, efficiency, collaboration, economic growth, and value.

Social

Connect communities, improve services, create opportunities, and empower individuals.



Synthesis of Building Blocks



INTEROPERABILITY

Data Models and Formats

Data Exchange APIs

Provenance and traceability



TRUST

Identity Management

Access & usage control policies

Trusted Exchange



DATA VALUE

Metadata & Discovery Protocol

Data Usage Accounting

Publication & Marketplace Services



GOVERNANCE

Overarching cooperation agreement

Operational (e.g. SLA)

Continuity model

Regulatory Framework

- Digital Markets Act
 - EU's law to make the markets in the digital sector fairer and more contestable.
- Digital Services Act (DSA)
 - Regulates online intermediaries and platforms such as marketplaces, social networks, content-sharing platforms, app stores, and online travel and accommodation platforms.
- AI Act
 - Regulates the use of AI
- Data Governance Act
 - Regulates the reuse of publicly/held protected data by promoting data sharing through the regulation of novel data intermediaries and by encouraging data sharing.
- Data Act
 - Gives both individuals and businesses more control over their data through reinforced portability rights, copying or transferring data easily from across different services.

The European Strategy for data at a glance


The **European Strategy for data** (2020) aims to make the EU a leader in a data-driven society.

The **Data Governance Act** (2020) facilitates data sharing across sectors and Member States.


The **Data Act** (2022) clarifies who can create value from data.

Ten **European common data spaces**, ranging from industry to mobility, from European Green Deal to energy and health.


◆ **Cheaper prices** for aftermarket services and reparation of their **connected objects**.
A factory robot breaks down.

TODAY	TOMORROW
 <p>Only the manufacturer can access the data, leaving no alternative for the company but to call them for repairing.</p>	<p>The user could request that a repair service that may be cheaper also gets access to the data.</p>

◆ **New opportunities** to use services relying on access to this data.
A farmer has equipment from different manufacturers (tractor, automatic irrigation system).

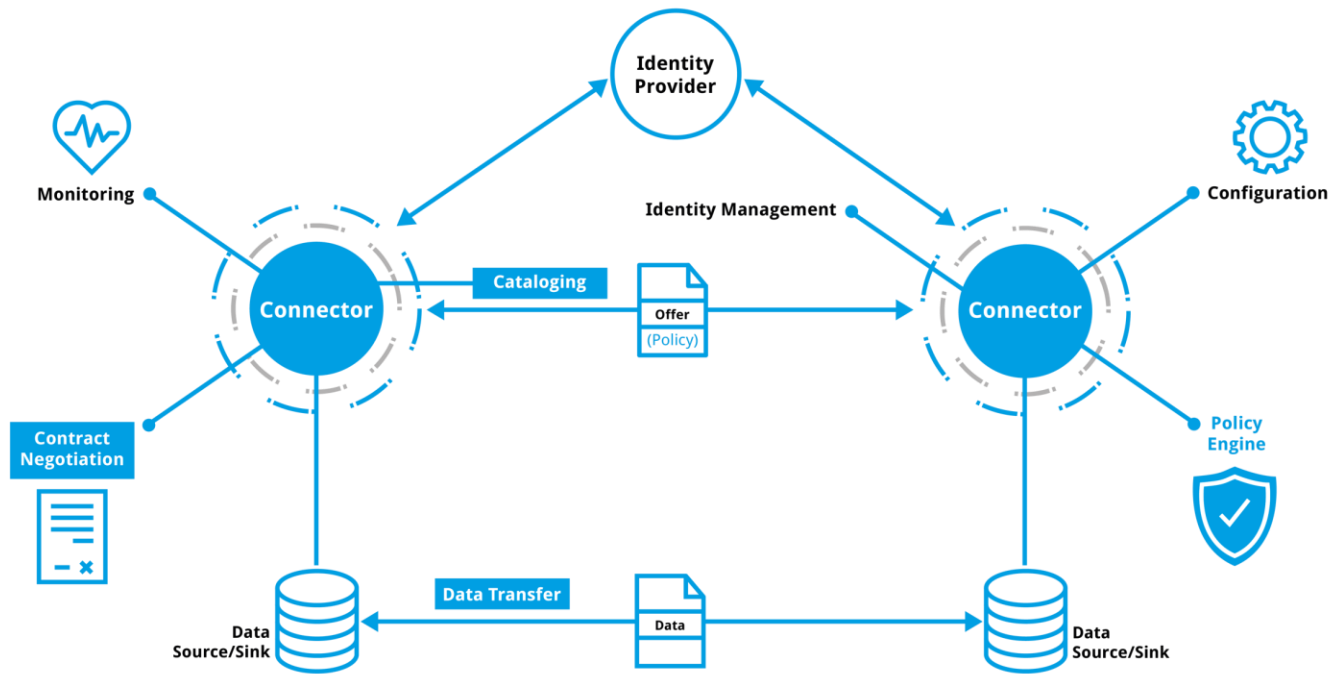
 <p>He cannot outsource the data analytics of its different equipment, the data is locked with each manufacturer.</p>	<p>He could receive customised advices from a company gathering data from the different equipment.</p>
---	--

◆ **Better access** to data collected or produced by a device.
A bar owner wants to serve better coffee, and the coffeemaker company wants to improve its product.

 <p>Only the company can access the data produced by the machine to design the next generation of coffeemakers but the bar owner cannot access information such as the quantity and temperature of water or coffee strength.</p>	<p>The Data Act clarifies that both parties can access all data collected by the machine.</p>
---	---

Source: <https://www.eu-data-act.com/>

Data Space Protocol



Source: <https://docs.internationaldataspaces.org/ids-knowledgebase/dataspace-protocol>



How **Datasets are deployed** as DCAT Catalogs and usage control is expressed as Open Digital Rights Language (ODRL) Policies.



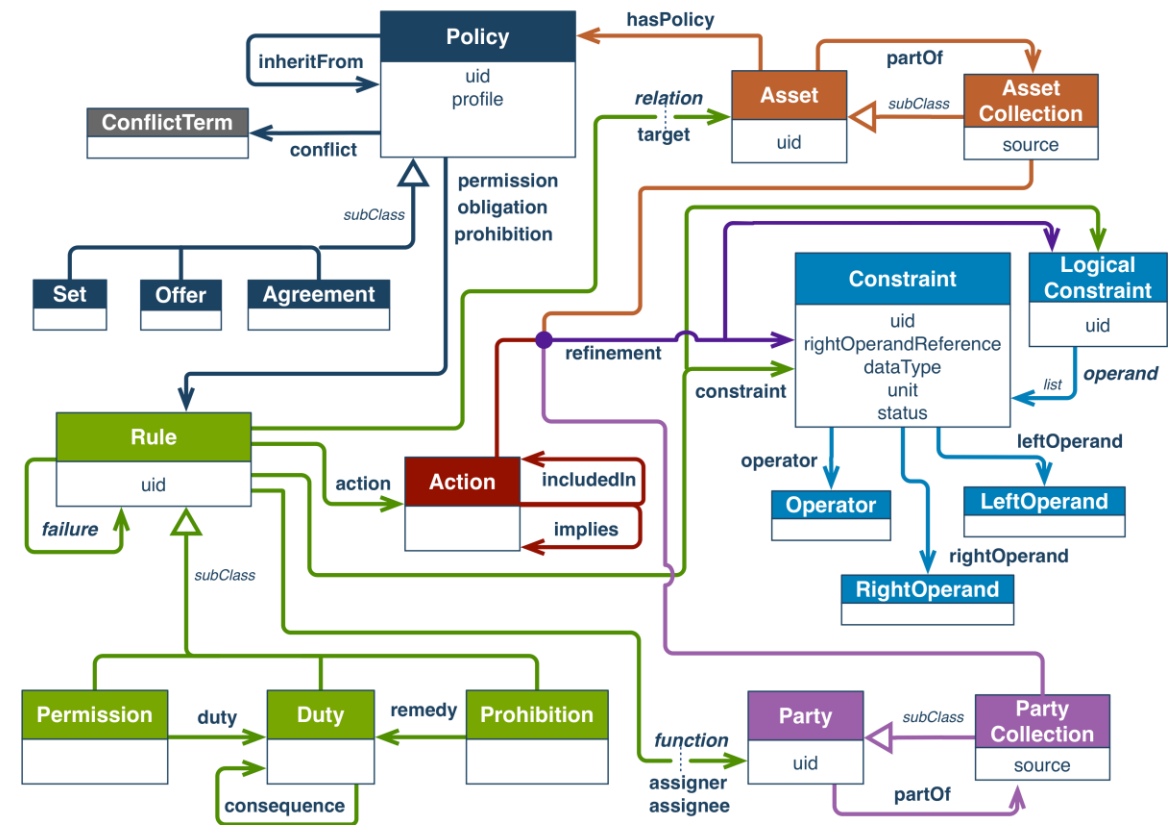
How **Agreements that govern data usage** are syntactically expressed and electronically negotiated.



How **Datasets are accessed** using Transfer Process Protocols.

Open Digital Rights Language Information Model

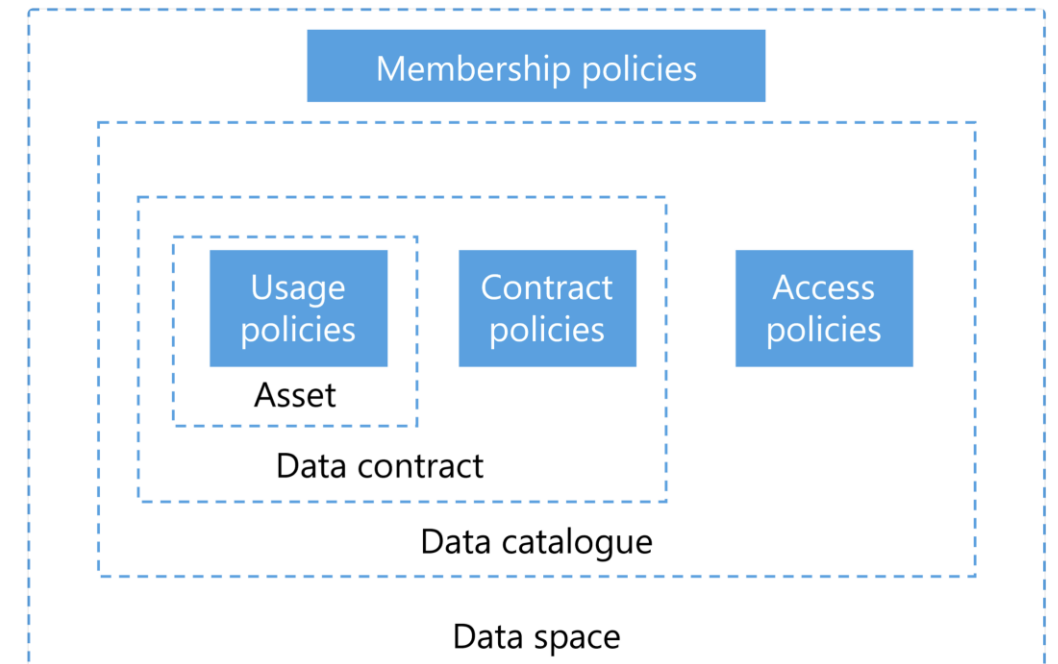
- Represents Policies that express Permissions, Prohibitions and Duties related to Asset usage
 - **Asset**
 - A resource or a collection of resources, subject to a Rule
 - **Policy**
 - A non-empty group of Permissions and/or Prohibitions and/or Duties
 - **Party**
 - An entity or a collection of entities that undertake Roles in a Rule
 - **Action**
 - An operation on an Asset
 - **Rule**
 - An abstract concept representing the common characteristics of Permissions, Prohibitions, and Duties
 - **Constraint/LogicalConstraint**
 - A boolean/logical expression that refines an Action and Party/Asset collection or the conditions applicable to a Rule



Source: <https://www.w3.org/TR/odrl-model/>

Different Policies in Data Spaces

- Policies
 - Expressing three possible restrictions, e.g. prohibitions, obligations, and permissions
 - Mandatory policies depend on the data space design and its requirements
 - Participants may always choose additional policies in their data contracts to further restrict access and use
- Constraints
 - Express a rule that can be combined into more complex rules, forming the applicable policy
- Membership policies
 - Required to join a data space
- Access policies
 - Control access to data contracts
- Contract policies
 - Control the contract terms and the usage of data
- Usage policies
 - Control how the receiving party can use the data



Source: https://docs.internationaldataspaces.org/ids-knowledgebase/idsa-rulebook/idsa-rulebook/3_functional_requirements

Classification and Regulatory Controls of Data: An Example

Explanation	Protection Need	Explanation
Public weather data	low	Some datasets are already publicly available and can be shared without enabling sensitive data about persons or business secrets.
Shipping information	medium	Some data is valuable at a large scale and is likely to be highly protection-worthy as it can give insights into business relations and transactions.
Personal health data	high	Personal health data is highly protection-worthy due to strong laws and potential danger to the individual in case of data misuse.
Machine operations data	high	Industrial data is also usually of high value due to the sensitive business information it represents.



DATA PACT



GATE Urban Data Space

Current status

An integrated data space environment comprising ALL components of the IDS Reference Architecture Model

- Based on current IDS reference implementations
- Certification of GATE Data Space Connector (Trust Level 1)

Initial scenario for the air pollution prediction

1. A data consumer collects data about a particular pollutant from two providers via an API and a database
2. Air pollution prediction models are offered as a service in the App store
3. The models are downloaded and executed in the data consumer connector

The certification body of the IDSA hereby certifies that the:

INTERNATIONAL DATA SPACES ASSOCIATION

GATE Dataspace Connector

of

GATE Institute, Sofia University

meets the requirements

Trust Level 1, Assurance Level 1 IDS Certification for connectors

The evaluation result and certification requirements are referenced in the annex to the certificate.

This certificate with the registration number IDS-CON-0001-TL1-AL1-2024 is valid until 01.08.2026.

Dortmund, 01.08.2024



Lars Nagel, CEO
International Data Spaces Association



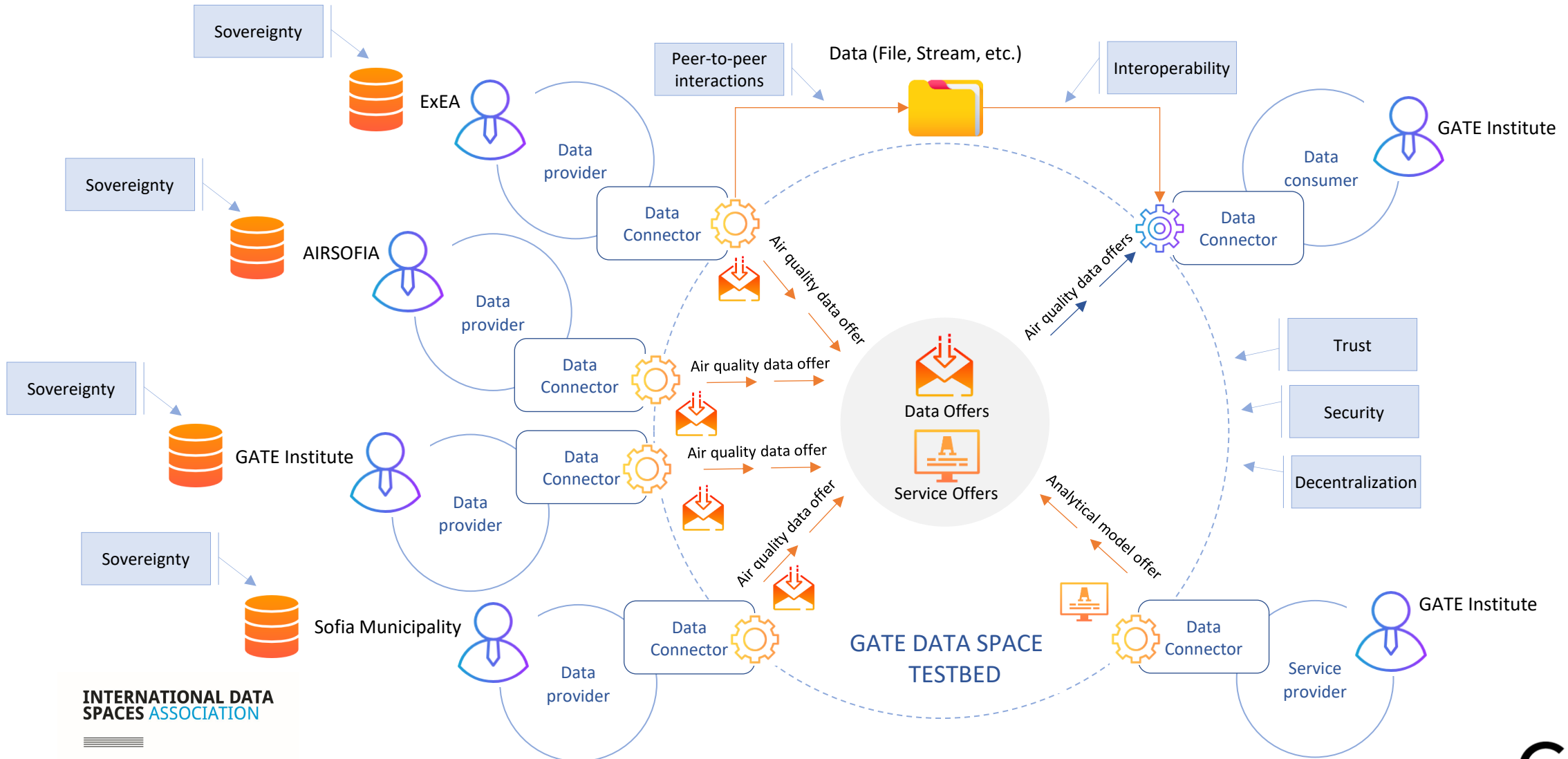
Thorsten Huelsmann, CFO
International Data Spaces Association

Certification Body
certification@internationaldataspaces.org



Certificate

GATE Data Space Testbed



Air Quality Data Modelling and Analytics

Compliant Data Modelling

- OGC CityGML standard
- CityGML ADEs
- FIWARE Smart Data Models
- MIMs

01

03

Temporal Air Quality Prediction

- ARIMA
- LSTM
- Transformers
- CNN-Linear-RNN
- LSTNet (CNN + LSTM)
- Ceemdan-LSTM
- Spatial attention + temporal attention bidirectional LSTM (SATABL)

Spatial Air Quality Prediction

- Gaussian auto-regressive model

02

04

Interactive 3D City Model

- LLAMA 3.1 (70B), QLoRA fine-tuned
- 3D City Model, CityGML

The 'Talking' 3D City Model Demo

CHATBOT @ GATE-AI

chat.gate-ai.eu/citychat/new

Zimbra: Inbox My GATE Mail GATE

GATEBOT

What is the weat...
What is the reas...
Replay the air po...

Hi there, I am Mapster, your friendly assistant at the GATE Institute. What is your next map-tastic job for me?

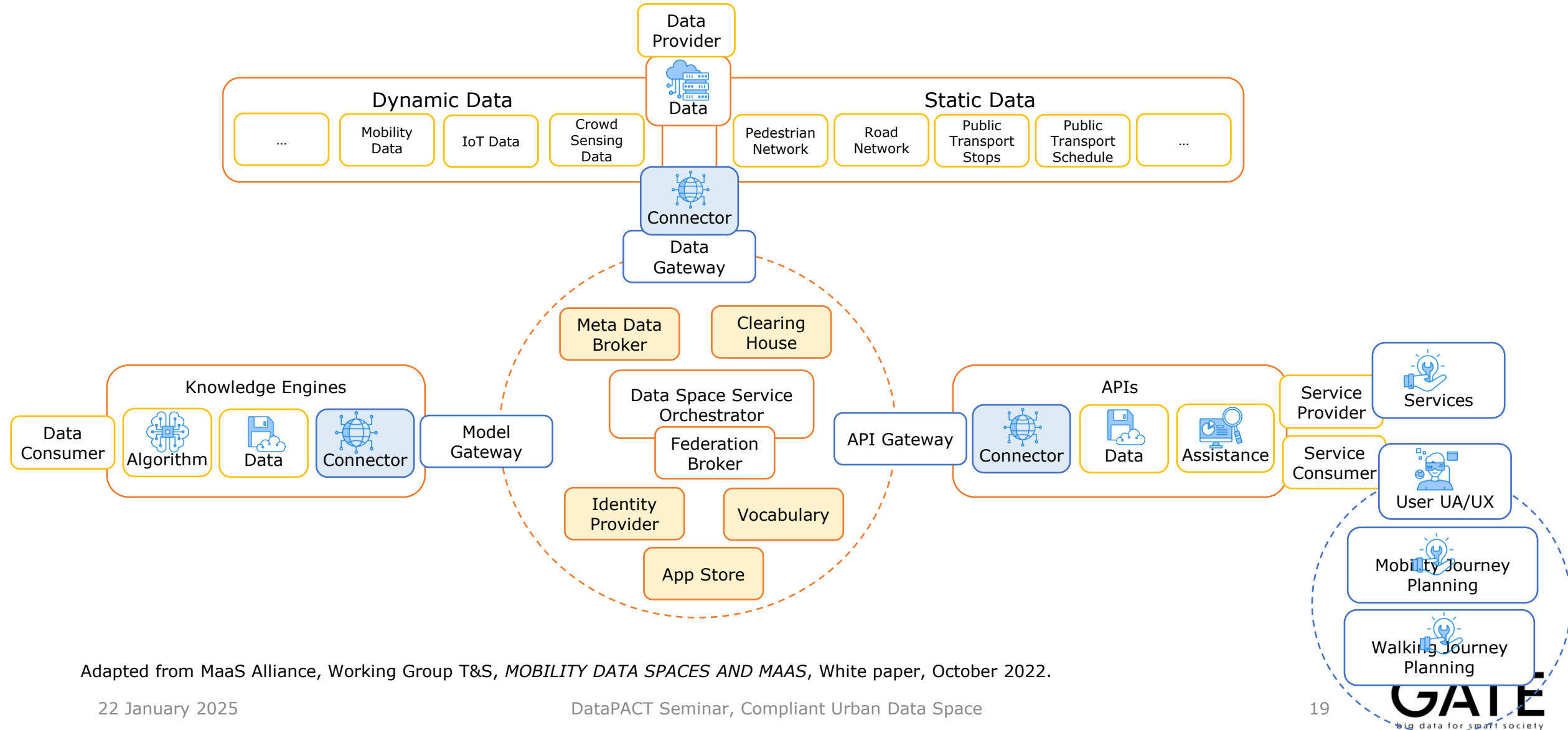
1x
Jan 13 2025
18:34:10 UTC

CESIUM ion Upgrade for commercial use - Google Data attribution

Jan 13 2025 20:00:00 UTC Jan 14 2025 00:00:00 UTC Jan 14 2025 04:00:00 UTC Jan 14 2025 08:00:00 UTC Jan 14 2025 12:00:00 UTC Jan 14 2025 16:00:00 UTC

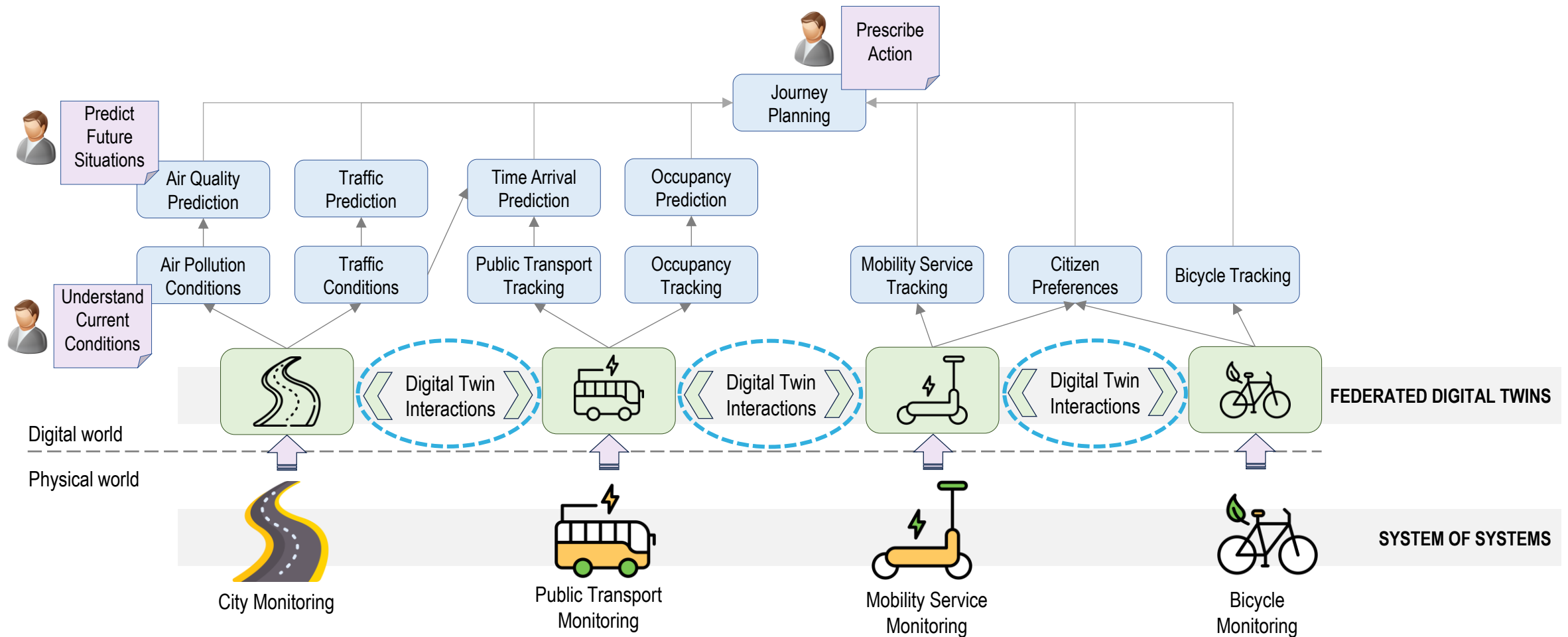
Type here or mash the mic.

Topology Overview of Mobility Data Space



Adapted from MaaS Alliance, Working Group T&S, *MOBILITY DATA SPACES AND MAAS*, White paper, October 2022.

GreenMob Use Case of Sofia



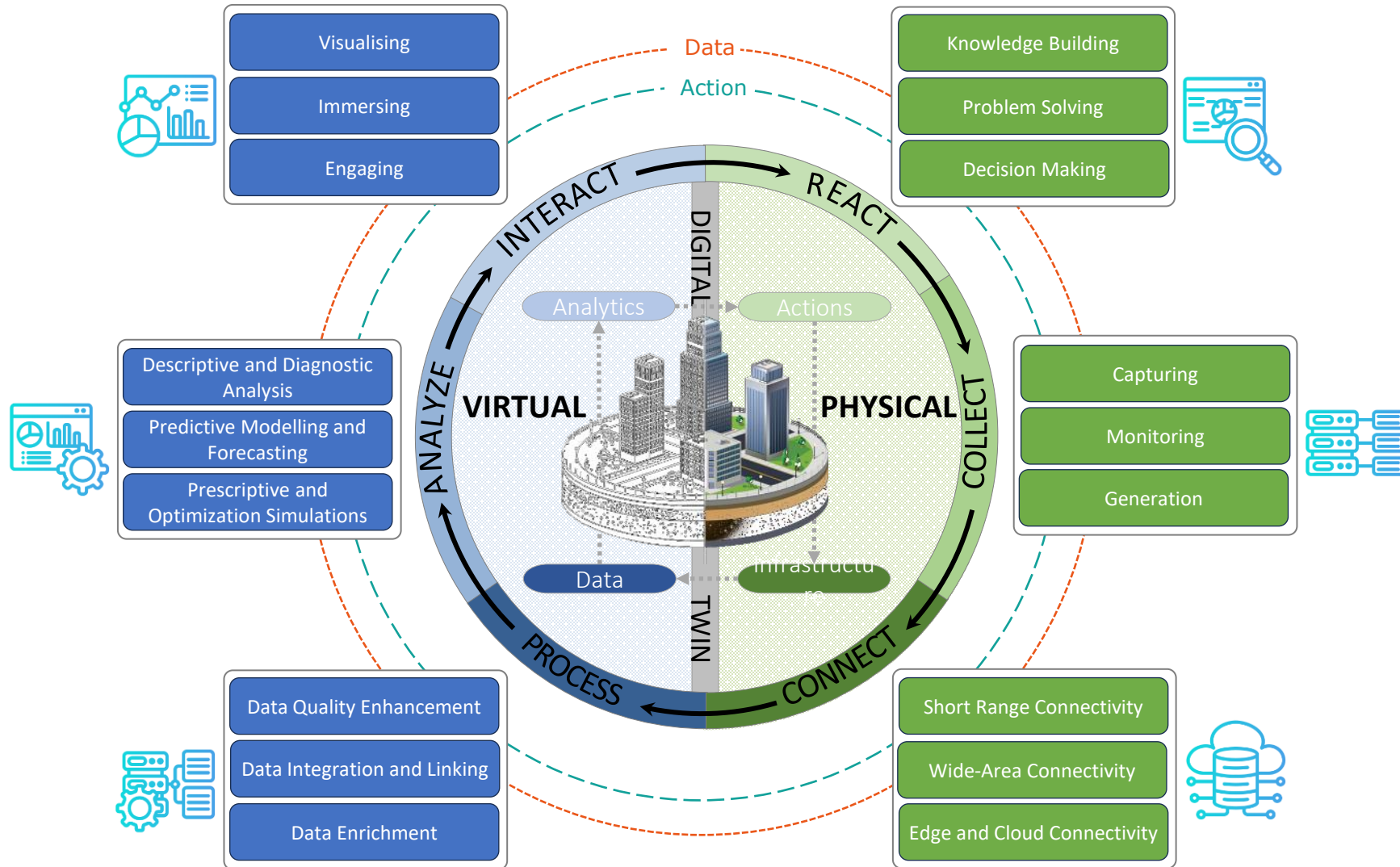
DataPACT Contribution to the Data Spaces

- Data providers
 - Checking compliance with legal, ethical, societal and environmental sustainability frameworks and policies for data to facilitate **easy, fair and unbiased data sharing** following existing and emerging regulations and guidelines.
- Data consumers
 - Checking compliance with legal, ethical, societal and environmental sustainability frameworks and policies for AI pipelines to secure the development of **AI systems**, especially those **with a high risk**.
- Actual data transferred between the systems
 - Checking compliance with the **structure, syntax or semantics of the data**, which is to the agreements of the Participants or the Dataspace.
- Compliance assessment of connectors, shortening the time for their certification
 - The IDS Connector defined in DIN SPEC 27070 is part of the German **standardization work, subject to international standardization** in ISO/IEC, CEN/CENELEC, IEEE, and W3C.
 - The IDS Certification launched in 2022 proves compliance with these requirements, offering different trust and assurance levels for connectors and operational environments.
- Connector additional functionalities
 - Implement additional internal compliance functionalities like **monitoring or policy engines**, as appropriate.

DataPACT Benefits

- Data Space
 - Implementing the technical, operational, and legal aspects of Urban Data Space
- PolicyR
 - Specify, negotiate and enforce policies for participants' data use in the data space.
- AssessR
 - Evaluate the shared data compliance with the legal, ethical, societal, and environmental sustainability regulations and frameworks.
- PipelineR and ApplyR
 - Create compliant data pipelines and develop compliant AI pipelines based on shared data.
- GreenR
 - Track the carbon footprint of data operations and transactions.

Data Space Enabled Local Digital Twins





Looking forward to the future!

Q&A

