Beyond data sharing: Realizing Value through Applications in Data Spaces

12/03/2024 14:00–15:30 (CET)

Daniel Alonso, Nuria De Lama, Thomas Hahn, Arturo Medela, Emmanouil G. Spanakis, Ed Curry, Tanguy Coenen, Hanno Focken, Rob Smeets, Aitor Corchero, Denia Kanellopoulou, Mihnea Tufis

DSBA ■ BDV BIG DATA VALUE ASSOCIATION FIWARE FOUNDATION gaia-x INTERNATIONAL DATA SPACES ASSOCIATION

DATA SPACES SUPPORT CENTRE

data Spaces Symposium
Unite. Innovate. Adopt.
Darmstadtium | Frankfurt region

Objectives of the session

- To discuss about different current and potential ways to generate value from data spaces (with data sharing and collaboration among stakeholders as main drivers)

- To identify technical enablers needed to materialize this value generation

- To hear from industry different sectors real use cases of value generation from data and data sharing, and how data spaces are contributing to them.

- To focus on AI / industrial AI as a paradigmatic example that can benefit from its connection with data spaces

Agenda of the session
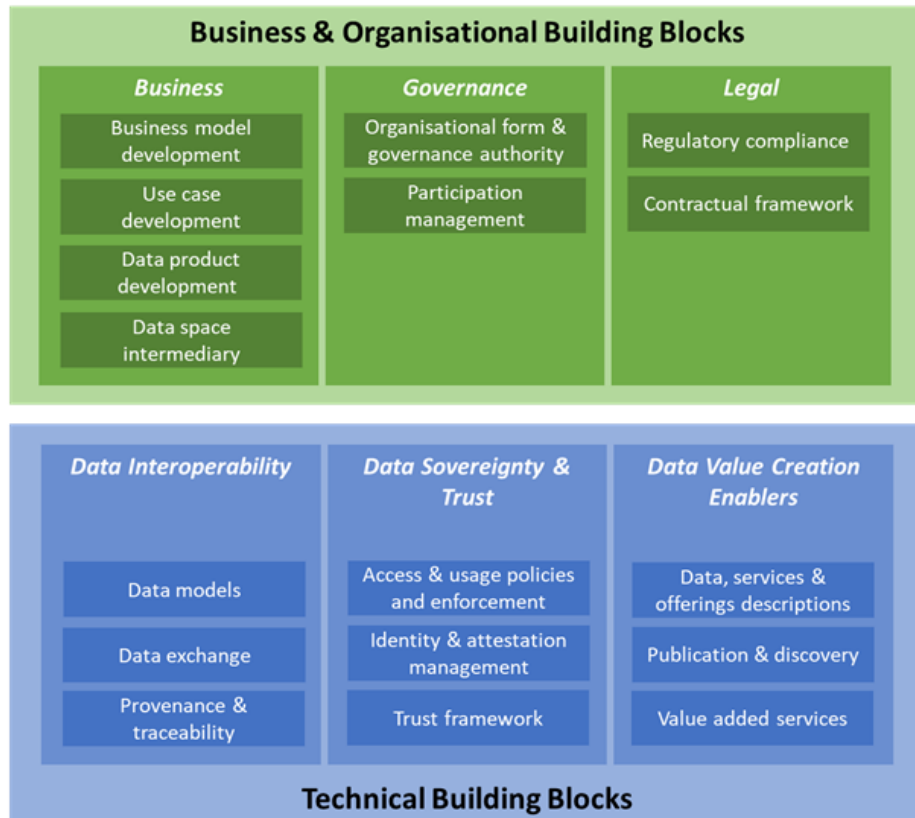
Speaker	
Introduction and setting-up the scene	
Daniel Alonso (BDVA), Nuria De Lama (IDC)	Intro and framework
Thomas Hahan (Siemens, BDVA)	Motivational picth
Part 1 – View from research and innovation (initial statement + panel discussion) -> moderator Daniel	
Arturo Medela	SEDIMARK Project
Emmanouil G. Spanakis	Trustee Project
Ed Curry	Generative AI and data spaces
Tanguy Coenen	Algorithm Centric Automated Data Integration
Minhea Tufis	PISTIS project
Part 2 – View from industry (initial statement + panel discussion) -> moderator Nuria	
Hanno Focken	Manufacturing sector
Rob Smeets	Healthcare domain
Aitor Corchero	Public administration
Denia Kanellopoulou	Industrial AI

BDVA data value generation framework

BDV

Business
Requirements, business models, etc ...

Data driven value generation

For specific sectors, based on business needs and models, using applications that rely on the use of data and / or data sharing through data spaces

Use cases

Conformity assessment

Sectors
Healhcare, mobility, manufacturing, agrifood, energy, finance, media, telecom, SSCC, security, earth observation

Applications

| AI | Digital twin | Marketplaces | Virtual worlds | Supply chain optimization | Energy-efficiency management | ... |

Data space

Data protection

Data quality

| Data analytics | Data processing | Data visualization |
| Data generation | Data acquisition | Data cleaning |

| Data Exchange | Description, publication and discovery |
| Data models and formats | Access and usage control |

AI tools

Sensing and perception	Action and interaction
Knowledge and learning	Reasoning & decision making
LLM	NLP

Data Spaces Support Center

Business & Organisational Building Blocks

Business	Governance	Legal
Business model development	Organisational form & governance authority	Regulatory compliance
Use case development	Participation management	Contractual framework
Data product development		
Data space intermediary		

Data Interoperability	Data Sovereignty & Trust	Data Value Creation Enablers
Data models	Access & usage policies and enforcement	Data, services & offerings descriptions
Data exchange	Identity & attestation management	Publication & discovery
Provenance & traceability	Trust framework	Value added services

Technical Building Blocks

Some examples of value-added services:

- Data visualization
- Data fusion
- Data enrichment
- Data and metadata quality assurance
- Anonymization / pseudo-anonymization
- Collaborative data analytics
- ML models hosting (models as a service)
- Federated learning
- Collaboration and knowledge sharing
- Training and education
- Data innovation labs

Some potential tecnical components identified:

- Services orchestrator
- Services validation
- Different interfaces:
- Monitoring
- Testing and quality assurance
- Load balancing
- Scalability
- Training and support

Data Spaces Symposium
Unite. Innovate. Adopt.

Source Paul Clarke - Eigenes Werk, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=37435469

Data Spaces Symposium
Unite. Innovate. Adopt.

March 1989

# Why Data Transparency is needed?

Example: Carbon Footprint
- ~10% of CO2 emissions from industry generated by own factories, ~90% by upstream/downstream supply chain [1]
- To reduce CO2 emissions, transparency along the entire value chain is necessary

1) depending on production

Example: Battery production
- Only if we use the potential of digitalization we can produce batteries more sustainably
- ~ 96% of the ingredients of a battery are recyclable
- Consistent and transparent use of digital twins along the complete cycle the design and production will be more sustainable
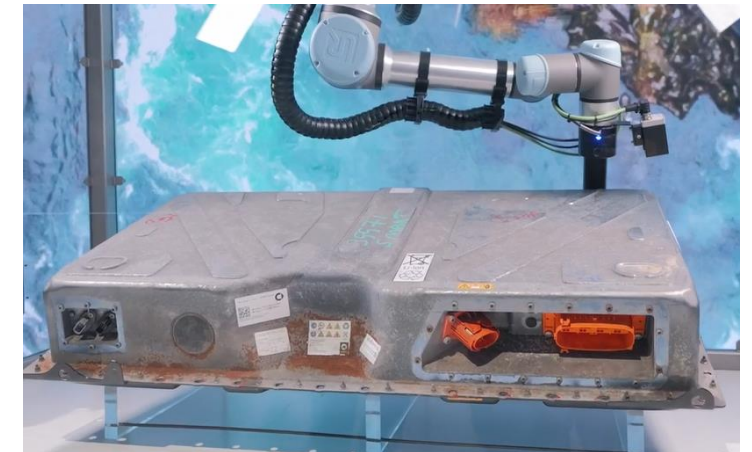
Manufacturing-X: Make Data Work

Goal

Manufacturing-X is a German **initiative** to digitalize the entire manufacturing and supply chains in industry. It aims at implementing the data space for Industrie 4.0 across industries on a **global scale**. The goal is to enable digital innovations for greater resilience, sustainability and competitiveness.

Approach

Companies jointly work together on use cases and shared services to enable data exchange across companies and products of different vendors. Manufacturing-X comprises Catena-X for the automotive industry and a proposed project proposal called Factory-X for Machine Builders and factory operators. It builds on common standards such as the Asset Administration Shell.

Customer value

- **Fast adaptation** of existing **standards** and **technologies**
- Creates **customer value** for concrete use cases
- Fast adaptation of **innovative** ideas and **business** models
- Opportunities for **SMEs** to gain access to technology and infrastructure as basis to provide their digital offerings
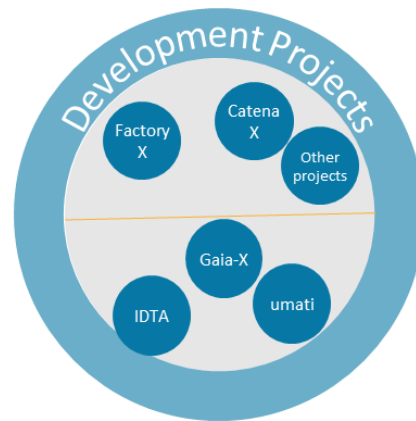
How industrial Data Ecosystems WORK?!

INFLUENCE
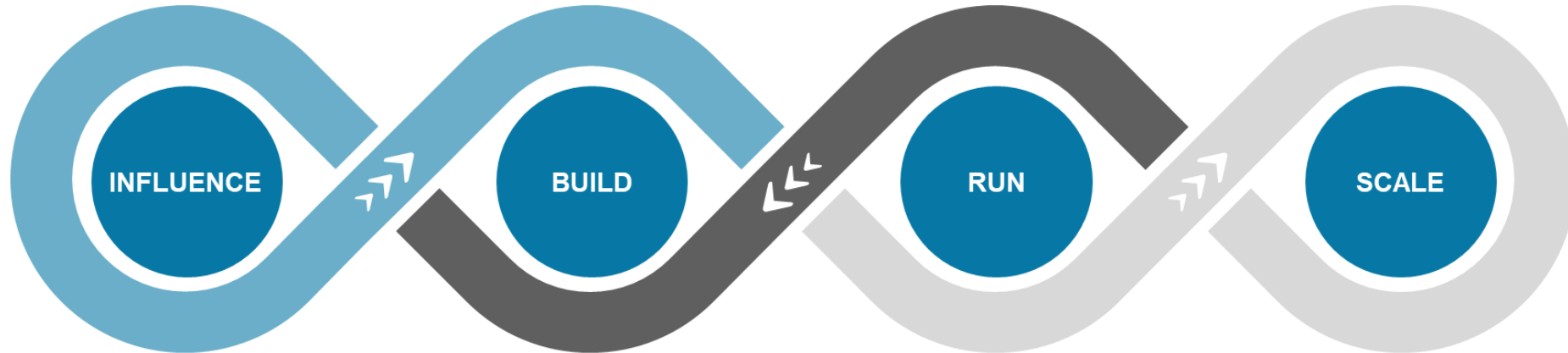
Rules & Governance
Overall Concept
Principles, Guidelines, Policies

BUILD

X-Association compatible Spec.
Development & Integration
Software | Services | Data Models

RUN

Secure and trusted
Operation of enabling services
Distributed Operating Model

SCALE

SME support activities
Consulting and training
Test beds and test support
Roll out of use cases
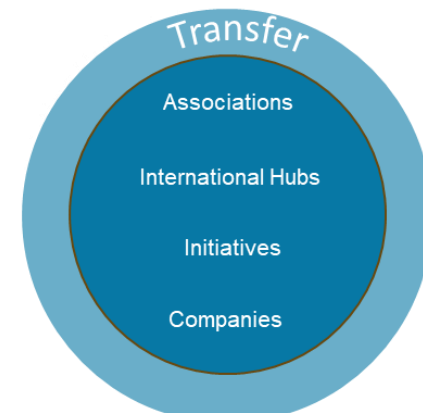
Existing Associations, Platforms, Communities & Standardization bodies

iMX

Development Projects

Factory X
Catena X
Other projects
Gaia-X
IDTA
umati

Certified Operating

Operating companies
Certifcation bodes
Conformity assessment bodies
Service providers
Market places

Transfer

Associations
International Hubs
Initiatives
Companies

Data Spaces Symposium
Unite. Innovate. Adopt.

What is SEDIMARK?

- SEDIMARK aims at designing and prototyping a **secure decentralised and intelligent** data and services **marketplace** that bridges remote data platforms and allows the efficient and **privacy-preserving sharing** of vast amounts of heterogeneous, high quality, certified data and services supporting the common **EU data spaces.**
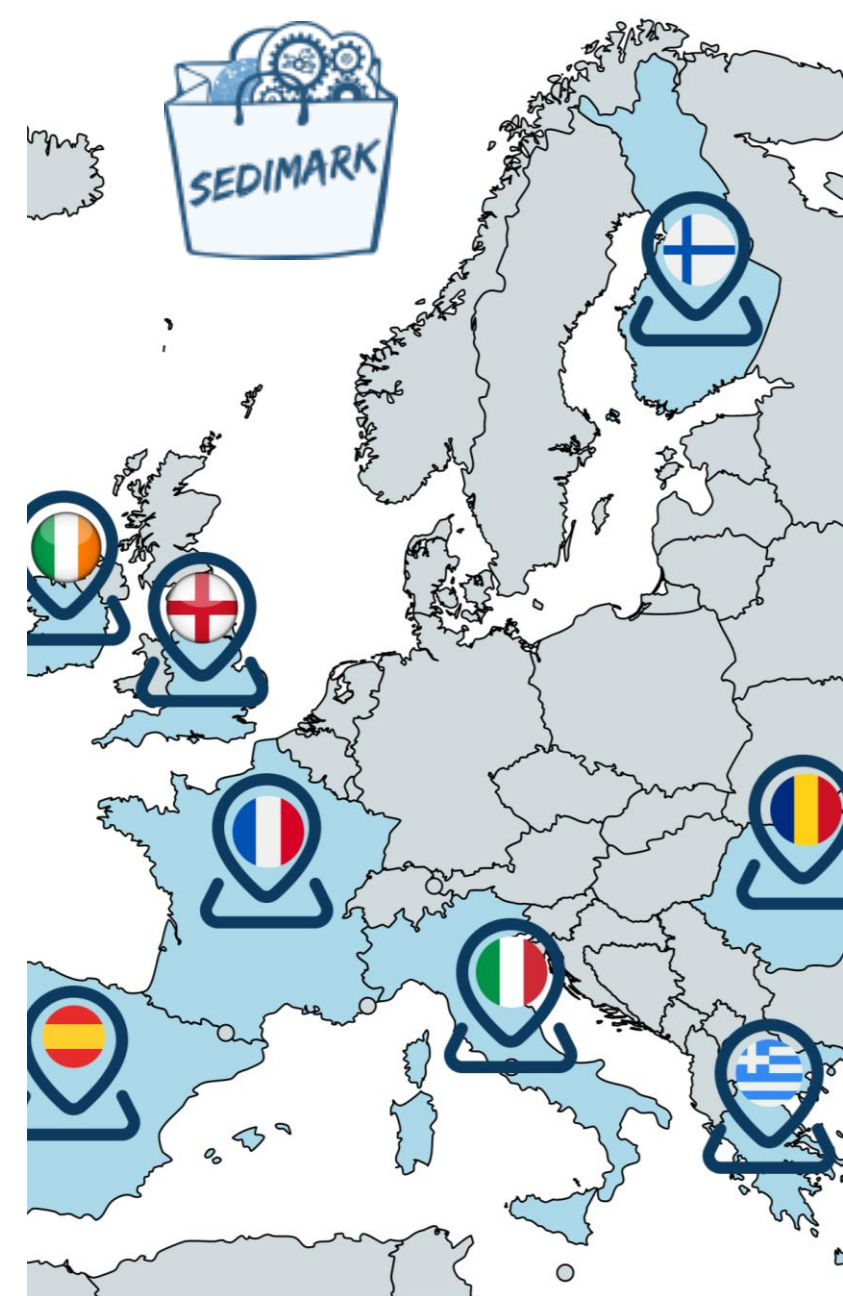

Mobility Digital Twin

Urban bike mobility planning

Valorisation of energy consumption data
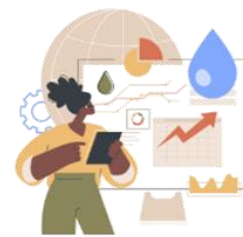
Valorisation of water-related data

Data Spaces Symposium
Unite. Innovate. Adopt.

SEDIMARK and data sharing

Key concept 🔑

- Data lifecycle control and data curation pipeline

- AI-based enrichment on top of semantically annotated data

- Greener solution as context information does not leave its own context

- Recommendation and data monetization schemes

- On-demand (distributed) data processing/model training

Going into the real world 🌐

- Adopt standards and vocabularies prominent in the **IoT domain** and emerging **data spaces** → allow **inter-disciplinary interoperability**

- **Optimise AI models** during training and inference → support providers and consumers

- Semantic enrichment, Data quality annotation, Model annotation, ML-oriented data quality annotation, AI model quality annotation, AI model validation, data validation/certification → **interoperability enabler.**

Goals/Ambition

TRUSTEE, in accordance with the principles of **responsible/trustworthy AI**, **proposes a novel secure-by-design, homomorphically enabled, federated platform in accordance with EU data strategy** and **EU reference architectures** capable of ensuring user-friendly, safe, trustworthy, compliant, fair, transparent, accountable, long-term data collection, storage, processing, querying, and delivery.
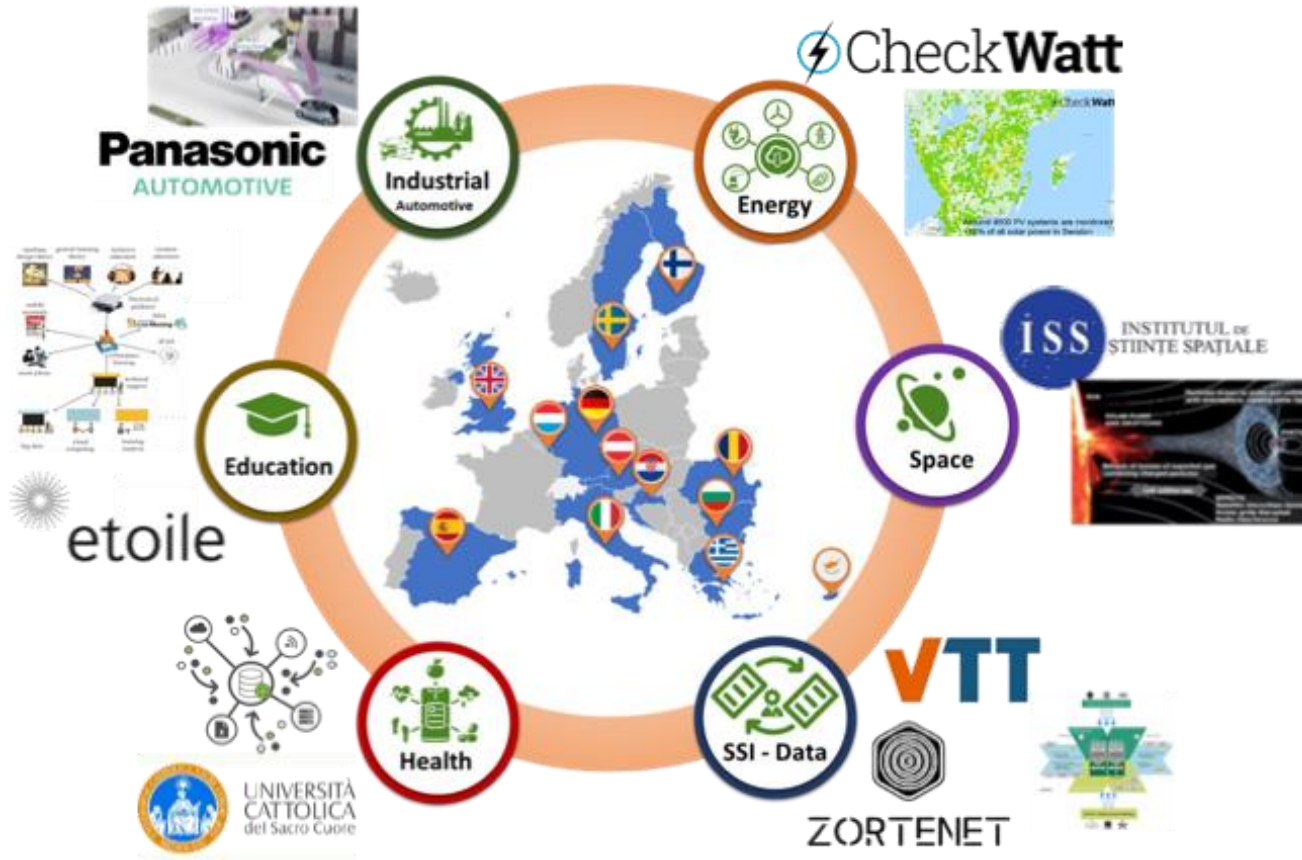
Pillars of innovation:
Collaboration | Trusted Data Solutions | Automate Testing and Monitoring | Data driven (AI enabled) services | Security and data sovereignty | Value adding apps | Interoperability | Ecosystem Harmonisation | Policies & legally compliant
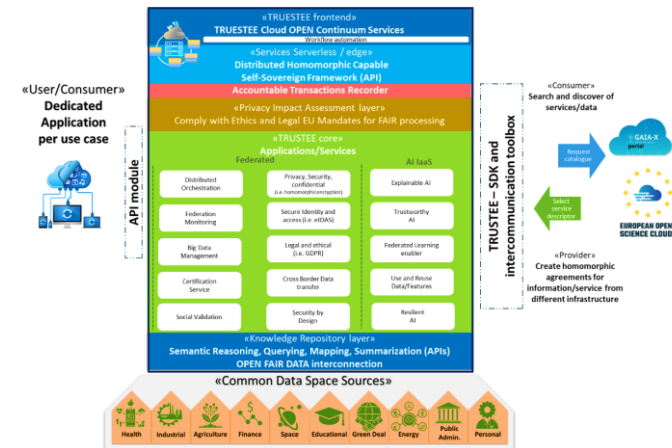
Application space – uses cases

TRUSTEE's
fully encrypted solution
will be **validated
through six different
use cases** supporting
GAIA-X, EOSC, EGI, etc.
**demonstrating a multi-
disciplinary, Pan-
European federated
FAIR and private data
ecosystem.**
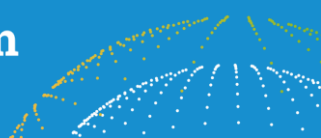
Value

1. Design a Secure and Trust framework and Reference Architecture to ensure end-to-end trust and privacy for stakeholders across European data

2. Design and develop a distributed Homographic-capable self-sovereign framework for accessing, sharing and manipulating data, compliant to national and EU legislation

3. Novel Big Data management and analytics infrastructure to facilitate use and re-use of data in data spaces across interdisciplinary science fields and business sectors
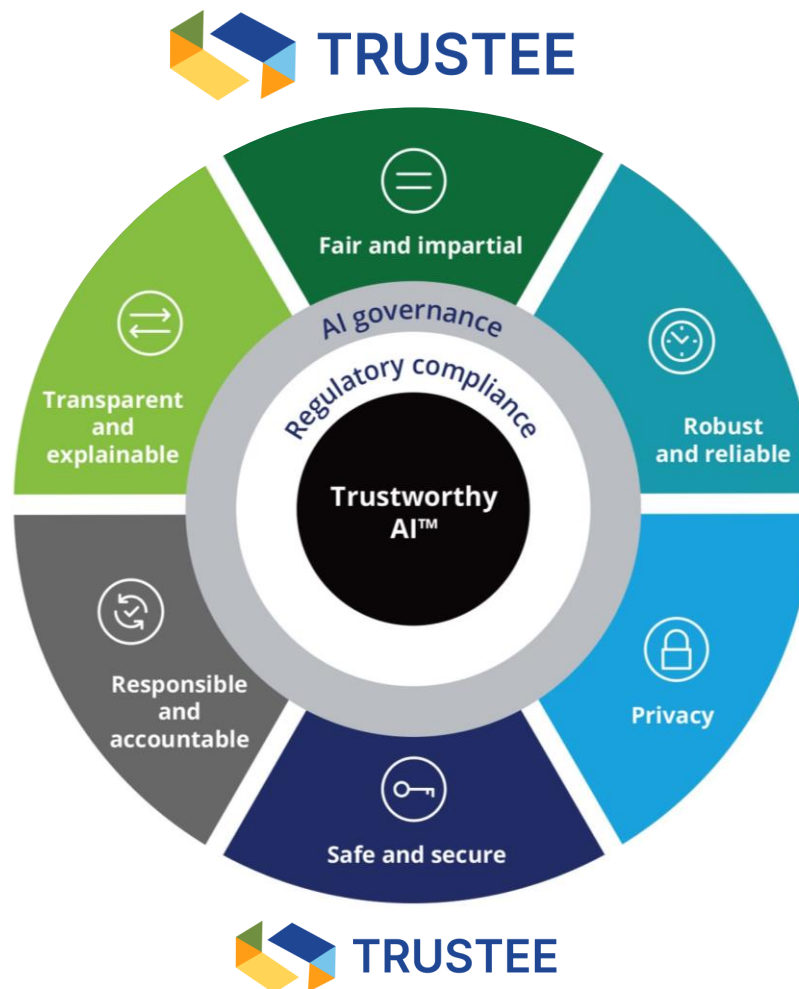
4. Analyze the ethical, legal, privacy and IPR issues for collection, storage, processing, querying, analytics and delivery of data enabling the European single market for data

5. Design and develop TRUSTEE explainable and trustworthy AI for efficient and robust use and re-use of data and metadata across interdisciplinary domains

6. Provide FAIR databases facilitating sharing and manipulation of data in compliance with prevailing and emerging legislation (e.g. GDPR)

TRUSTEE

AI governance
Regulatory compliance

Trustworthy AI™

Fair and impartial
Robust and reliable
Privacy
Safe and secure
Responsible and accountable
Transparent and explainable

TRUSTEE

Value creation for healthcare

People
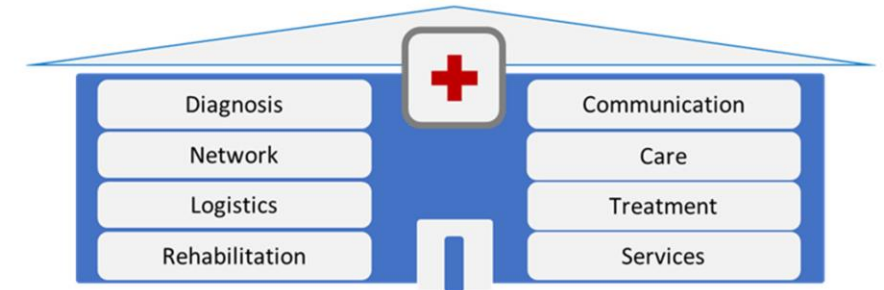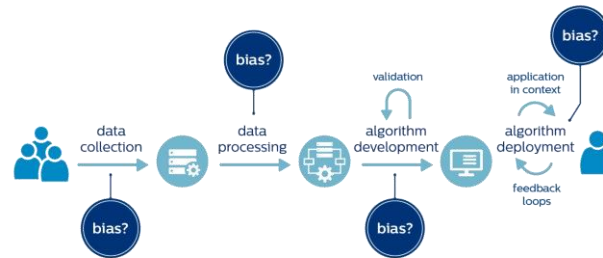Convince people that data sharing will advance science and engineering, and also their own scientific work.

Metadata and ontologies
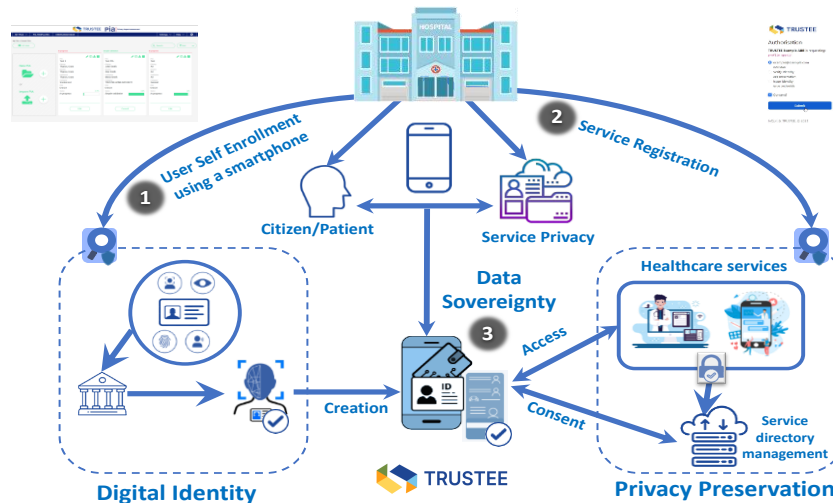Develop metadata schemas, parsers, converters and ontologies.

Data processing and analysis
Develop and offer ontology-derived workflows, a materials encyclopedia and an AI toolkit.

Infrastructure
Consider the storage, retrieval, transfer and processing of exponentially growing data volumes. Develop software for centralized as well as local servers and for a federated data infrastructure.

data collection → data processing → algorithm development (validation) → algorithm deployment (application in context, feedback loops)

bias?

Diagnosis | Communication
Network | Care
Logistics | Treatment
Rehabilitation | Services

Specific application potential regarding AI applications in hospitals

① User Self Enrollment using a smartphone
② Service Registration
③ Data Sovereignty

Citizen/Patient
Service Privacy
Healthcare services

Digital Identity
Creation
Access
Consent
Service directory management
Privacy Preservation

	Superior challenges	COVID-19 // Pandemic aggravating factors	AI solutions and opportunities
Diagnosis	Increasing medical complexity / Increasing diversity of diseases	Further supply of the regular cases / Need of short reaction time	Advanced support decision making tools
Care	Increasing chronic diseases / Increasing number of older, frailer, multi-morbid patients	Need for contactless medical care	Pathways modelling / Automated simulation & prognosis / Robots (care quality level & improve security)
Logistics	Increasing economical pressure / Shortage of skilled medical staff	Need of networking and rapid transmission of information	Automated planning of resources / Autonomous systems for knowdledge & information management

Edward Curry

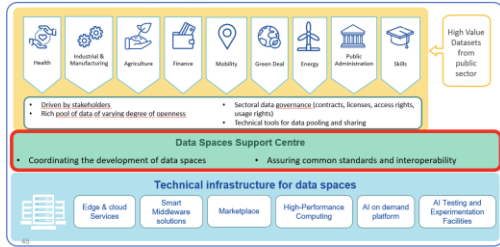I have been researching the underlying technology for data spaces for the last decade...
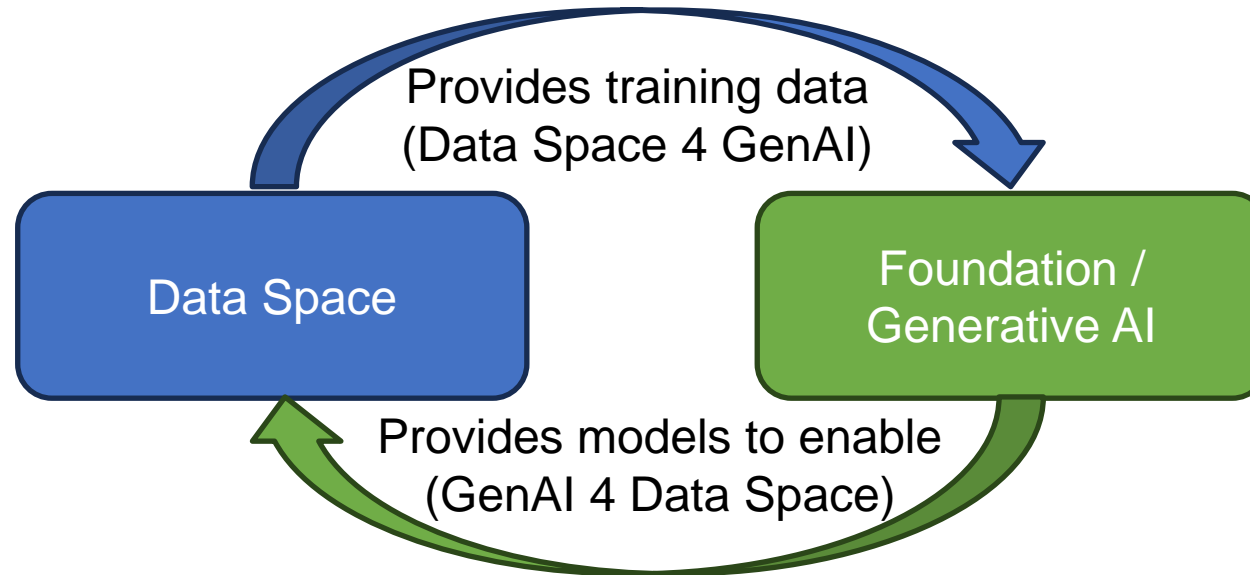
Data Spaces Symposium
Unite. Innovate. Adopt.
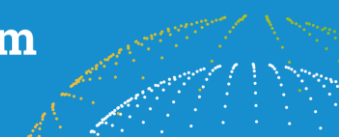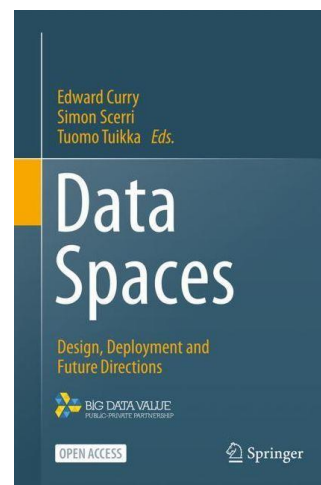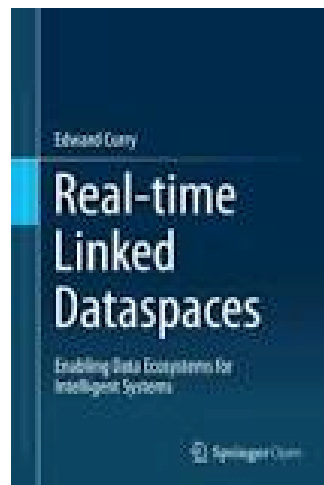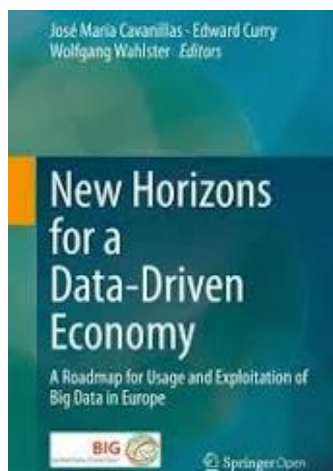
Symbiotic Relationship between Data Spaces and AI....
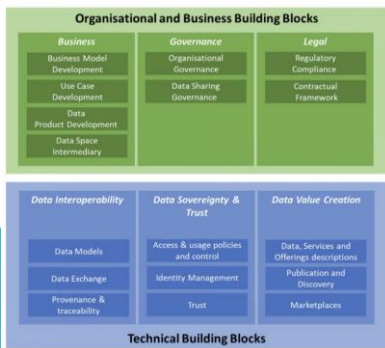
High Quality Data
Community of Data Users/Owners
Data Goverance/compliance

Provides training data
(Data Space 4 GenAI)

Data Space

Foundation / Generative AI
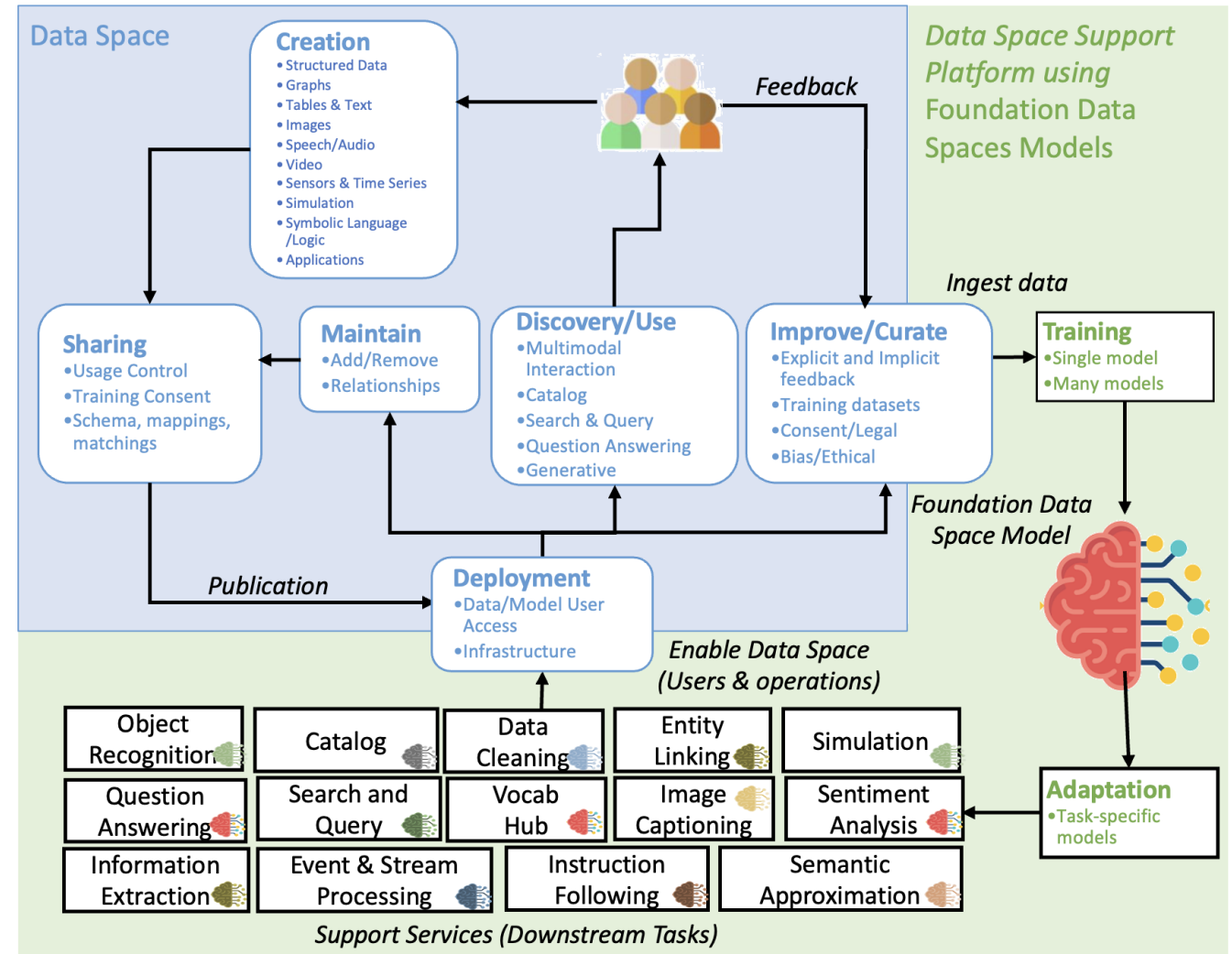
Provides models to enable
(GenAI 4 Data Space)

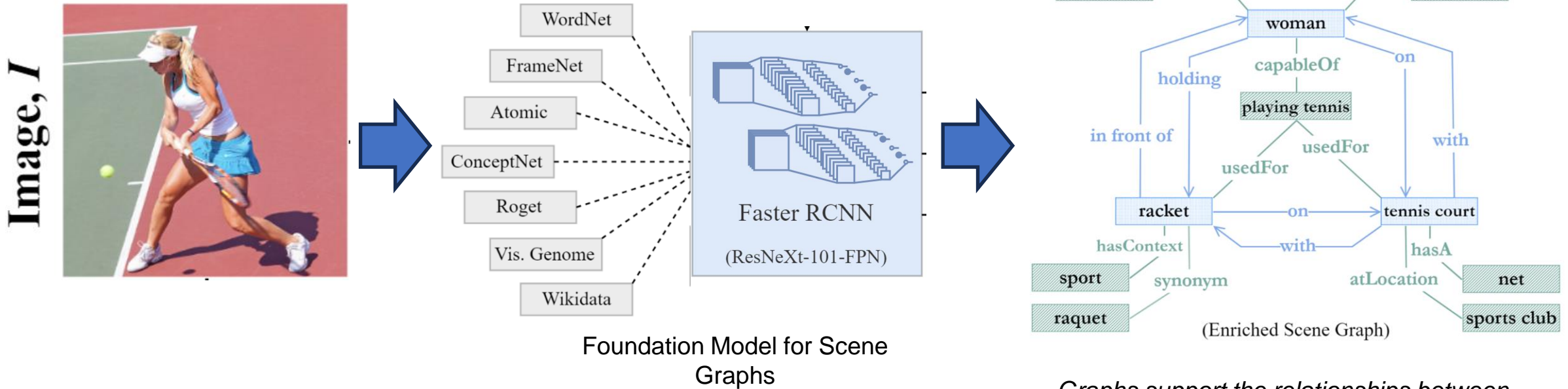Extract Value from Data
Faster innovation cycles
Synthetic Data
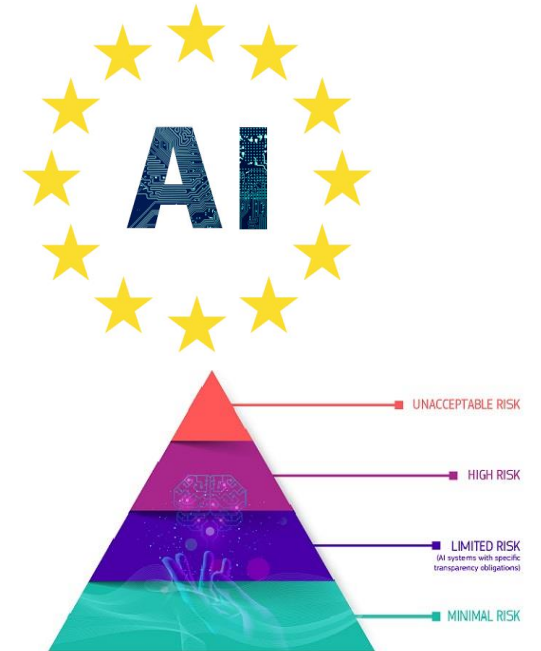
Foundation Data Space Models.....

A foundation data space model is any model that is trained on broad data (generally from the data space and using self-supervision at scale) that can be adapted (e.g., fine-tuned) to a wide range of downstream tasks to support the life cycle of the data space.
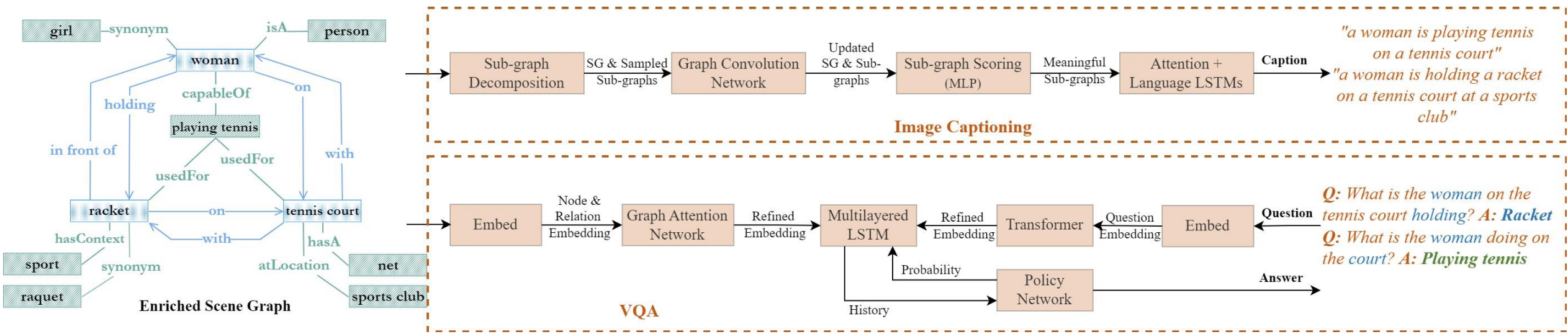
E. Curry, M. Timilsina, T. Zaarour, M. Al-QATF, R. Haque, "Foundational Data Space Models: Bridging the AI and Data Ecosystems (Vision Paper), Proceedings of the *2023 IEEE International Conference on Big Data (Big Data)*, Sorrento, Italy, 2023

Insight

Scene Graph Foundation Model for Multimodal Semantic Interoperability...

Foundation Model for Scene Graphs

(Enriched Scene Graph)

Graphs support the relationships between data space resources

.J. Khan, J.G. Breslin, E. Curry, "Expressive Scene Graph Generation using Commonsense Knowledge Infusion for Visual Understanding and Reasoning," Extended Semantic Web Conference (ESWC) 2022, Crete, Greece, May-June 2022.

Insight

Downstream Semantic Tasks:
- Image Captioning & Multimodal
- Question Answering

.J. Khan, J.G. Breslin, E. Curry, "Expressive Scene Graph Generation using Commonsense Knowledge Infusion for Visual Understanding and Reasoning," Extended Semantic Web Conference (ESWC) 2022, Crete, Greece, May-June 2022.

Insight

Research Directions

Unified Data-AI Ecosystem and Business Model

- Single data and AI life cycle
- Incentivized to share data, models, and computing infrastructure.
- Shared development toolkits

Data Space Interoperability

- Model-based "good-enough" semantic interoperability
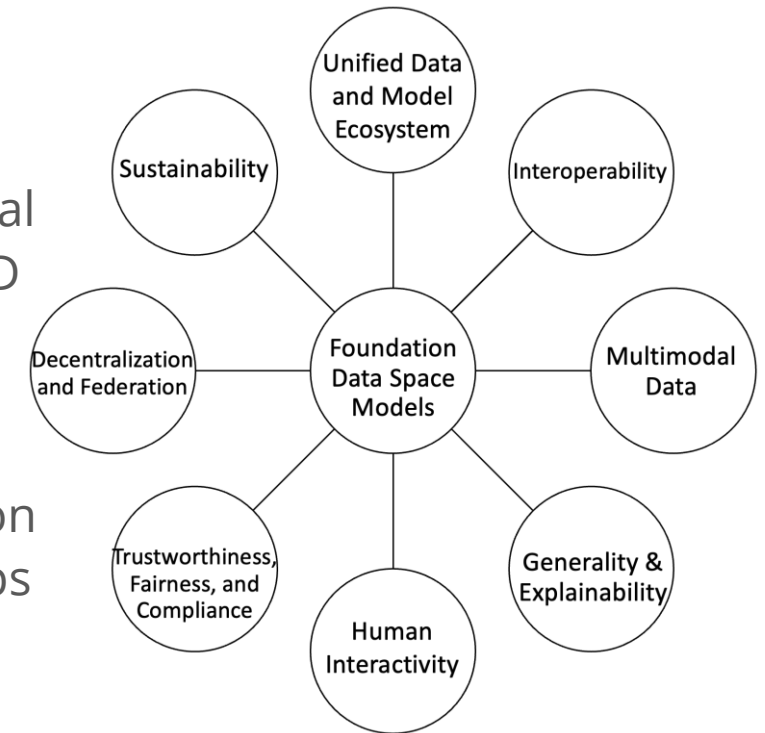- Share semantic interpretation models between dataspace (domain-specific)

Multimodal Data

- Establish connections and relationships
- Cross-modal information retrieval
- From real-time sensor data to 3D geospatial models

Robustness

- Model monitoring and adaptation
- Unified data-model lifecycles/Ops

Trusted, Ethical, and Compliant

- Intrinsic biases
- Integrity & transparency
- AI and data governance

.J. Khan, J.G. Breslin, E. Curry, "Expressive Scene Graph Generation using Commonsense Knowledge Infusion for Visual Understanding and Reasoning," Extended Semantic Web Conference (ESWC) 2022, Crete, Greece, May-June 2022.

Insight

DATA SPACES SHOULD ALLOW FASTER INTEGRATION OF MORE DIVERSE DATASET INTO ALGORITHMS… BUT WE'RE NOT THERE YET

ALGORITHM-CENTRIC AUTOMATED DATA INTEGRATION

The PISTIS Project

Federated data management interoperability & governance

Federated secure data sharing

Data valuation and monetisation

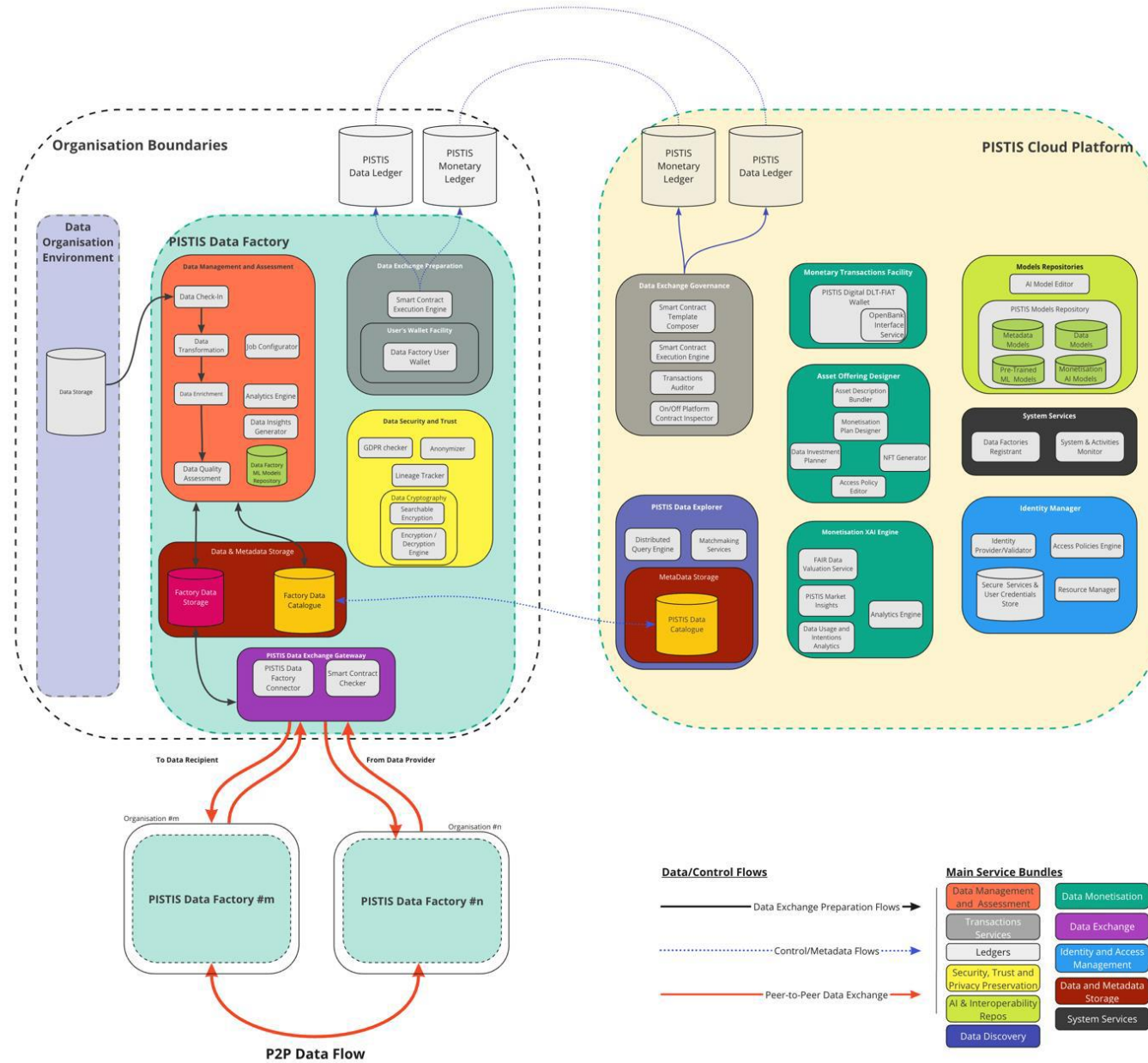Data sharing skills cultivation

MOBILITY and URBAN PLANNING

ENERGY

AUTOMOTIVE

Catena-X was founded
… to tackle our industries most pressing issues, together.

TRUSTED SUSTAINABILITY

Report Product Carbon Footprint (incl. primary data) — *as of 2025 annual report*

Scale Circularity with Product & Battery Passports — *as of 2024 / 25*

Environmental & Social Governance

RESILIENT SUPPLY-CHAINS

Master Data Services & Certificates

Ensure Material Flow with Demand / Capacity Mgt. — *today*
- Short
- Long-Term

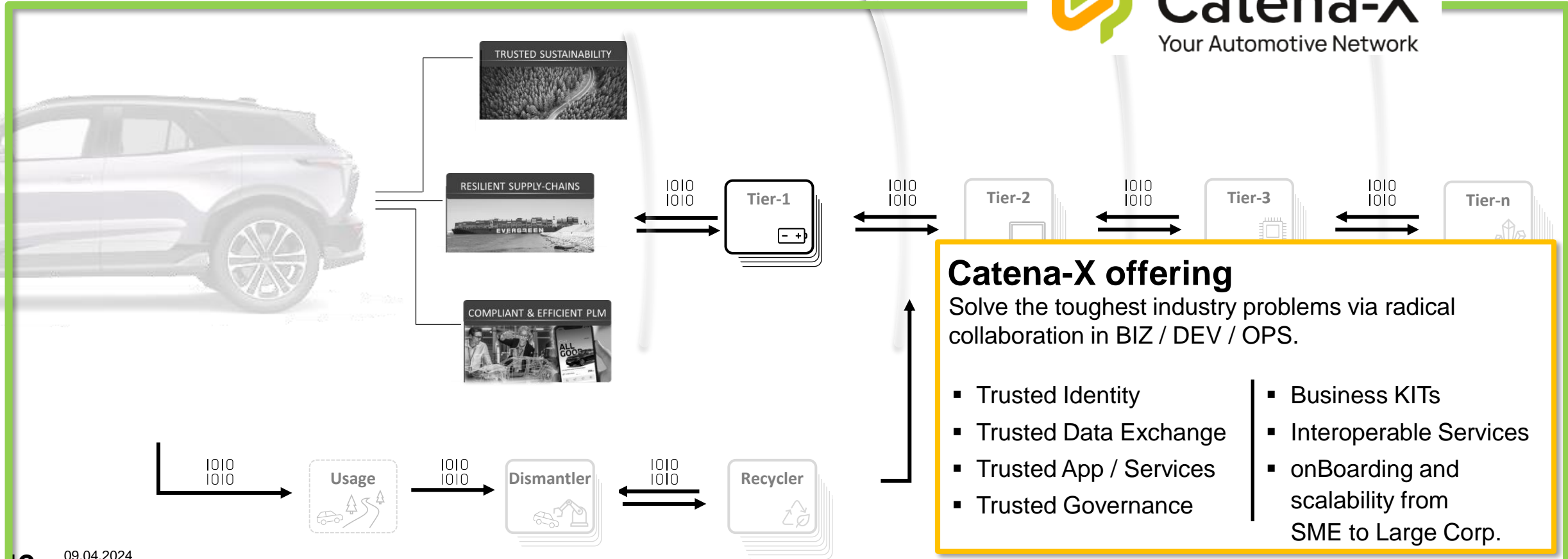**COMPLIANT & EFFICIENT PLM**

Traceability of Parts & Contain non-compliant parts

Live Quality Loops & Analysis — *today*

Behavior Twins

Catena-X Value Proposition

Provide a trusted, compliant, global and interoperable environment for the industry, to create and utilize data driven value chains for dedicated business processes.

Catena-X
Your Automotive Network

TRUSTED SUSTAINABILITY

RESILIENT SUPPLY-CHAINS

COMPLIANT & EFFICIENT PLM

Tier-1

Tier-2

Tier-3

Tier-n

Usage

Dismantler

Recycler

Catena-X offering
Solve the toughest industry problems via radical collaboration in BIZ / DEV / OPS.

- Trusted Identity
- Trusted Data Exchange
- Trusted App / Services
- Trusted Governance

- Business KITs
- Interoperable Services
- onBoarding and scalability from SME to Large Corp.
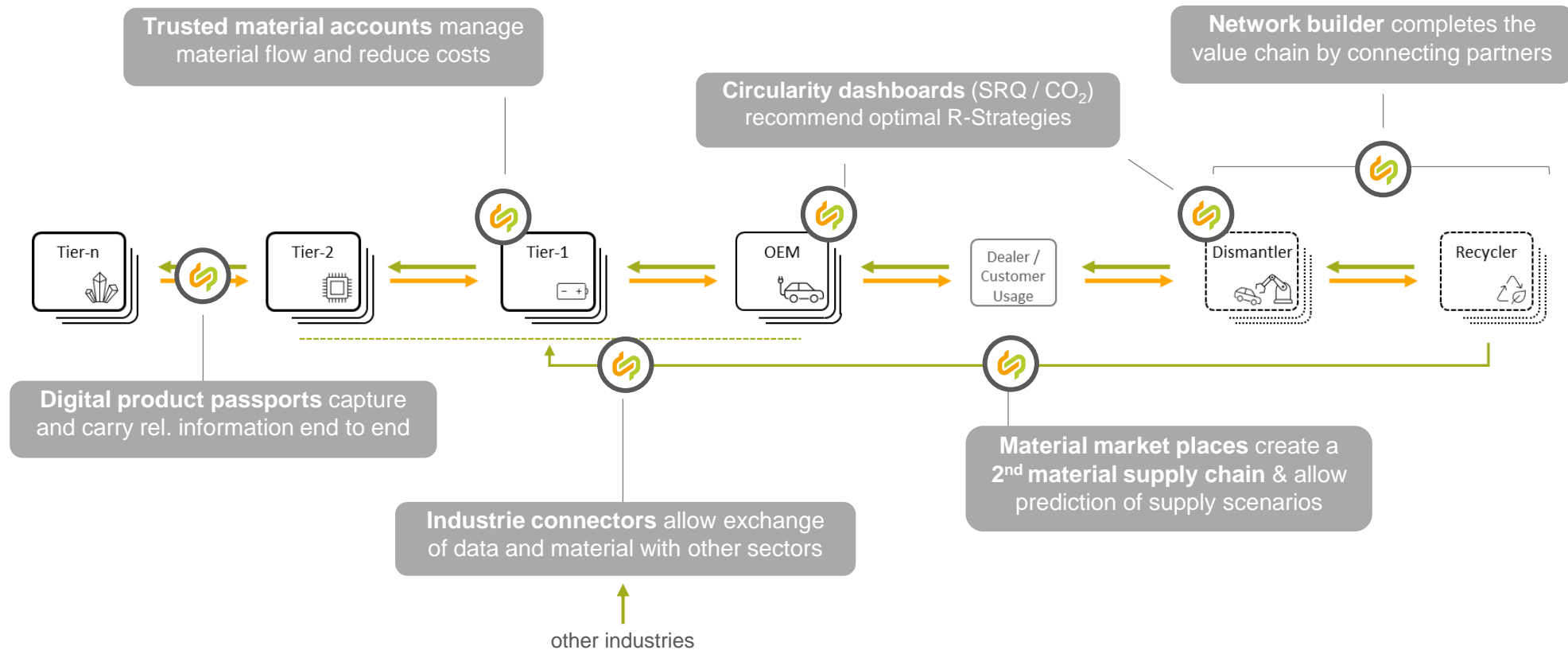
Circular Economy

Fast Facts

100 billion
tons of material resources consumed annually

< 9%
of resources are recycled or reused

~ 30%
secondary material ratio per vehicle

$4 Trillion
demand for battery materials up to 2050

>30.000
to be connected industry partners (to close loops)

Catena-X Standards & Artifacts

Connect partners, streamline flow of information and create solution portfolio to enable scalable value derivation

Trusted material accounts manage material flow and reduce costs

Network builder completes the value chain by connecting partners

Circularity dashboards (SRQ / CO_2) recommend optimal R-Strategies

Tier-n   Tier-2   Tier-1   OEM   Dealer / Customer Usage   Dismantler   Recycler

Digital product passports capture and carry rel. information end to end

Material market places create a **2nd material supply chain** & allow prediction of supply scenarios

Industrie connectors allow exchange of data and material with other sectors

other industries

Questions?
Get in touch with us!

Catena-X Automotive Network e.V.
Reinhardtstraße 58
10117 Berlin
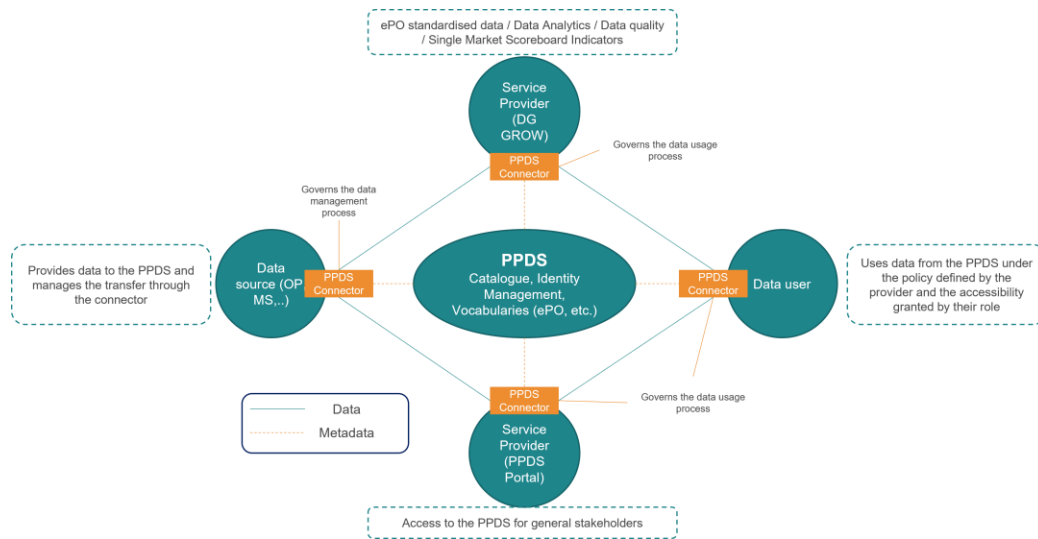
info@catena-x.net

Catena-X
Your Automotive Network

Hello! My name is Aitor Corchero, and I would like to discuss about DS in public administrations

I am Semantic Domain Expert at NTT Data focused on bringing Semantic interoperability to public sector

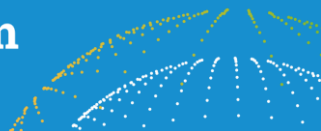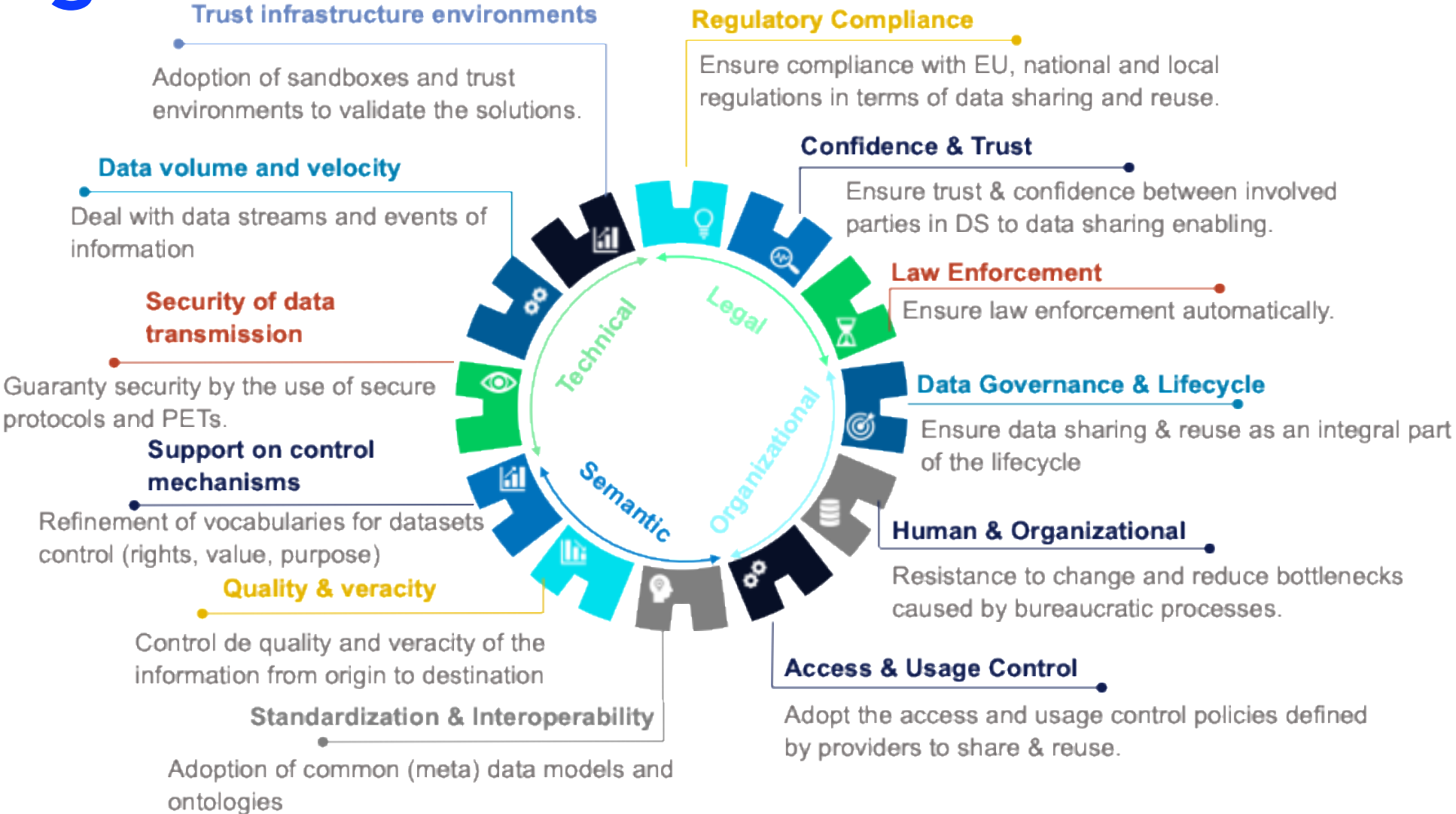**Data Spaces Symposium**
Unite. Innovate. Adopt.

Challenge & Needs Data Spaces in Public Services

1. **Active involvement of public administrations and companies** providing services to such administrations
2. **Improve data interoperability and quality** to enable further exploitation of data through analytical services;
3. **Elaboration and generation of digital services** to ensure the implementation of governance procedures in an efficient manner;
4. **Establishing certification mechanisms** for applications and services to establish trusted application catalogues;
5. **Automatically ensuring law enforcement** verification and control processes (e. g. GDPR).

Key Messages

Trust infrastructure environments

Adoption of sandboxes and trust environments to validate the solutions.

Data volume and velocity

Deal with data streams and events of information

Security of data transmission

Guaranty security by the use of secure protocols and PETs.

Support on control mechanisms

Refinement of vocabularies for datasets control (rights, value, purpose)

Quality & veracity

Control de quality and veracity of the information from origin to destination

Standardization & Interoperability

Adoption of common (meta) data models and ontologies

Regulatory Compliance

Ensure compliance with EU, national and local regulations in terms of data sharing and reuse.

Confidence & Trust

Ensure trust & confidence between involved parties in DS to data sharing enabling.

Law Enforcement

Ensure law enforcement automatically.

Data Governance & Lifecycle

Ensure data sharing & reuse as an integral part of the lifecycle

Human & Organizational

Resistance to change and reduce bottlenecks caused by bureaucratic processes.

Access & Usage Control

Adopt the access and usage control policies defined by providers to share & reuse.

The European AI On-Demand Platform (AIoD)
AIoD as Bridge & Catalyst between the European AI research and industry
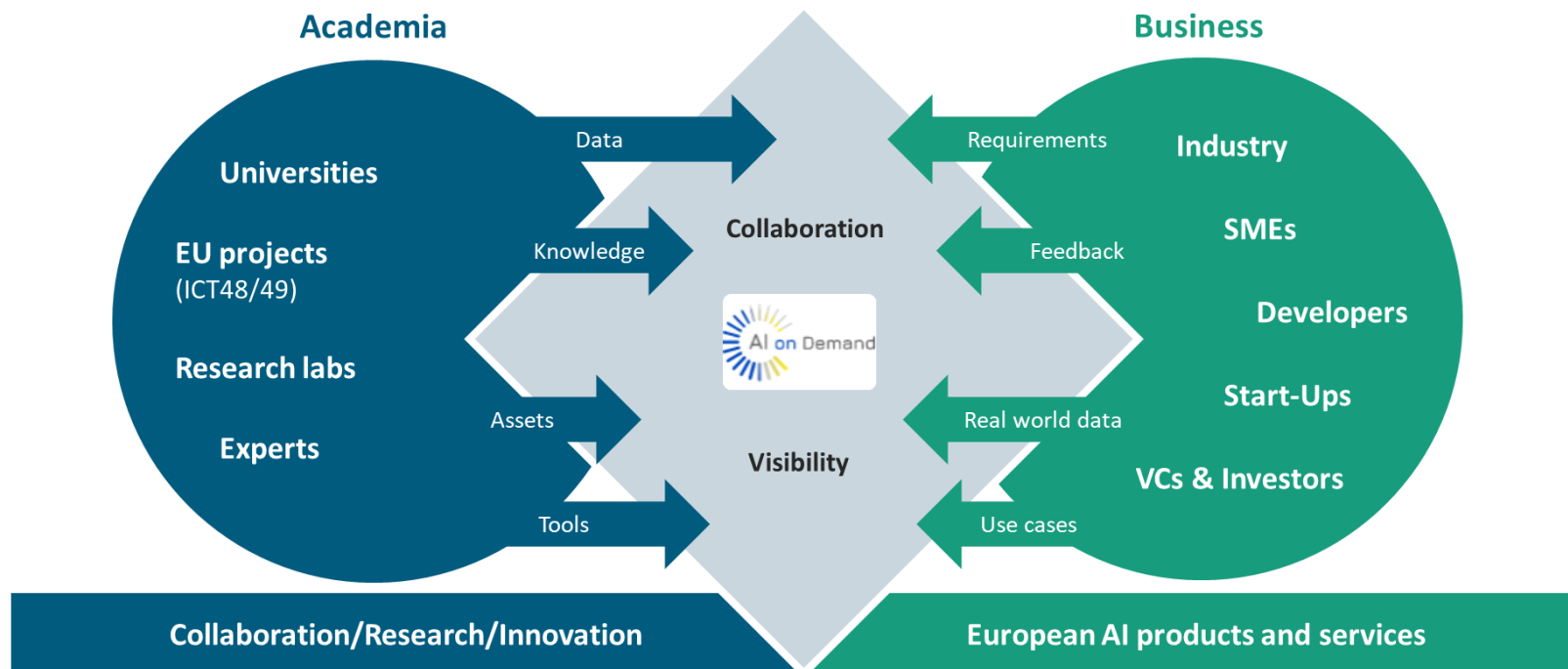

- **one-stop shop** for exchanging and building **trustworthy AI resources and applications**

- support **European enterprises** and the public sector

- ensure European **sovereignty** in AI, especially in **Generative AI**

Deploy AI

European AI-Technology: Cloud native Open Source AI-Technology Access
i.e. DevOps tools / Application Builder / Docker / European Large Language Models

European DATA: Seamless access to European Data Spaces
i.e. Health, Public Administration, Skills, Manufacturing, etc.
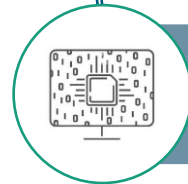
European High Performance Computing Ressources
Seamless access to European HPC's for testing and AI-training

EU AI-Trustworthiness
EU regulatory compliance and trustworthiness support

European AI-Community:
Qualified AI Consulting Network / Access to talents / Training Resources / Professional Services

Data Spaces Symposium

DATAWEEK²⁴
JOIN.LEARN.SHARE.GET VALUE

15:30

International Manufacturing-X Council: Make Data Work!

Davide dalle Carbonare, Oscar Lázaro, Thomas Hahn,
Sicco Lehmann-Brauns & Sergio Gusmeroli