

# Data Spaces Symposium

11:00

**Data spaces  
unlocking AI innovation**

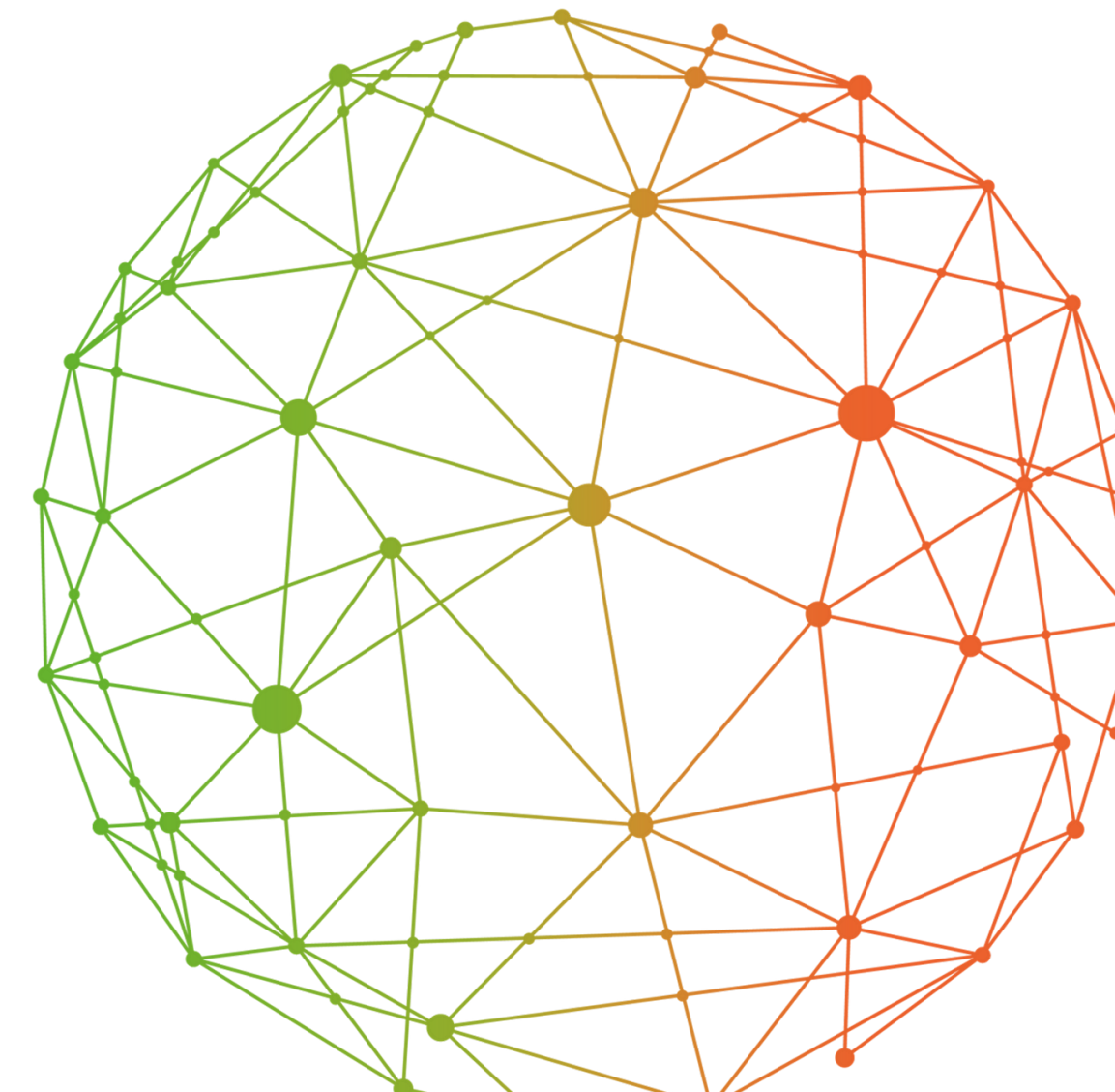
---

Focus session [AI]

# The EU Data and AI Value Chain

## Data Spaces unlocking AI Innovation

Data Spaces Symposium 2025



The Data Spaces Support Centre receives funding from the European Union Digital Europe Programme under grant agreement n° 101083412

# Industrial AI

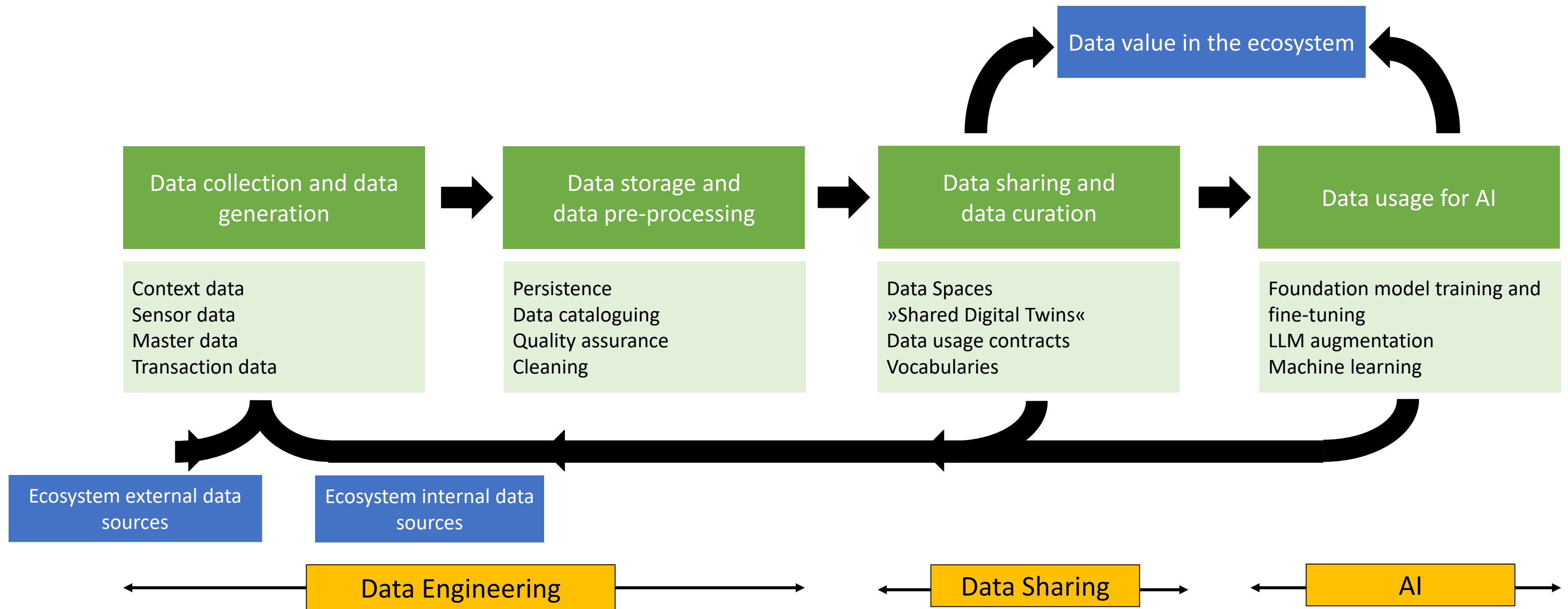
## Analyzing typical use cases

	1 Predictive AI User	2 Predictive AI Creator	3 Generative AI User	4 Generative AI “Booster”	5 Generative AI Co-Creator
Use case	IE consumes predictive AI services from a vendor for smart maintenance, logistics etc.	Using external (e.g. customer) data to offer predictive AI customer services for maintenance, intralogistics etc.	IE uses generative AI to automate manufacturing, business, and development processes	IE enriches foundation models (using RAG, fine-tuning etc.) to improve operational efficiency, customers services etc.	IE teams up with peers and contributes data to jointly develop a shared foundation model
IE Role	– User	– Creator	– User	– User – Creator	– Co-Creator – User
Ecosystem	– IE – Service provider	– Customers – IE	– IE – LLM provider	– IE – LLM provider	– IE – other IEs – AI service provider

Legend: LLM – Large Language Model; AI – Artificial Intelligence; IE – Industrial Enterprise.

# Integrated Data and AI Value Chain

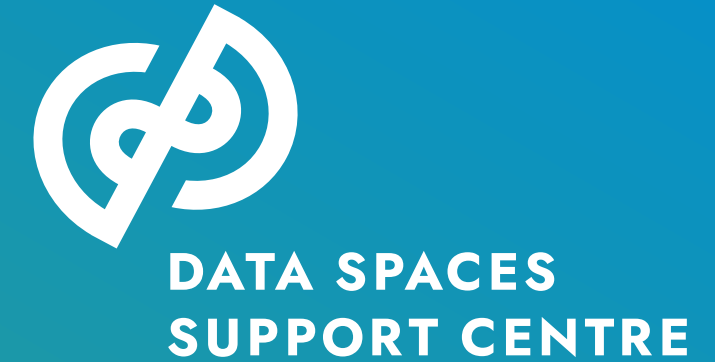
AI is nothing without data



Source: adopted from OECD (2020). Legend: LLM – Large Language Model; AI – Artificial Intelligence.

# European Competitiveness

## The need to speed up



- The Draghi Report proposes a sector-specific AI strategy: "EU Vertical AI Priorities Plan":
- Shared AI model development across sectors: Strategic AI integration in 10 key industries (automotive, energy, healthcare, etc.)
- Cross-industry data pooling to overcome Europe's lack of large datasets ("for free").
- Balance in supporting European cloud industry with securing key technologies amid US dominance.
- Key challenges: Companies hesitate to share data (competition concerns, lack of incentives, regulatory uncertainty)

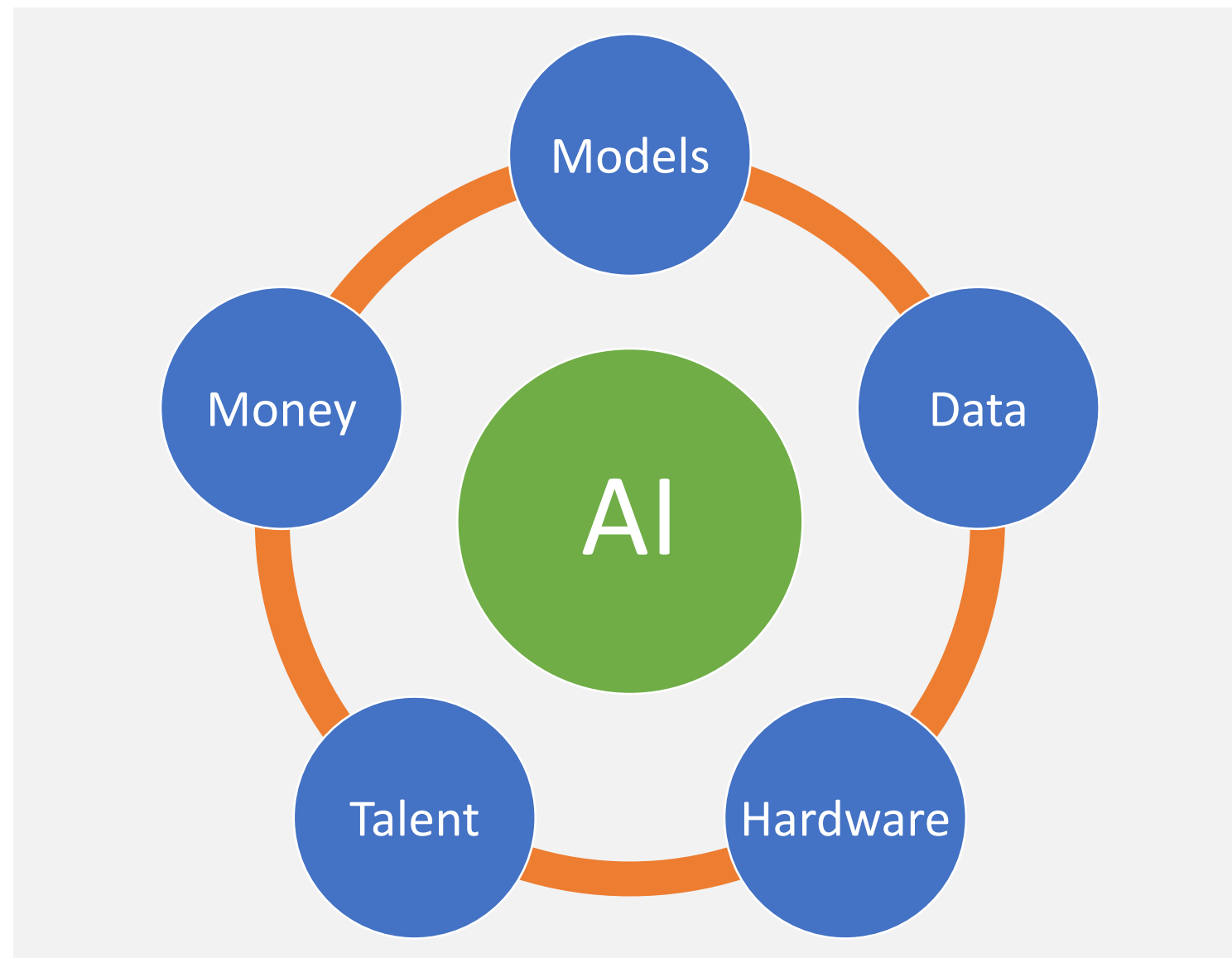
**"The EU should promote cross-industry coordination and data sharing to accelerate the integration of AI into European industry."<sup>1</sup>**

The Data Spaces Support Centre receives funding from the European Union Digital Europe Programme under grant agreement n° 101083412.

1) European Commission: The Draghi report on EU competitiveness (2024). - [https://commission.europa.eu/topics/eu-competitiveness/draghi-report\\_en](https://commission.europa.eu/topics/eu-competitiveness/draghi-report_en)

# Resources for Successful AI

## Understanding the EU AI Value Creation System



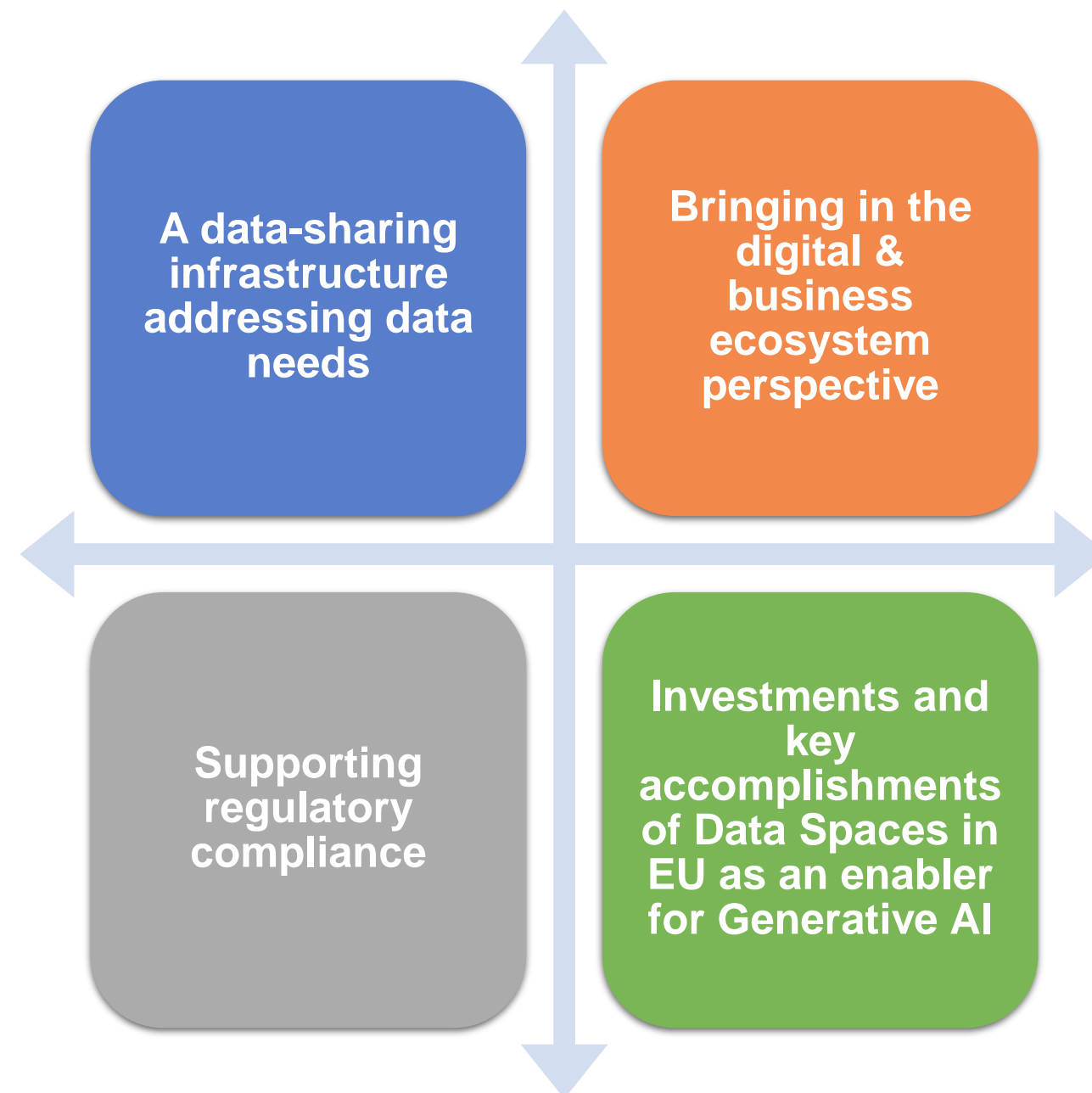
- Regarding all relevant AI resources, Europe's position is good, partially even excellent
- However, unlike in other economic areas, control over these resources is not in the hand of a few or the state, but distributed among various stakeholders
- Thus, the European Data Union shall embrace the concept of sharing resources and activate the ecosystem
- A prerequisite for this is trust among ecosystem members and sovereignty over data
- Thus, data spaces are a central enabler for the European Data Union

# DSSC White Paper

## Data Spaces facilitate the AI Value Creation System



### Data spaces' transformative potential for Generative AI



The paper is openly available through [dssc.eu](https://dssc.eu)





**DATA SPACES  
SUPPORT CENTRE**

# The EU Data and AI Value Chain

Boris Otto · Warsaw · 12 March 2025

The Data Spaces Support Centre receives funding from the European Union Digital Europe Programme under grant agreement n° 101083412.

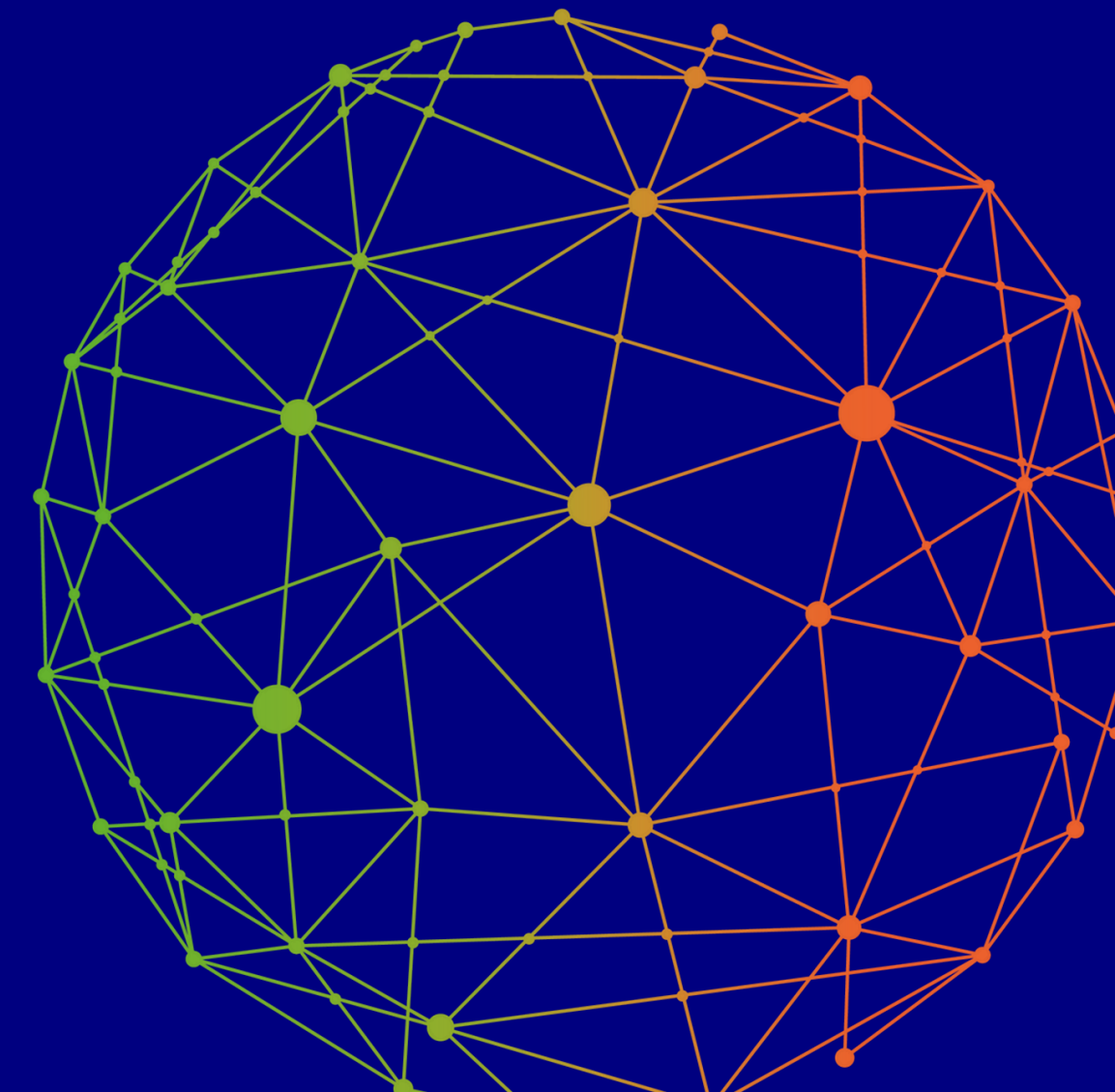




# AI driven Polish economy

Piotr Mieczkowski

Data Spaces Symposium 2025



DSBA



INTERNATIONAL DATA SPACES ASSOCIATION



DATA SPACES SUPPORT CENTRE



Funded by the European Union

The Data Spaces Support Centre receives funding from the European Union Digital Europe Programme under grant agreement n° 101083412

# Polish AI ecosystem in nutshell

01

## Leading cities

AI development concentrated in six large metropolitan areas. Warsaw is the AI Capital of Poland

02

## AI Companies

420+ Polish dynamic AI companies with revenue mix from Poland and abroad. This includes i.a. nomagic.ai, ElevenLabs, Lingaro, Synerise, Netguru, Tooploox, RTB House, STX Next, Predica, Spyrosoft and many more!

03

## R&D | GDS

40+ global companies that placed R&D, GDS, SDC in Poland. This includes i.a. Amazon, Aptiv, Citi, Google, Intel, EY, Nokia, Nvidia, Ringier Axel Springer Tech, Samsung, T-Mobile, TomTom, UBS

04

## AI Policy 2020

This includes i.a. creating AI Virtual Research Institute, Three centers of AI excellence, New AI Institute - NCBR Ideas, 190 mln USD funding for applying AI, Open data policy  
In 2025 AI policy will be under Review.

05

## Institutions

Set of institutions supporting the development of AI in Poland. This includes i.a. NCBR, NCN, IPI PAN, NASK PIB, OPI PIB, Ideas Institute, Digital Poland Foundation with aipoland initiative

06

## Science

Several Scientific AI societies. In 2018, Polish Initiative for the Advancement of Artificial Intelligence (PP-RAI) was formed as an umbrella organization

07

## AI events

A series of leading events on artificial intelligence that bring together experts and researchers in the field. This includes i.a. PP-RAI, ICAISC, ML in PL, Data Science Summit, PyData Warsaw

08

## AI Communities

A dozen of dynamic communities and study groups that meet regularly, which a.i. include Warsaw.ai, MLGdańsk, Deep Learning Labs, Robotics Association, SKALP, GMUM, Data Science Warsaw, DataOps Poland

# Some leading AI compnies from Poland

allegro

COMARCH

lingoro

SALESmanago

ardigen

SYNERISE

netguru

TOOPLOOX

VRTUSLAB

sages

RTB HOUSE

Applica

deepsense.ai  
BIG DATA SCIENCE

Future Processing

PGS SOFTWARE

digica

BRAINLY

DE SOFTWARE

edrone

Infermedica

ONWELO

STXNEXT

Nethone

predica.

spyrosoft

alphamoon

SOFTWAREMILL

neurosys

nomagic

neoteric

neptune.ai

VoiceLab

LEKTA

yosh.AI

addepto

sotrender

BRAIN SCAN

cosmose AI

EMPLOCITY

SAG\*

StethoMe


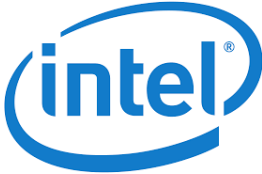






















GetResponse

MIM SOLUTIONS

senti one

codewise

# Examples of leading AI R&D, GDS, SDC centres in Poland

 Nokia	 Intel	 UBS	 Google	 Samsung R&D Institute	 IBM	 Roche	 Accenture
 Epam	 BCG Gamma	 SII Group	 EY	 Capgemini	 ABB	 Aptiv	 TCL Research Europe
 Citi	 P&G	 Amazon	 IQVIA	 ING	 TomTom	 Ocado Group	 OLX Group
 Atos	 Nvidia	 T-Mobile	 Bosh	 ByteDance	 Pearson OKI	 Revolut	 Nielsen IQ
							

# Selected institutions supporting the development of AI in Poland

In Poland, there are several institutions supporting the development of artificial intelligence. These range from basic research to implementation to research on the wider impact of the technology on society.



Funding for basic research into artificial intelligence



Funding the implementation of artificial intelligence

**NASK**

Research on artificial intelligence



Research on artificial intelligence



Conducting fundamental research



Supporting digital and AI initiatives in Poland



Research on artificial intelligence, promoting artificial intelligence in society



Research on artificial intelligence



Supporting artificial intelligence in law



Supporting artificial intelligence in science



Research on artificial intelligence



Research on artificial intelligence



# PLLuM – Polish Large Language Model

PLLuM (Polish Large Language Model) is a project that aims to create a Polish-language large language model and an intelligent assistant based on this model. Key features of PLLuM are its reliance on Polish-language content, its **open and free nature**, and its **compliance with the principles of ethical and responsible artificial intelligence**.

The model is being developed by a consortium of **six leading Polish research units**, four of which are based in Warsaw, including the National Research Institute NASK, the Information Processing Centre of the National Research Institute, the Institute of Computer Science Basics of the Polish Academy of Sciences and the Institute of Slavic Studies of the Polish Academy of Sciences. The consortium leader is the Wrocław University of Technology.

Warsaw, as the administrative centre of the country, can benefit from PLLuM applications, especially in the context of the planned intelligent assistant, which aims to increase the availability of public services. **The PLLuM became made public for widespread use in february 2025.**



# Bielik.ai – bottom-up movement for Polish LLM

Bielik is a Polish language model created by the SpeakLeash a.k.a. a 'Spichlerz' community present in Warsaw and Wrocław in collaboration with the AGH University of Science and Technology in Kraków (Cyfronet computers were used to train the model). The model is a grassroots response to international solutions such as ChatGPT, but with an emphasis on better support for the Polish language. Bielik enables conversations, answering questions and generating various content, including literary content such as poems and stories.

Its main characteristics:

- Adaptation to the specific needs of Polish-speaking users.
- Openness, which allows for further development of the model by the community.

However, despite its advantages, Bielik struggles with some challenges, such as the accuracy of the information provided and the recognition of well-known characters. These difficulties are due to the limited amount of data in Polish and the smaller number of parameters compared to larger models such as GPT-3.












Nevertheless, Bielik is a project that is constantly evolving and has the potential to reach even higher levels of sophistication in the future. Despite its limitations, it represents an important step towards adapting AI to local cultural and linguistic needs. It proves that local initiatives can significantly contribute to the development of AI in Poland, with the potential to further refine the model as more data and experience are acquired.

Specialists in Warsaw are playing a key role in the development of the Bielik project. Many of the people involved in the project live and work in the capital, showing that the initiative has a broad reach and is not geographically limited. Experts from Warsaw bring valuable experience and expertise to the project, which supports the development of the Polish language model.

> **SpeakLeash**  
/'spix.lɛʂ/ a.k.a Spichlerz

# plGRID – HPCs network

The Polish PLGrid infrastructure is managed by the PLGrid Consortium, established in January 2007 and comprising six scientific entities.

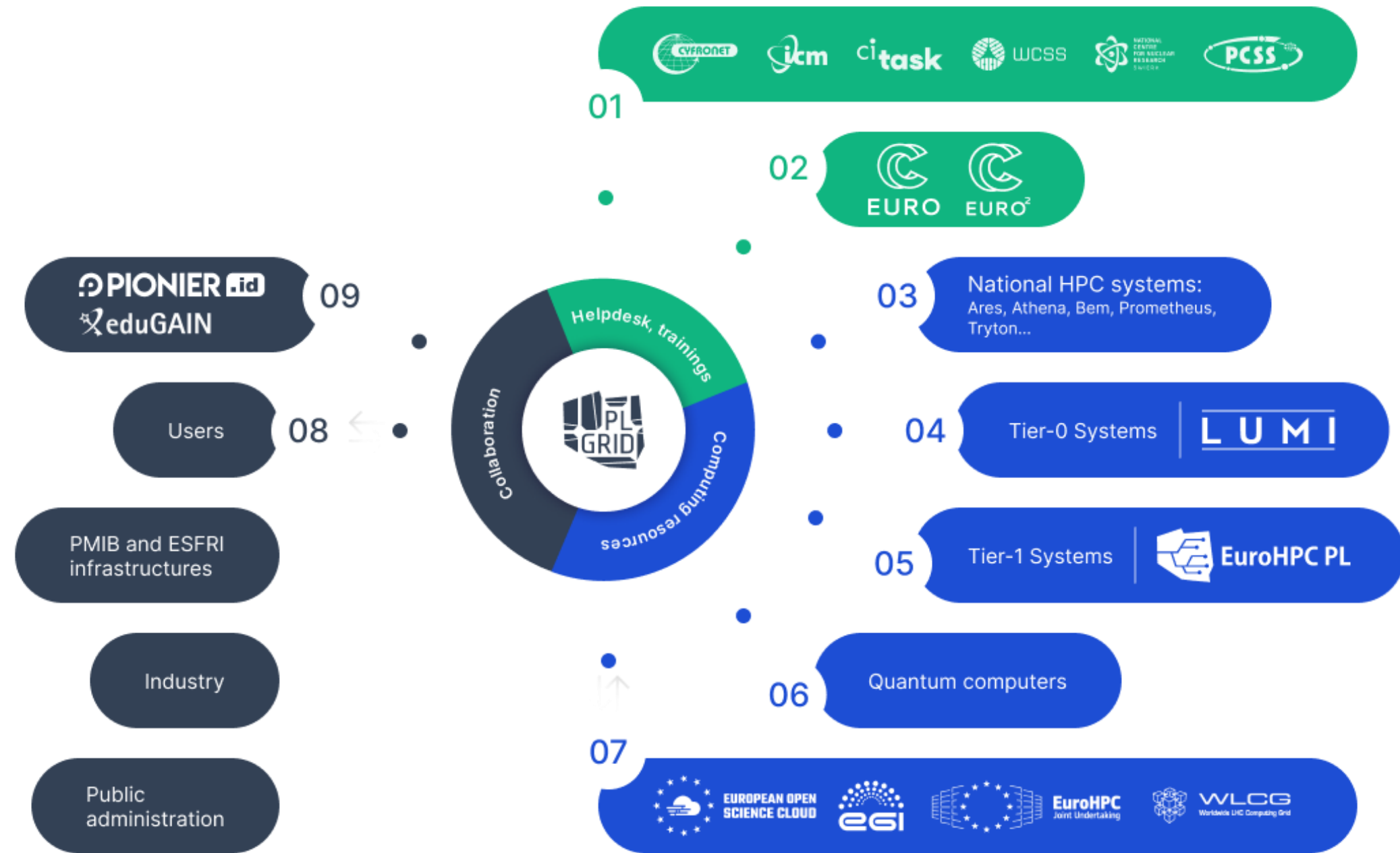
  Akademickie Centrum Komputerowe CYFRONET AGH w Krakowie	  Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego w Warszawie	  Poznańskie Centrum Superkomputerowo Sieciowe
  Centrum Informatyczne Trójmiejskiej Akademickiej Sieci Komputerowej w Gdańsku	  Wrocławskie Centrum Sieciowo- Superkomputerowe	  Narodowe Centrum Badań Jądrowych w Otwocku-Świerku



# pGRID – HPCs network

Helps Polish researchers conducting advanced computing and simulations

PLGrid is an environment combining world-class computing resources and specialist expertise, developed to support the research and development sphere in Poland. It facilitates the solution of important research problems and provides tools to accelerate the development of innovative technologies.

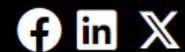


# AI Factories – ACK Cyfronet AGH

## W AGH powstanie Fabryka Sztucznej Inteligencji

06/11/2024

sztuczna inteligencja



Blisko 70 mln zł rządowej subwencji ma otrzymać AGH na realizację projektu budowy pierwszej w Polsce Fabryki Sztucznej Inteligencji. Powstanie ona w Akademickim Centrum Komputerowym Cyfronet AGH. Projekt współfinansuje Komisja Europejska.

*– Inwestycja w budowę Fabryki AI w Cyfronecie to kluczowy krok w kierunku umocnienia pozycji Polski w europejskim ekosystemie sztucznej inteligencji. Dzięki wsparciu finansowemu ze strony państwa oraz współpracy z Komisją Europejską rozwiniemy infrastrukturę obliczeniową, która pozwoli naszej uczelni i krajowi wkroczyć na nowy poziom zaawansowania technologicznego. – powiedział Rektor AGH prof. Jerzy Lis. – Jestem przekonany, że ten projekt przyniesie nie tylko znaczące korzyści naukowe, ale także pozytywnie wpłynie na rozwój gospodarki opartej na innowacjach. Wyrazy uznania należą się także dyrekcji Cyfronetu oraz całemu eksperckiemu zespołowi tej jednostki, którzy będą zaangażowani w projekt o ogromnym potencjale i znaczeniu dla przyszłości technologicznej Polski.*

Fabryka AI będzie centrum badawczo-rozwojowym, które pozwoli polskim naukowcom i naukowczyniom rozwijać, współtworzyć, testować i wykorzystywać najnowsze technologie oparte na sztucznej inteligencji. Powstanie poprzez rozbudowę oraz wzmocnienie komputerowych i obliczeniowych zasobów Cyfronetu.

Realizacja projektu ma rozpocząć się jeszcze w tym roku i zakończyć w 2025.

# Data Spaces Symposium

## AI factories and the data challenge

---

Juan Pelegrin

# Data Spaces unlocking AI Innovation

## AI Factories and the data challenge

Data Spaces Symposium 2025

Till Riedel (KIT)

Jeanette Nilsson (RISE)

Roberta Turra (Cineca)

Daniel Alonso (BDVA)

DSBA



BDV

BIG DATA VALUE ASSOCIATION

FIWARE FOUNDATION

gaia-x



INTERNATIONAL DATA SPACES ASSOCIATION

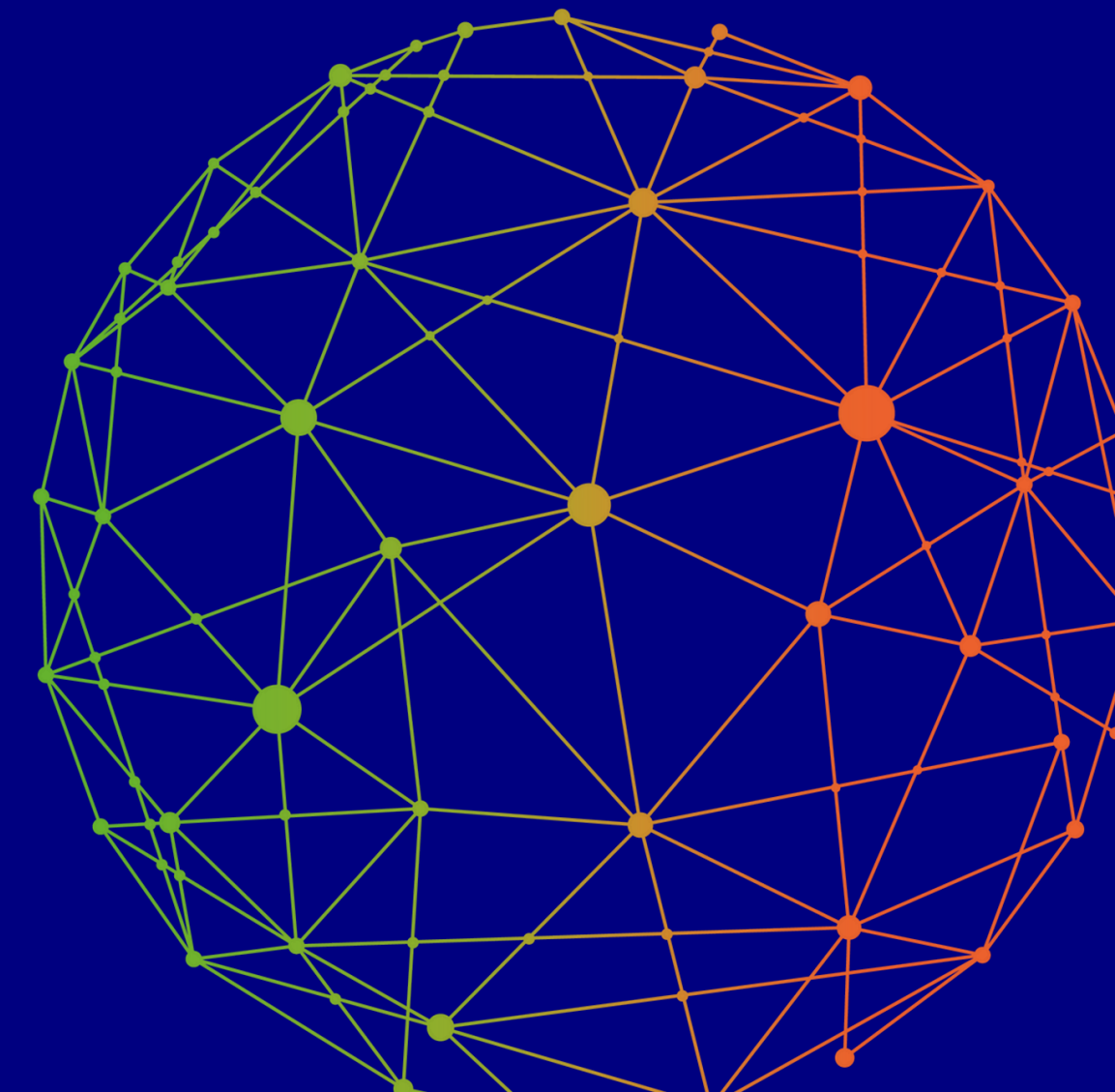


DATA SPACES SUPPORT CENTRE



Funded by the European Union

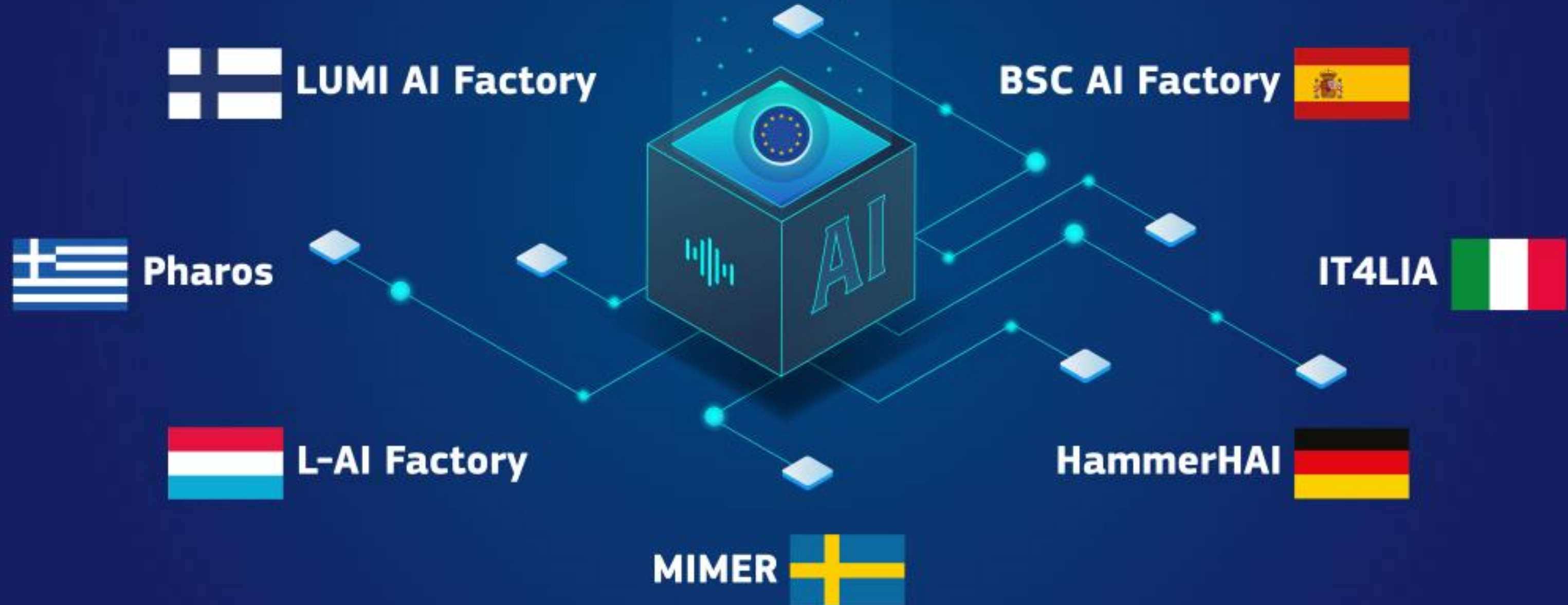
The Data Spaces Support Centre receives funding from the European Union Digital Europe Programme under grant agreement n° 101083412





**EuroHPC**  
Joint Undertaking

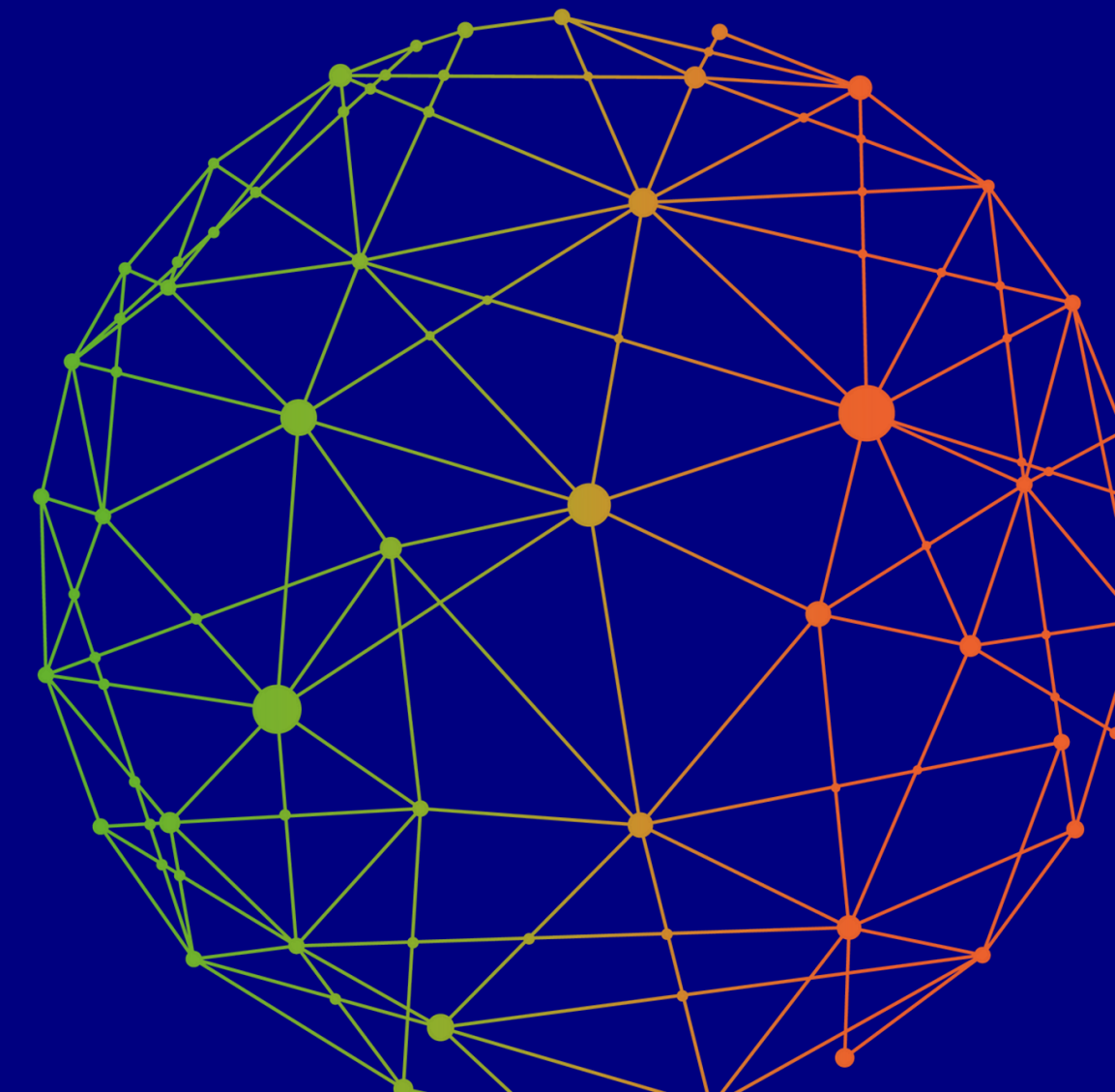
# 1st EuroHPC AI Factories



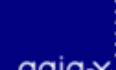
# MIMER - The Swedish AI Factory

## Data Spaces unlocking AI Innovation

Data Spaces Symposium 2025



DSBA



The Data Spaces Support Centre receives funding from the European Union Digital Europe Programme under grant agreement n° 101083412

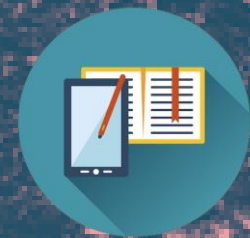
APPROACH

# Kick-start from ENCCS

CUSTOMER JOURNEY



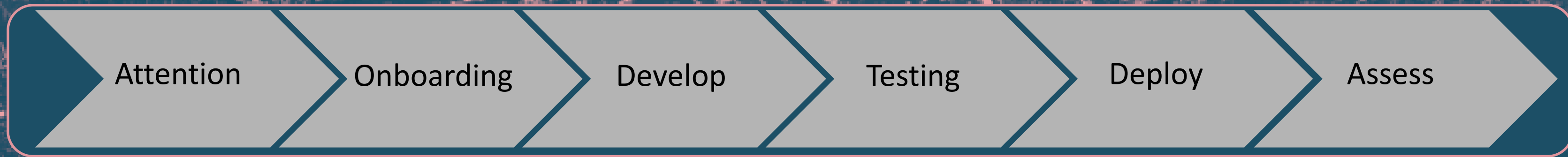
HPC access



Training



Consultancy



FOCUS AREAS

Gaming

Life Science

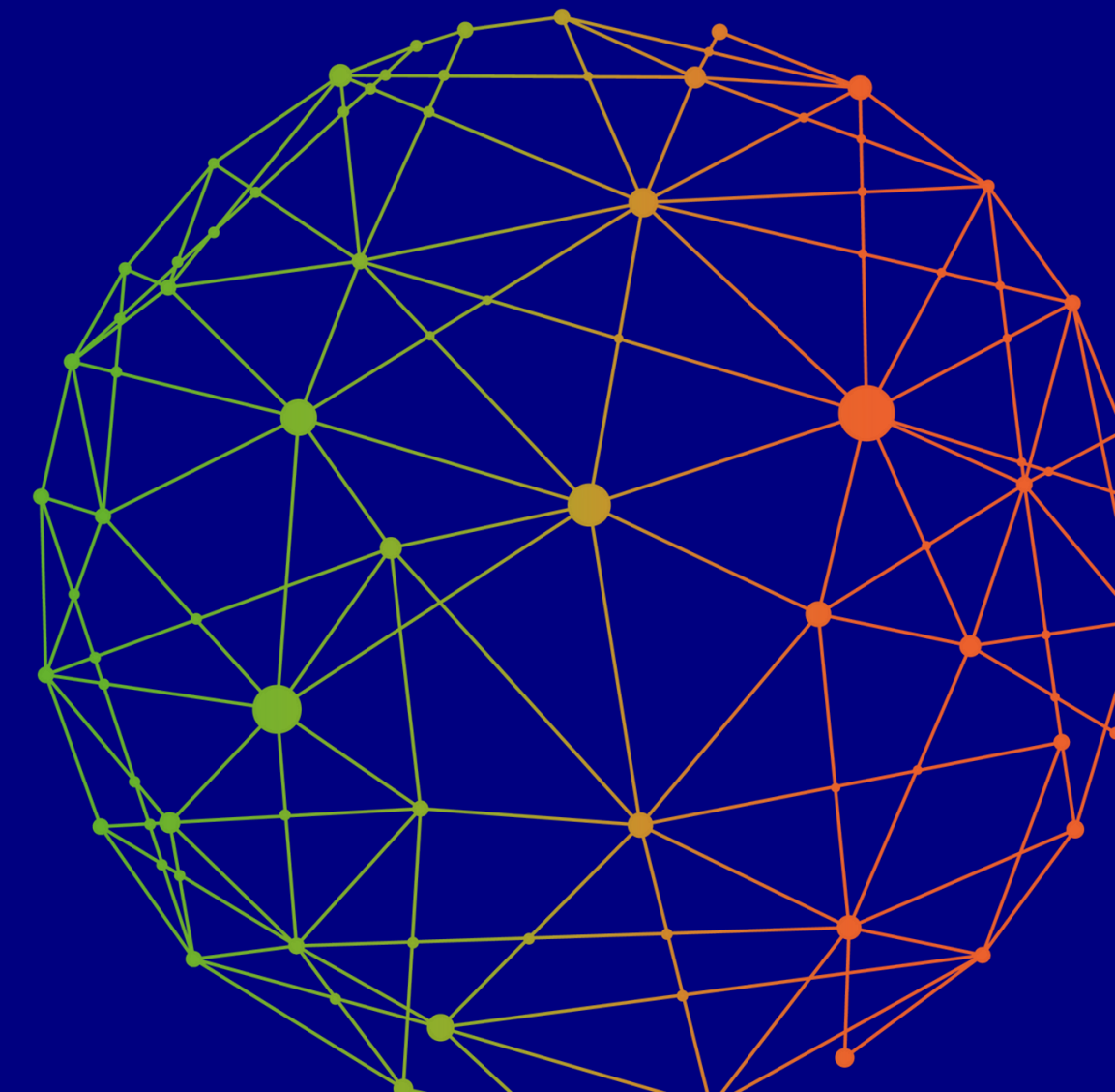
Autonomous systems

Material Sciences

# IT4LIA - The Italian AI Factory

## Data Spaces unlocking AI Innovation

Data Spaces Symposium 2025



DSBA



INTERNATIONAL DATA SPACES ASSOCIATION



DATA SPACES SUPPORT CENTRE



Funded by the European Union

The Data Spaces Support Centre receives funding from the European Union Digital Europe Programme under grant agreement n° 101083412



# The Italian AI Factory: IT4LIA

## AI FACTORY USERS

**AI for Science**  
*Academia, Research*

**AI for Innovation**  
*Startups, SMEs, Industries, Private  
Entities*

**AI for Public Good**  
*Public Administrations*

## AI GATES

## AI FACTORY SERVICES

**DATA-RELATED  
SERVICES**

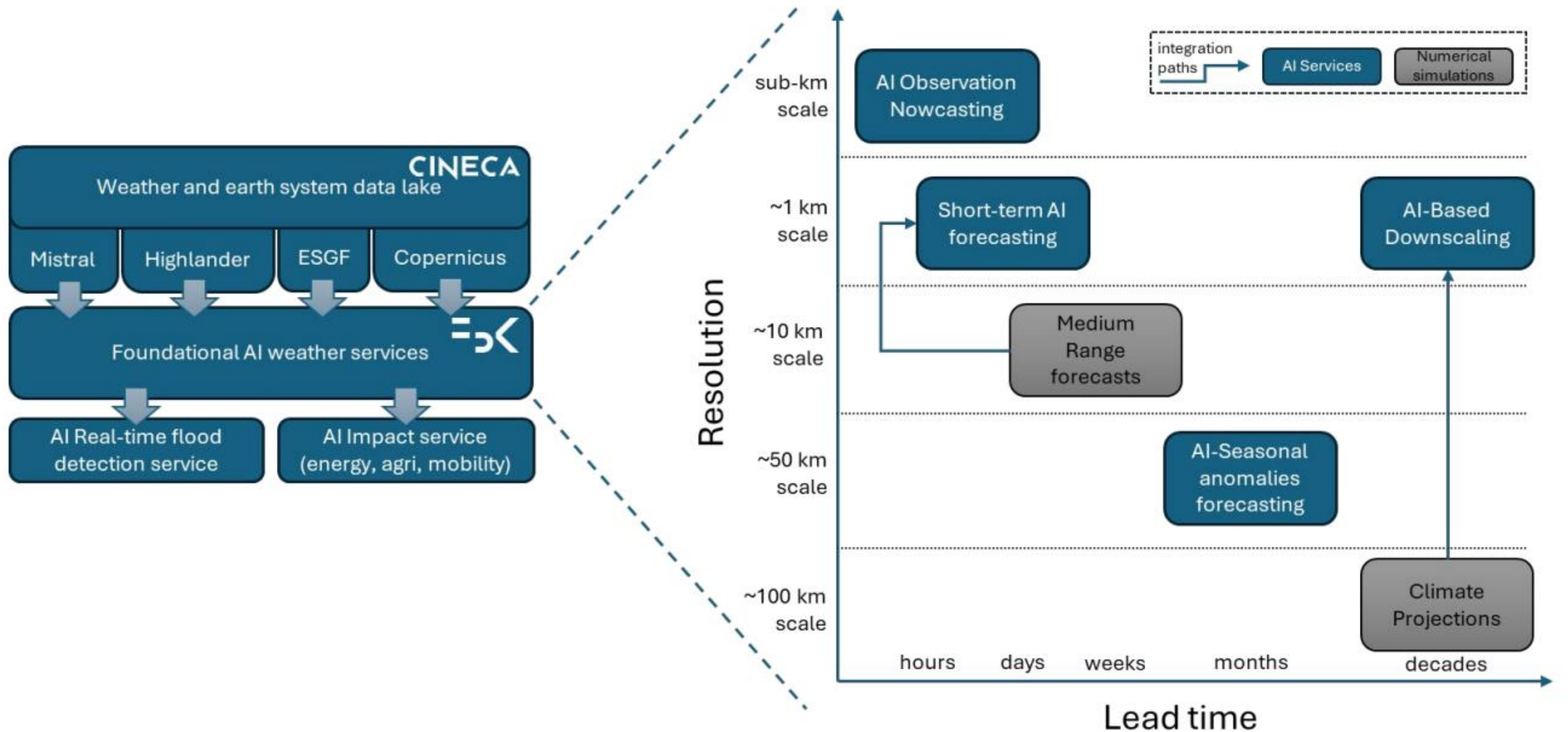
**HORIZONTAL SERVICES**

**VERTICAL SERVICES**  
Earth-Agrifood - Cyber  
-Manufacturing

**AI SKILLS**

## AI OPTIMIZED SUPERCOMPUTERS

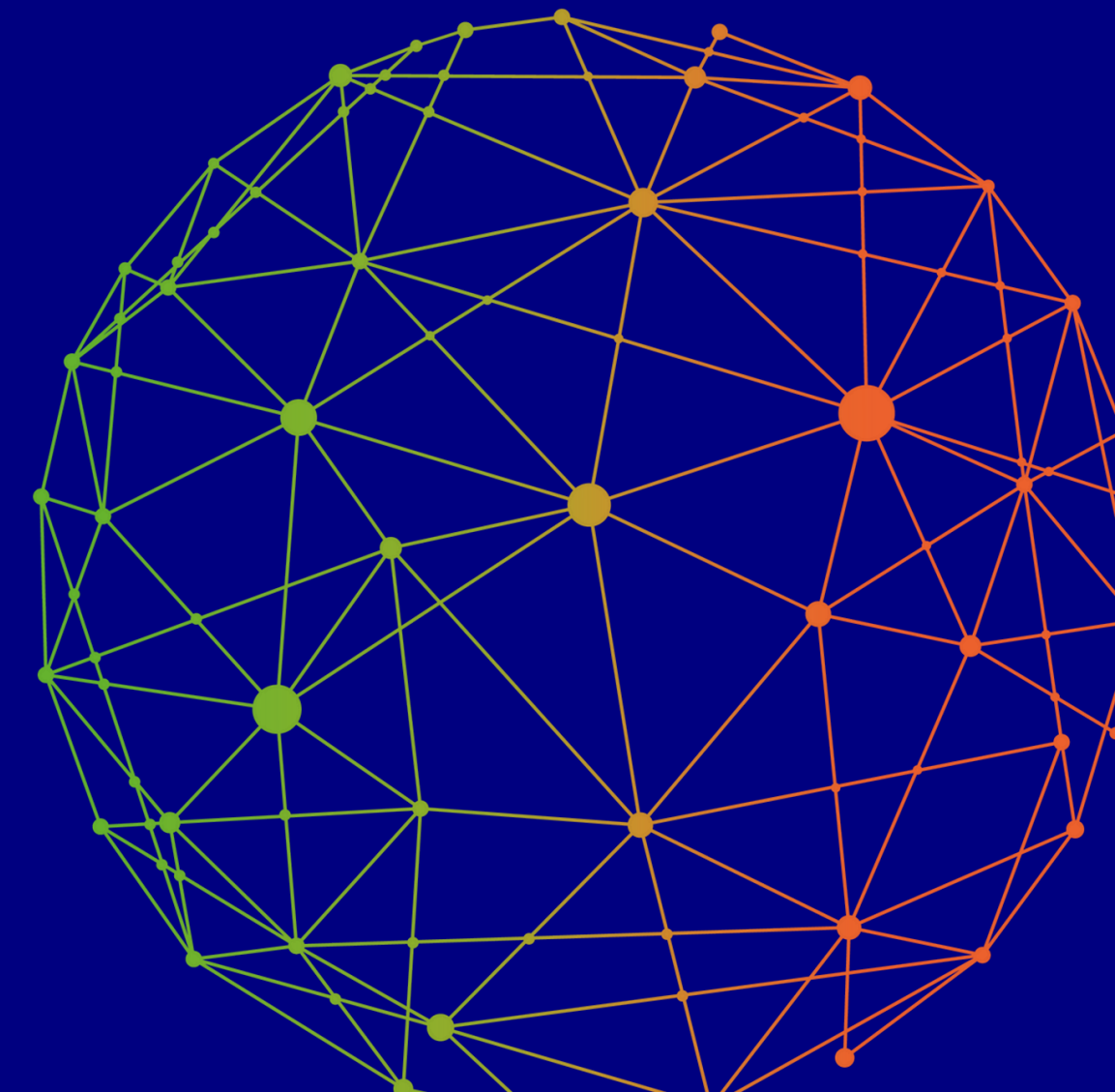
# The services of IT4LIA: example on Weather



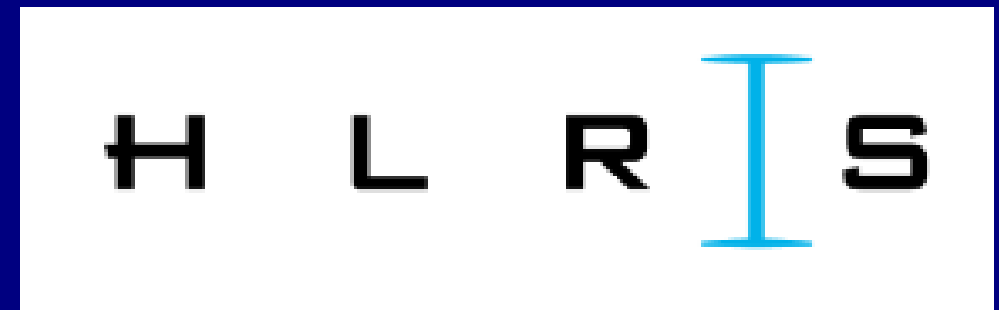
# HammerHAI - German AI Factory

## Data Spaces unlocking AI Innovation

Data Spaces Symposium 2025



The Data Spaces Support Centre receives funding from the European Union Digital Europe Programme under grant agreement n° 101083412



Coordinated and hosted by HLRS (Stuttgart)

## ■ Goals

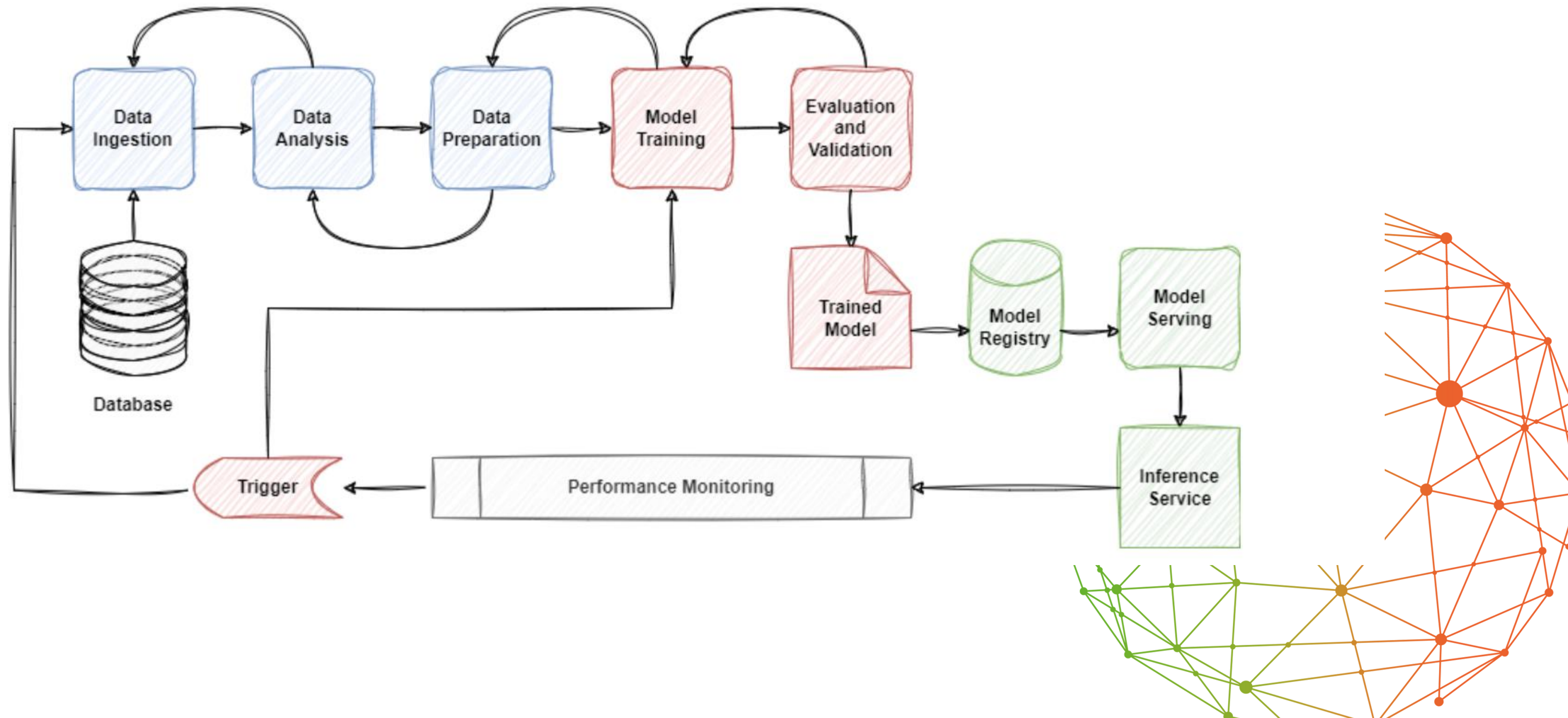
- Accelerate innovation and drive economic growth (focus: **manufacturing, engineering**)
- Support development of AI-based products and services

## ■ Key Features

- Large-scale, **AI-optimized** infrastructure
- **Secure**, local, and trustworthy platform
- **Cloud-like technologies** for easy migration and scaling
- Access to **workflow templates, pre-trained models, and datasets**



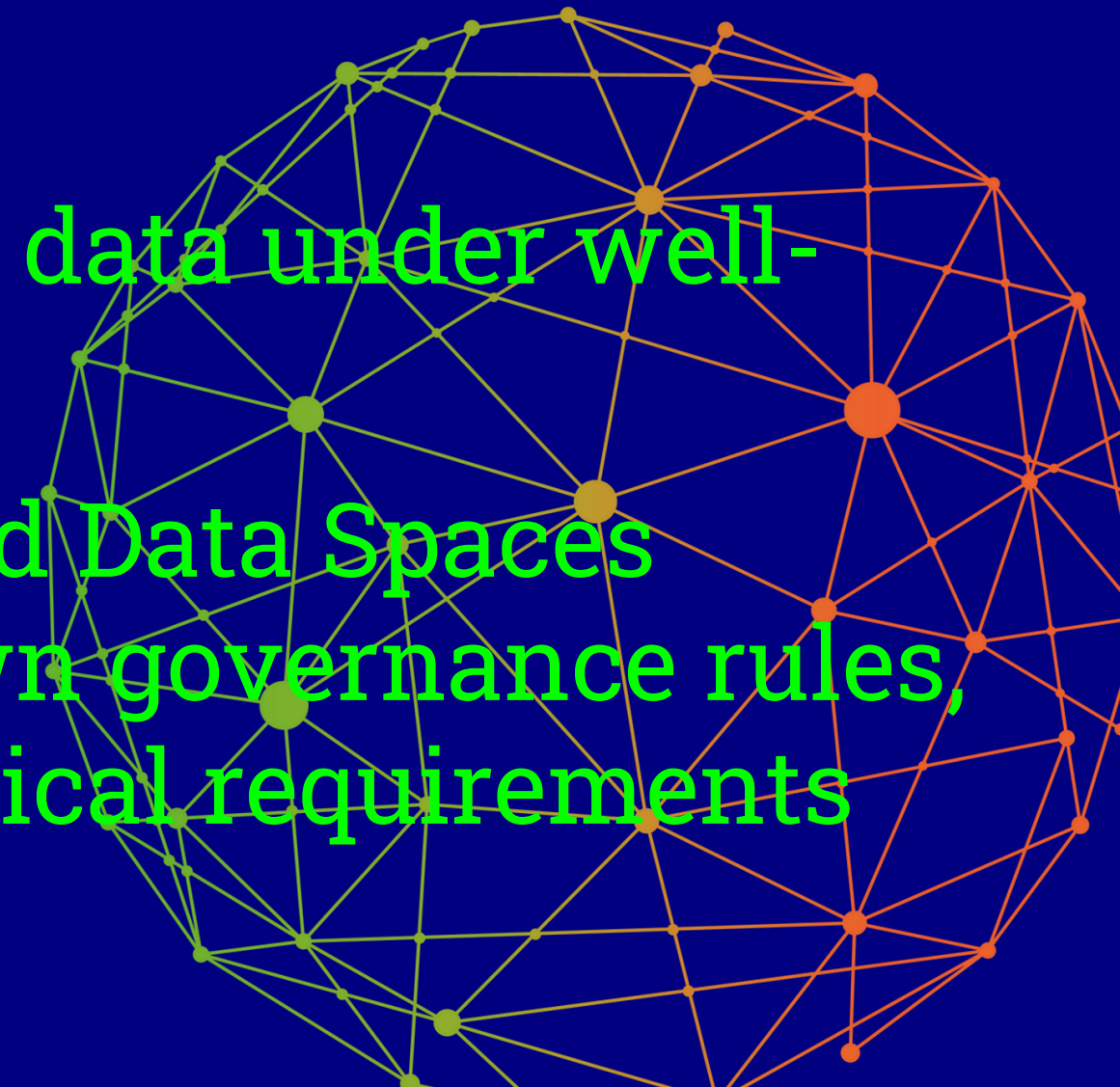
# Support for the entire AI Lifecycle on HPC



# AI Factories and the data challenge



- **AI Factories** -> strategic to position Europe at the forefront of AI innovation
- **Access to vast amounts of high-quality, purpose-specific data** essential to ensure the correctness, accuracy, and effectiveness of AI models
- **Data Spaces**: unlock access to private industrial data under well-defined conditions and usage policies
- **Not a plug and play connection**: AI Factories and Data Spaces operate as distinct ecosystems, each with its own governance rules, strategic objectives, business models, and technical requirements



# Turning the challenge into an opportunity!





Jeanette Nilsson  
RISE

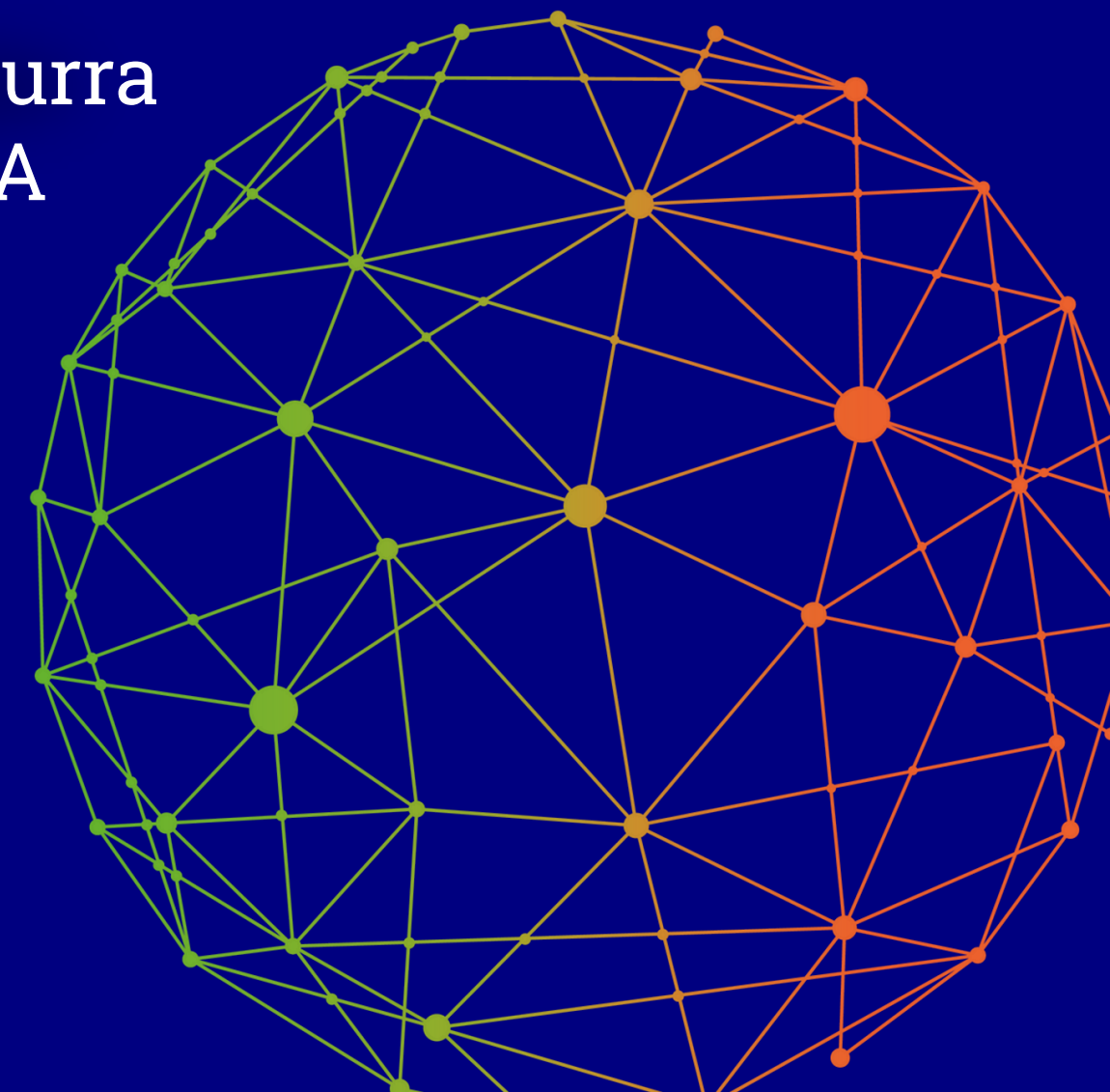


Till Riedel  
KIT



Roberta Turra  
CINECA

# Panel discussion



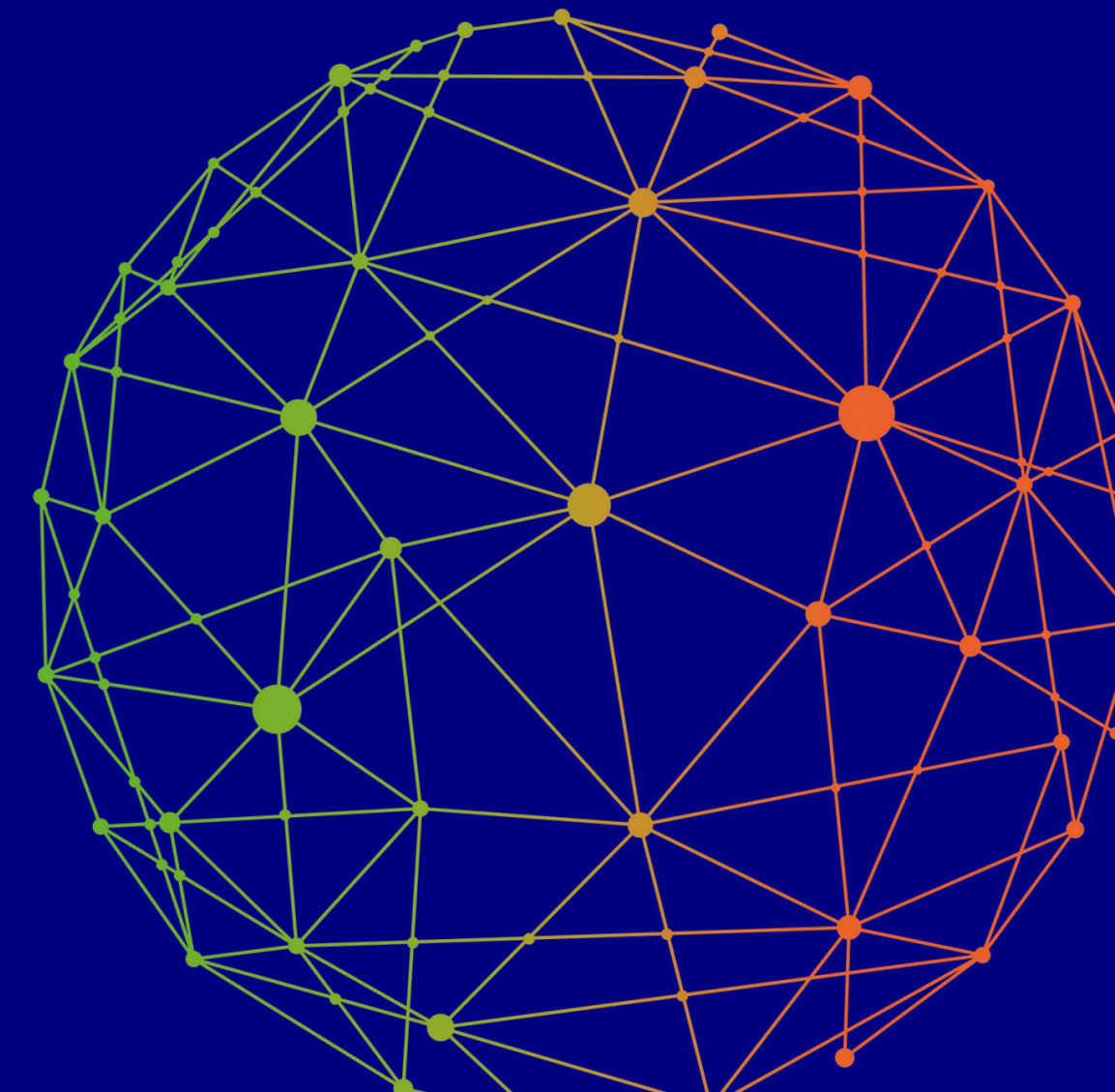


# Know the Unknown

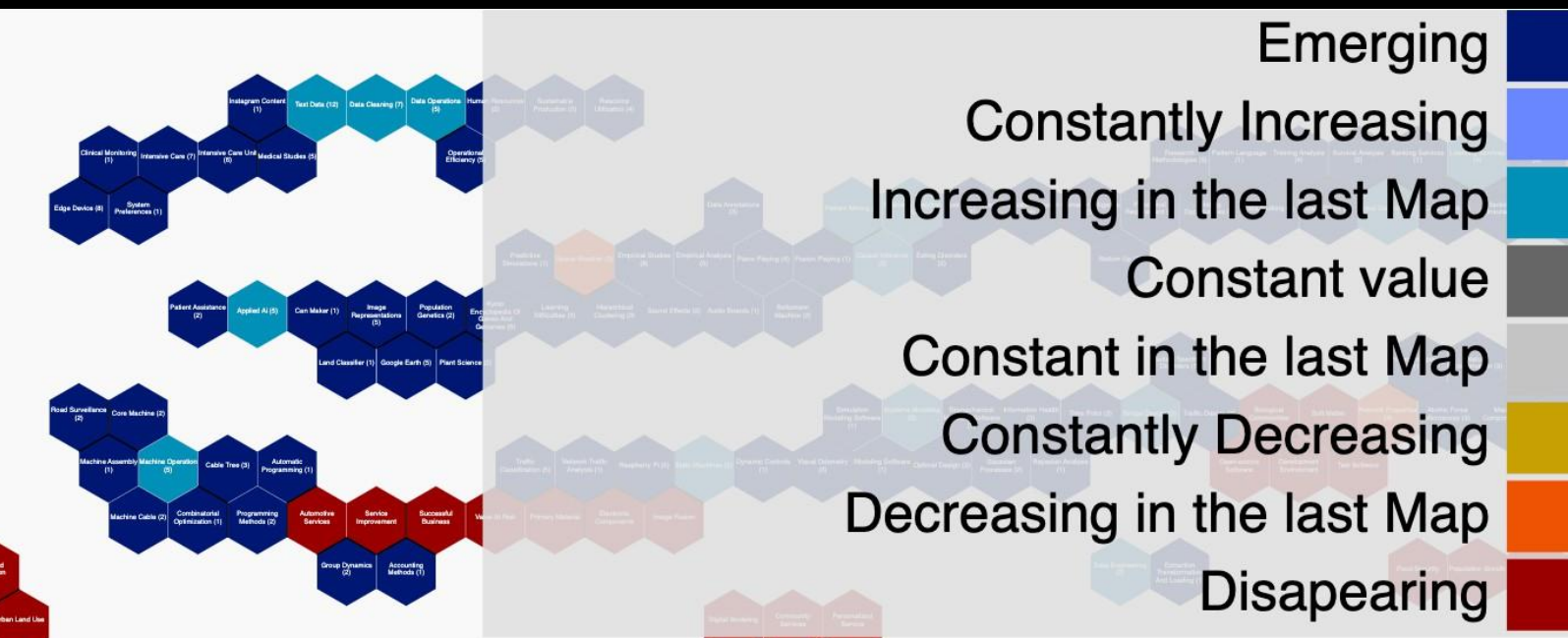
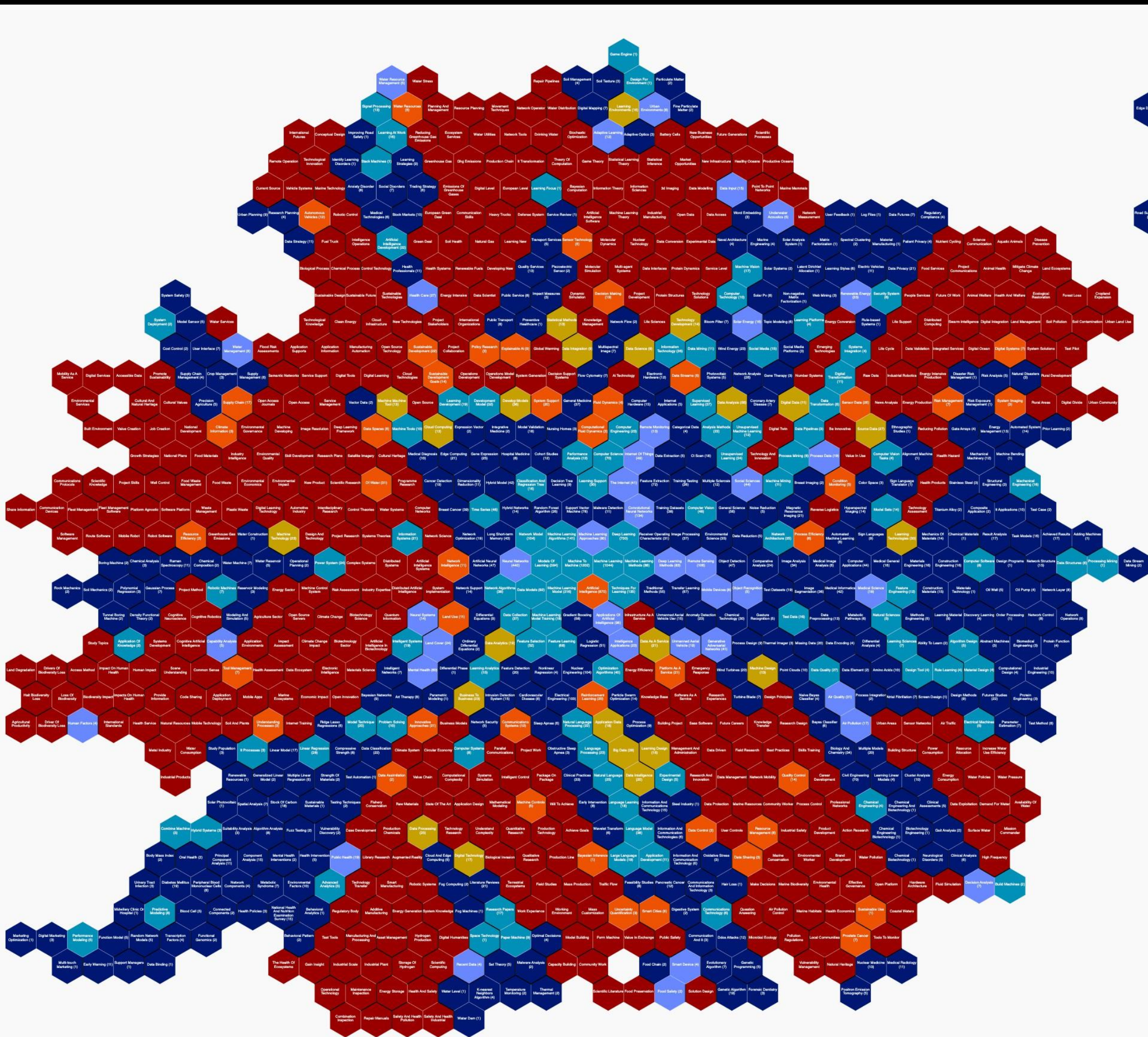
## AI innovation in Data Ecosystems

Data Spaces Symposium 2025

Harri Ketamo  
Founder, Headai

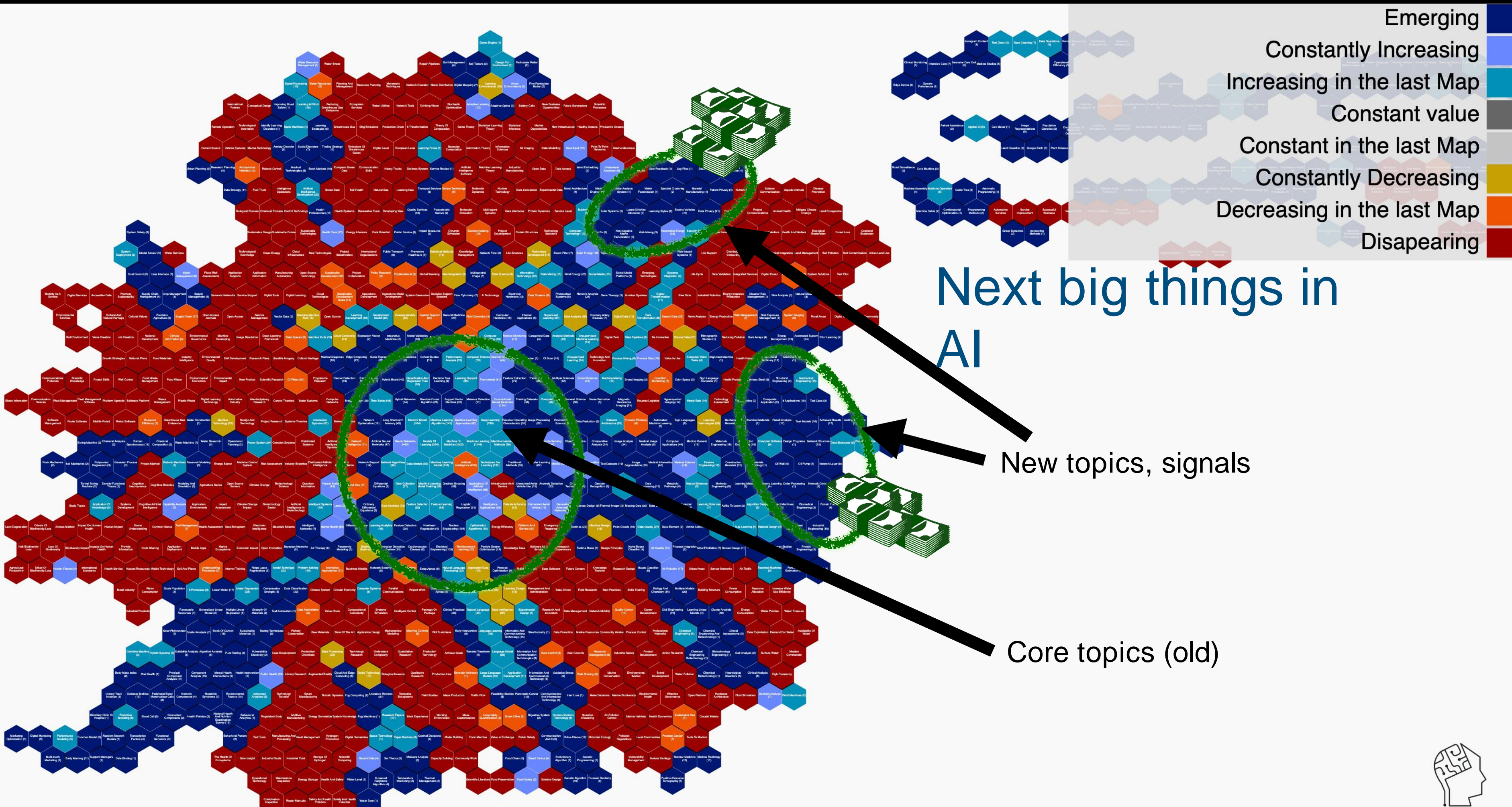


The Data Spaces Support Centre receives funding from the European Union Digital Europe Programme under grant agreement n° 101083412



Emerging  
 Constantly Increasing  
 Increasing in the last Map  
 Constant value  
 Constant in the last Map  
 Constantly Decreasing  
 Decreasing in the last Map  
 Disappearing

# Next big things in AI



Emerging

Constantly Increasing

Increasing in the last Map

Constant value

Constant in the last Map

Constantly Decreasing

Decreasing in the last Map

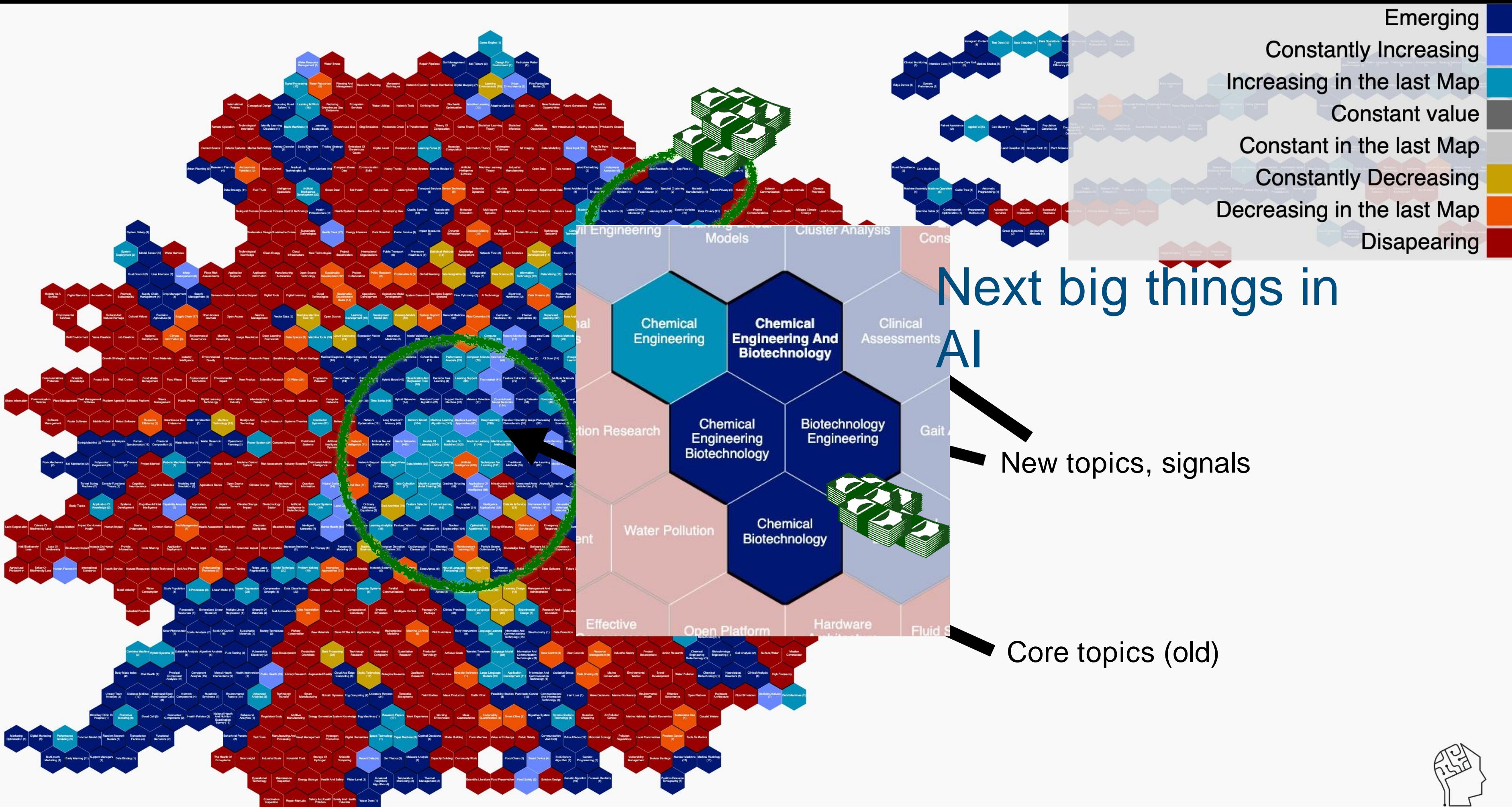
Disappearing

Next big things in AI

New topics, signals

Core topics (old)





Emerging

Constantly Increasing

Increasing in the last Map

Constant value

Constant in the last Map

Constantly Decreasing

Decreasing in the last Map

Disappearing

# Next big things in AI

New topics, signals

Core topics (old)



# The future of European competitiveness

Part A | A competitiveness strategy for Europe

SEPTEMBER 2024

arisa

## AI Skills Needs Analysis

AN INSIGHT INTO THE AI ROLES AND SKILLS NEEDED FOR EUROPE

31 May 2023

Co-funded by the European Union

THE WORLD BANK

## Philippines Economic Update

JUNE 2022 EDITION

### Strengthening the Digital Economy to Boost Domestic Recovery

LEADSx2030  
ADVANCED DIGITAL SKILLS

Home About Us LEADS 2022-2024 ADS Cluster Insights News Events Contact us

## DRIVING EUROPE'S DIGITAL COMPETITIVENESS

DS4Skills  
Data Space for Skills

ABOUT US NEWS & EVENTS RESOURCES CONTACT US TOOLS Stay Updated

### Building a human-centric and trusted ecosystem to share and access skills data

We prepare the ground for the future deployment of the first-time-ever European Data Space for Skills that will offer practical benefits to organisations and individuals in the digital age.

Discover More

37

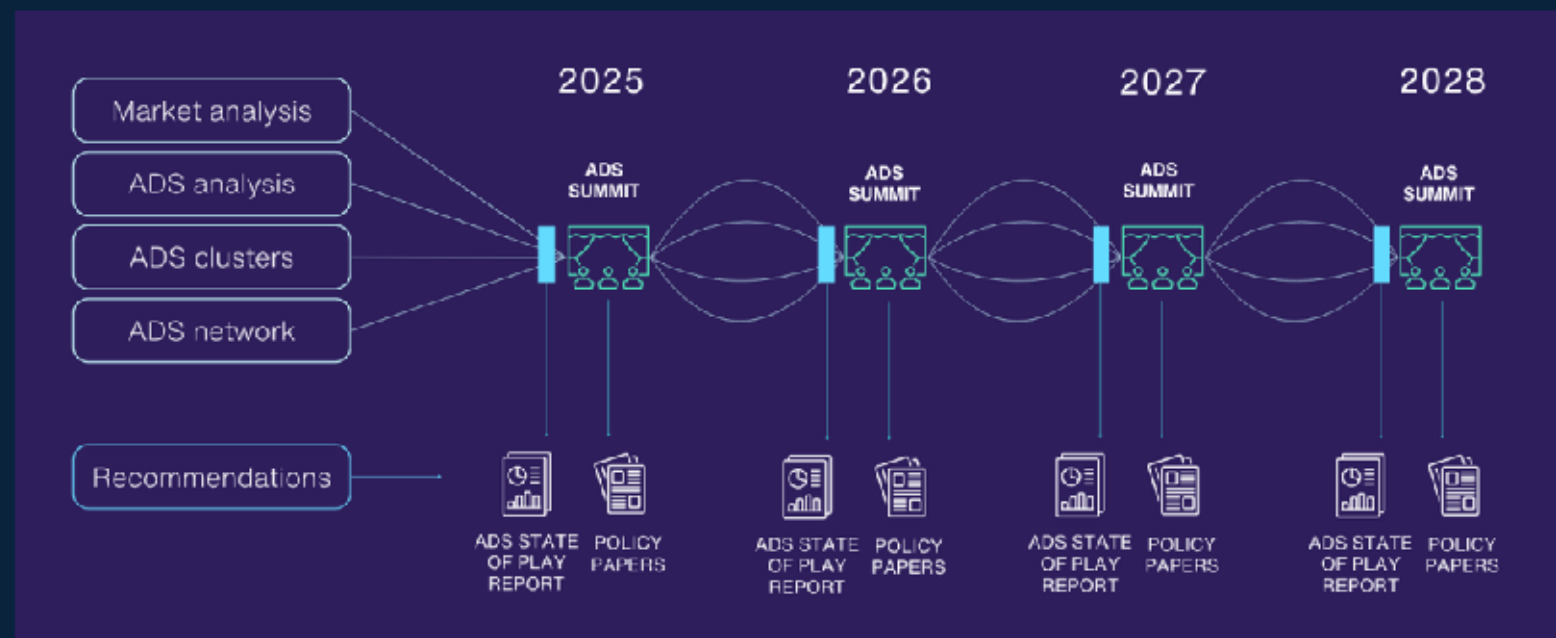
Funded initiatives

26

EU Member States

420

Participating organisations



# Semantic Layers for Dataspaces

## AI innovation in Data Ecosystems



Data Spaces Symposium 2025

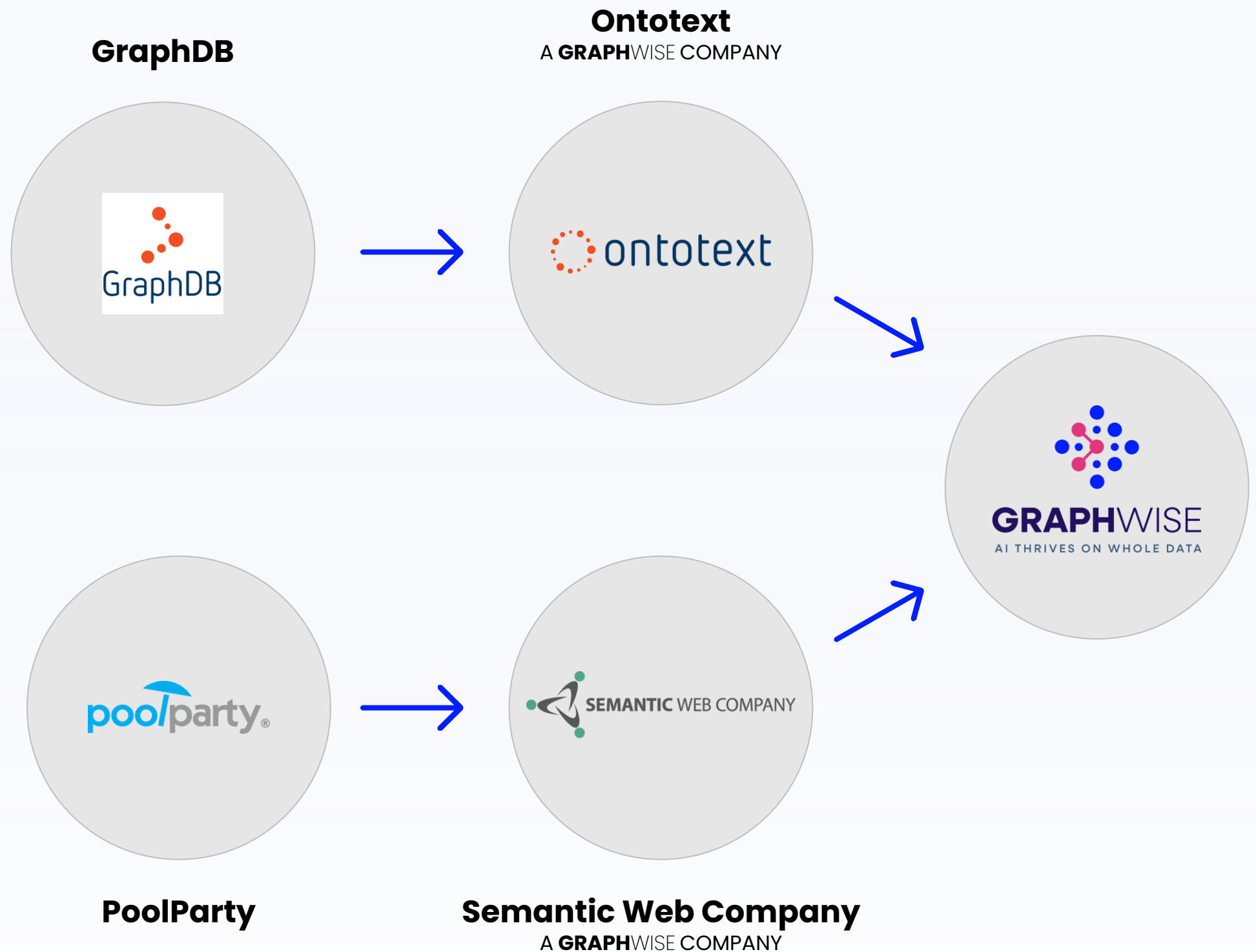


The Data Spaces Support Centre receives funding from the European Union Digital Europe Programme under grant agreement n° 101083412

# One Stop Shop for Graph AI

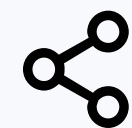


Graphwise has the platform and tools, the methodologies, and the people to deliver world-class end-to-end solutions to enterprises.





# Graphwise in a Nutshell



The most advanced  
**Graph AI platform**



**200**  
**employees**

200 employees across North  
America, Europe and APAC



**Main offices in  
Vienna, Sofia and  
New York**



**200+**  
**customers**

More than 200  
customers combined



**50%**  
**of our customers**

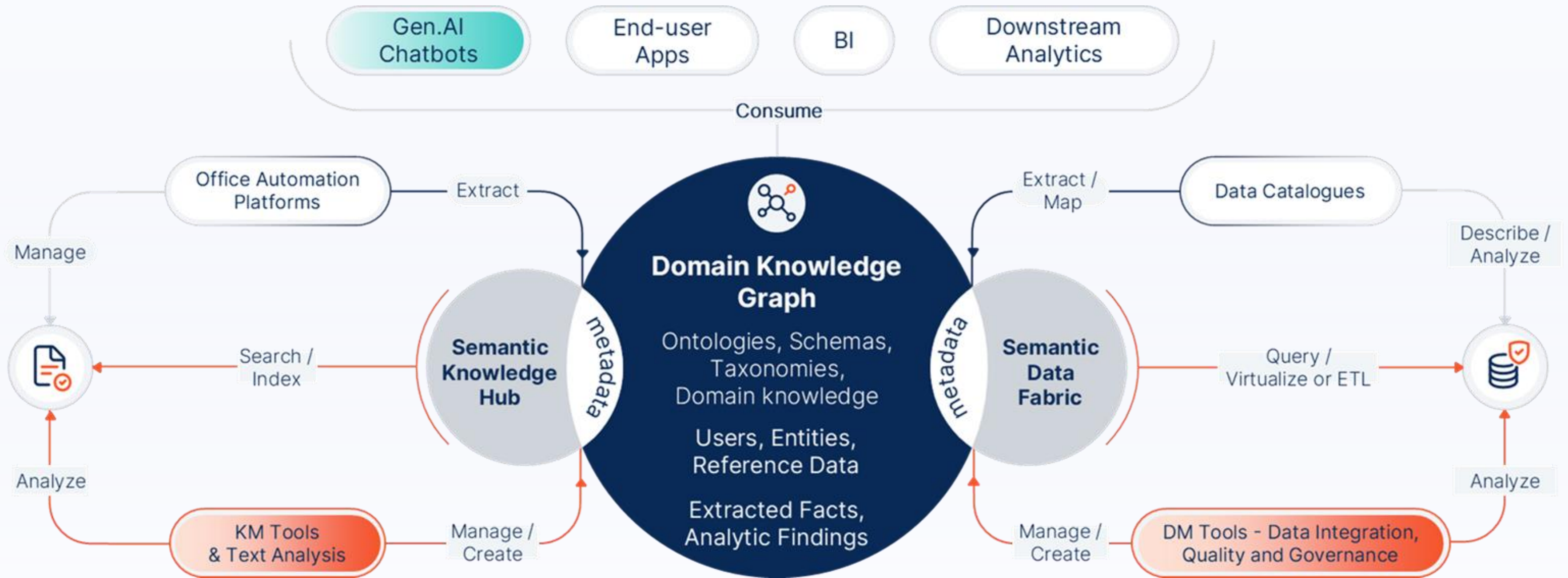
have already joint  
deployments



**30+**  
**partners**

Rich ecosystem  
of 30+ partners

# Semantic Layer = Content Hub + Data Fabric



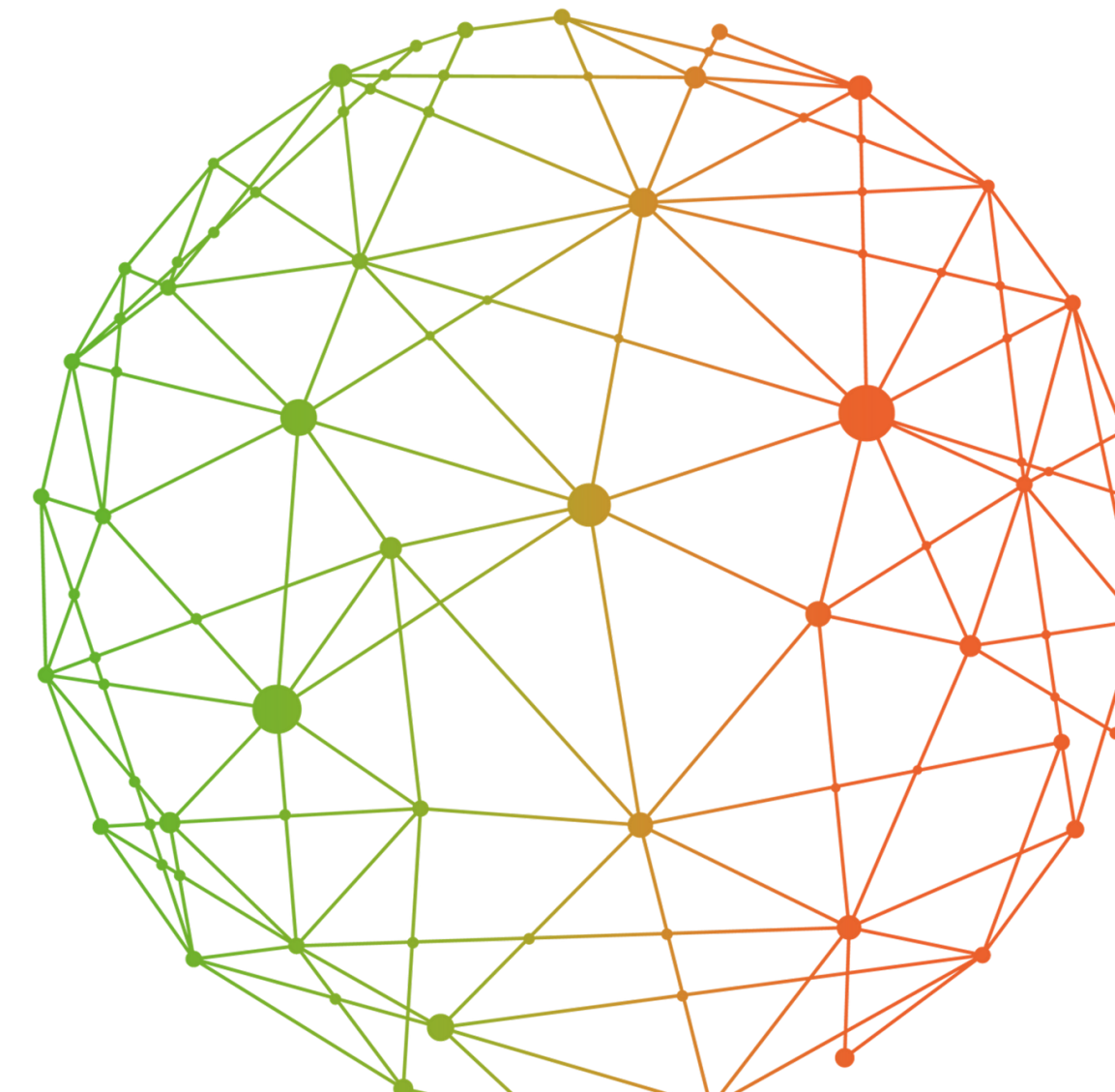
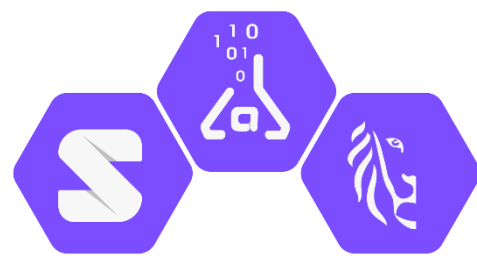
# Semantic Interoperability in Dataspaces

- Interoperability is **important**:
  - Great value for dataspace **consumers**: harmonization of dataset form and meaning
  - Precondition to dataspace **federation**: makes little sense to federate data without harmonization
- Interoperability is **elusive**:
  - Previously we wrote about the need for semantic harmonization of business (payload) data
  - But it turns out that semantic metadata in dataspaces also has numerous problems
- In the **UNDERPIN project** we use in-depth metadata (not only datasets but also columns) to enable
  - Dataset **discovery** (semantic faceted search)
  - **Automatic ingest** to Influx time-series database
  - More flexible **predictive analytics** apps
  - More flexible **data dashboard/visualization**
- References
  - Semantic Problems in Dataspaces. Vladimir Alexiev. AIOTI Workshop on Semantic Interoperability for Digital Twins, Sophia Antipolis, France, February 2025. [paper](#), [presentation](#)
  - Semantic Representation Challenges in AAS, ECLASS and SAMM. Vladimir Alexiev, Ken Wenzel, Hossein Rimaz. AIOTI Workshop on Semantic Interoperability for Digital Twins, Sophia Antipolis, France, February 2025. [paper](#), [presentation](#)
  - Raising the Role of Vocabulary Hubs for Semantic Data Interoperability in Dataspaces. David, R.; Ivanov, P.; and Alexiev, V. In Third workshop on Semantic Interoperability in Data Spaces, Budapest, Hungary, October 2024. [proceedings](#), [slides](#), [pdf](#)
  - Semantic Interoperability for Data Spaces. Alexiev, V. In SEMIC: Data Spaces in an Interoperable Europe, December 2022. [presentation](#)
  - Data Spaces vs Knowledge Graphs: How to Get To Semantic Data Spaces? Alexiev, V. In Data Spaces & Semantic Interoperability Workshop, Vienna, Austria, July 2022. [Paper](#), [slides](#), [video](#), [blog](#)

# Personal data sharing in the agentic age

## Tanguy Coenen - imec

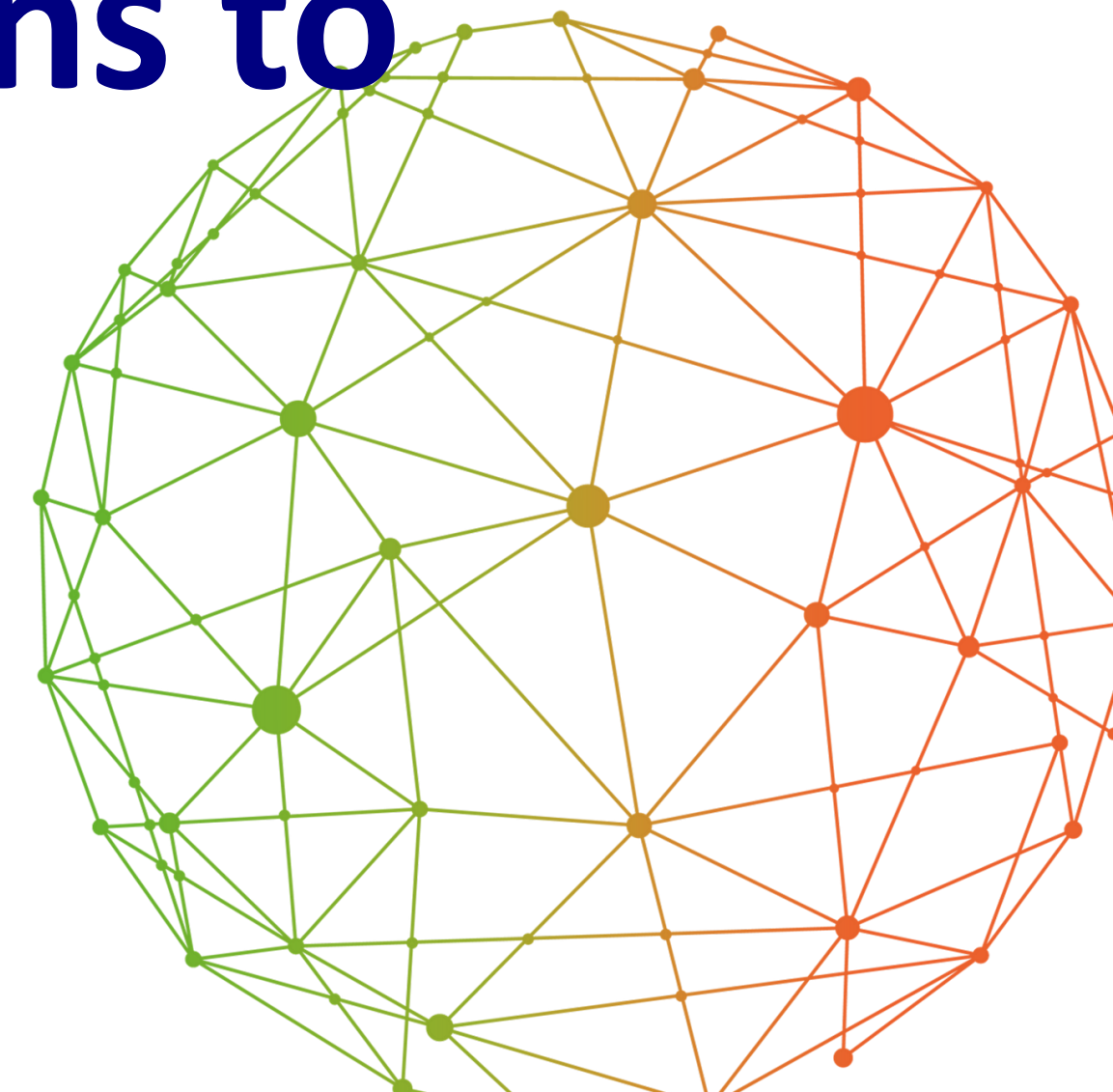
Data Spaces Symposium 2025



# The age of (personal) agents is coming



***An agent* refers to an autonomous system that can independently perceive its environment, make decisions, and execute actions to achieve specific goals.**

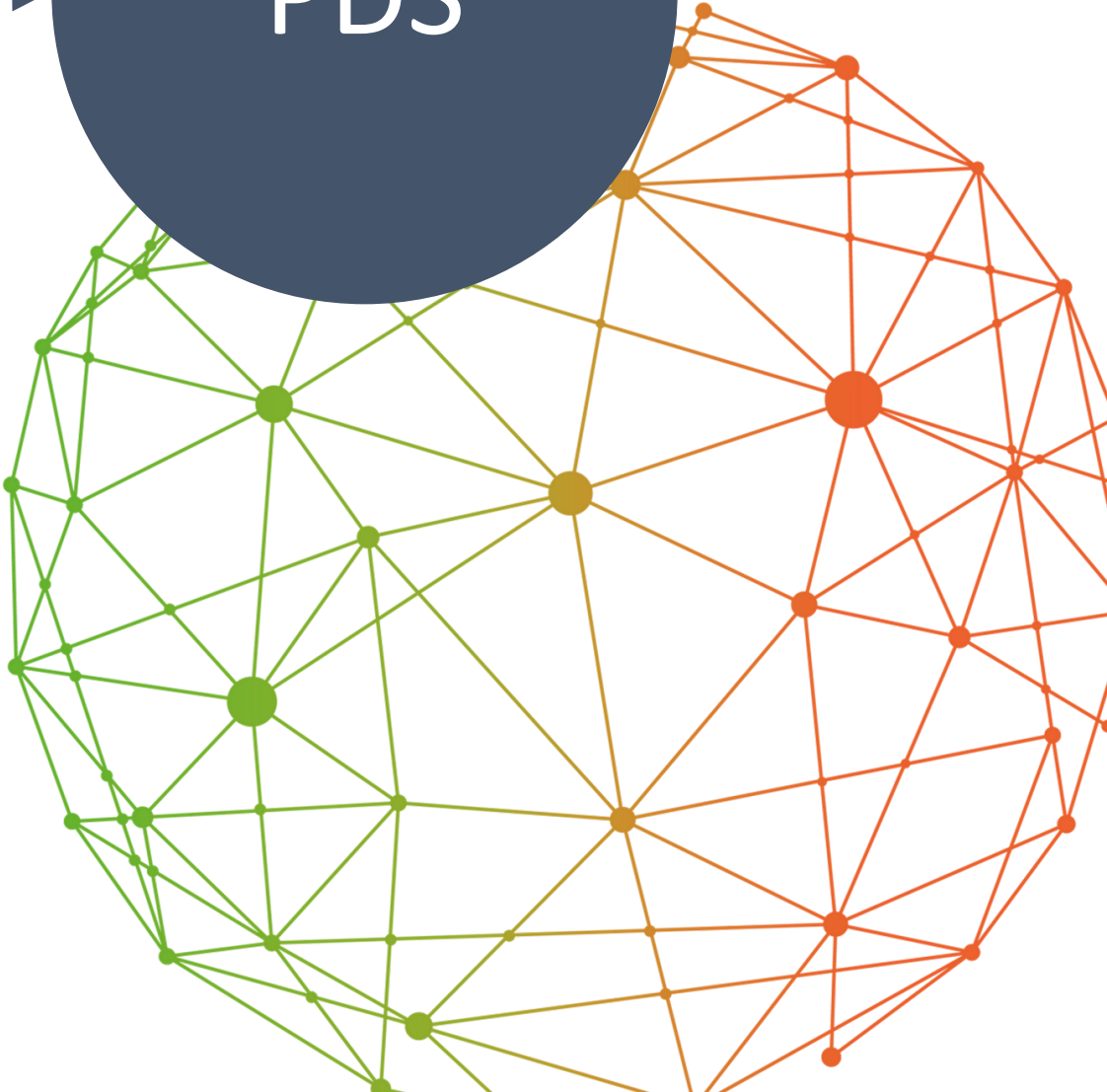


# Problem statement

- Increasing AI hardware in consumer devices
- Increasing rise of “agents”, that support the end-user in various tasks
- Increasing push towards EU sovereignty
- Increasing complexity in the EU regulatory landscape

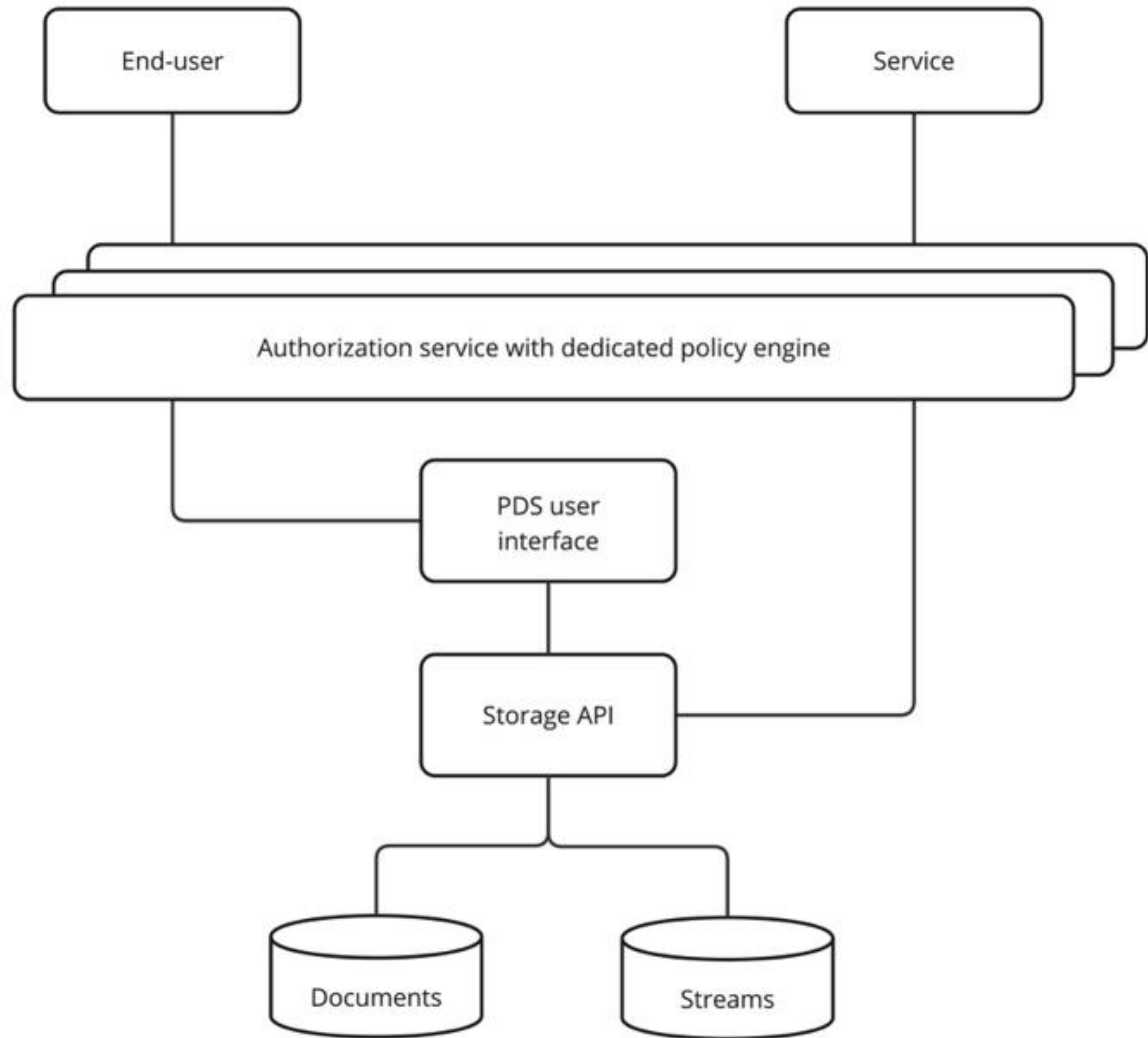
=> How to develop a single, integrated **personal data store** that fulfils these **regulatory and ethical obligations**, guarantees **user trust and data ownership**, and leverages **(edge) AI** to deliver meaningful, personalized services without compromising **privacy or security**?



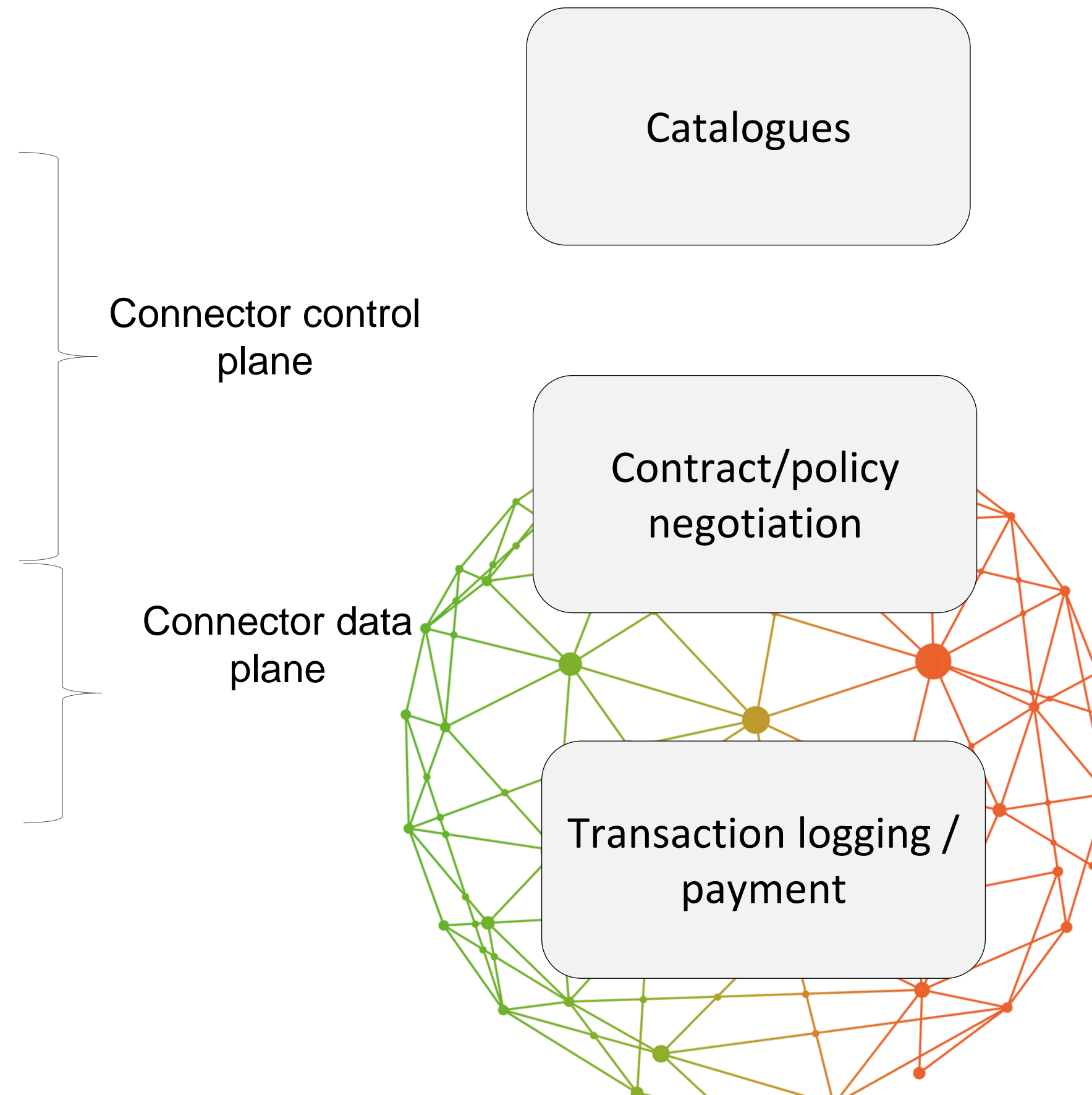
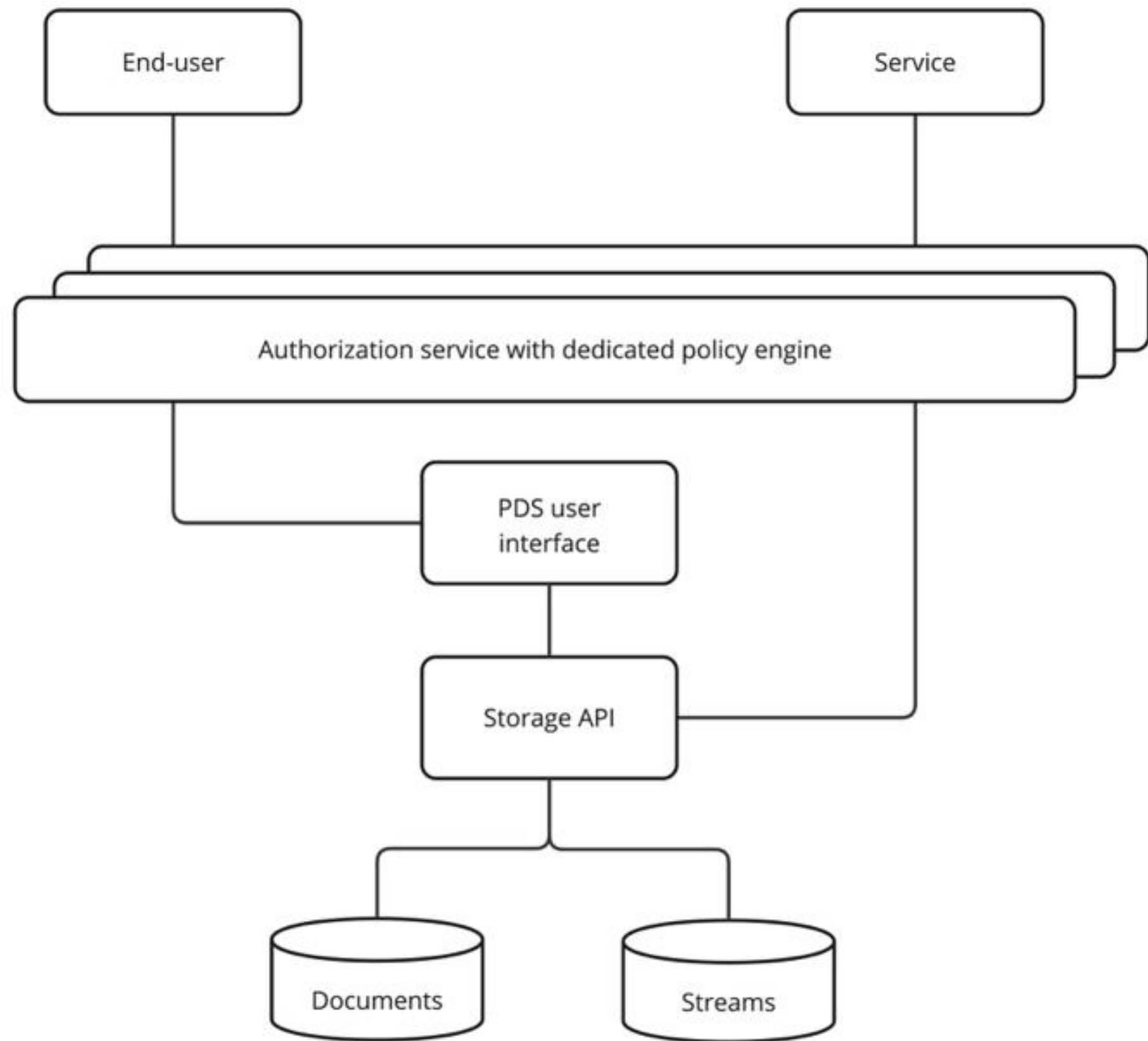




# Architecture



# Data space compatibility



# THE AGENTIC HOME IOT PROTOTYPE



Result

User

User prompt

LLM agent
Policies
Vector DB

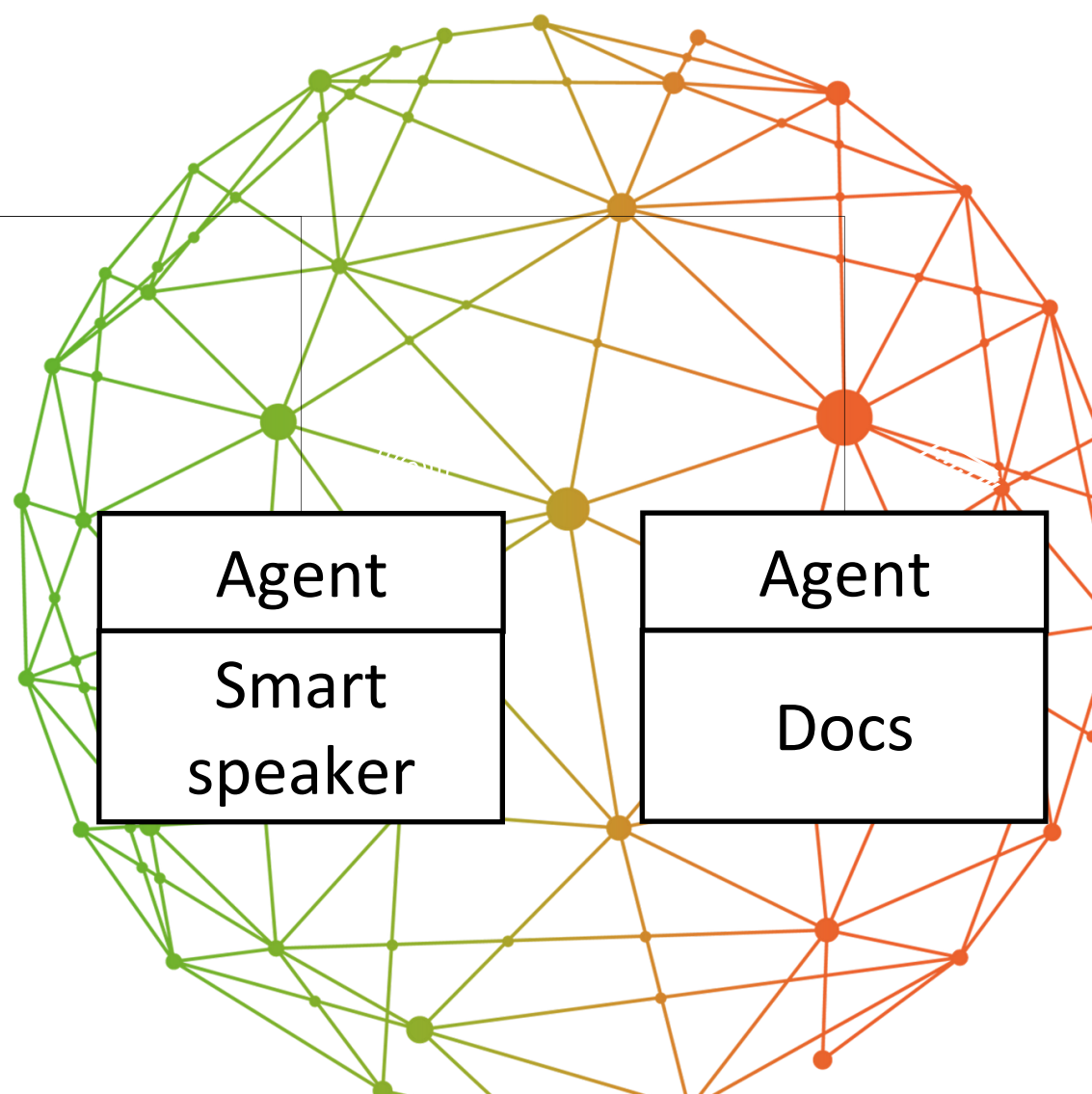
Agent
Fridge

Agent
Car battery

Agent
Dish-washer

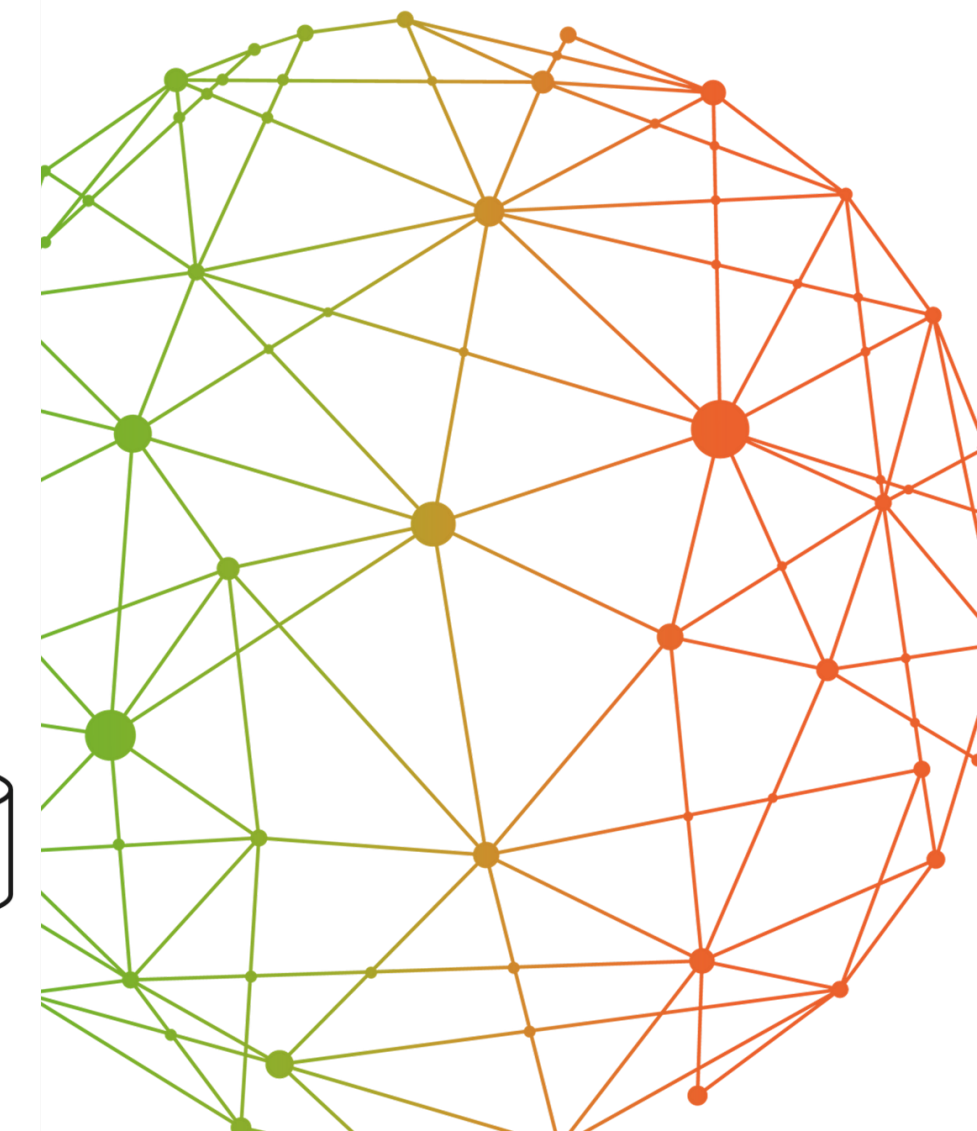
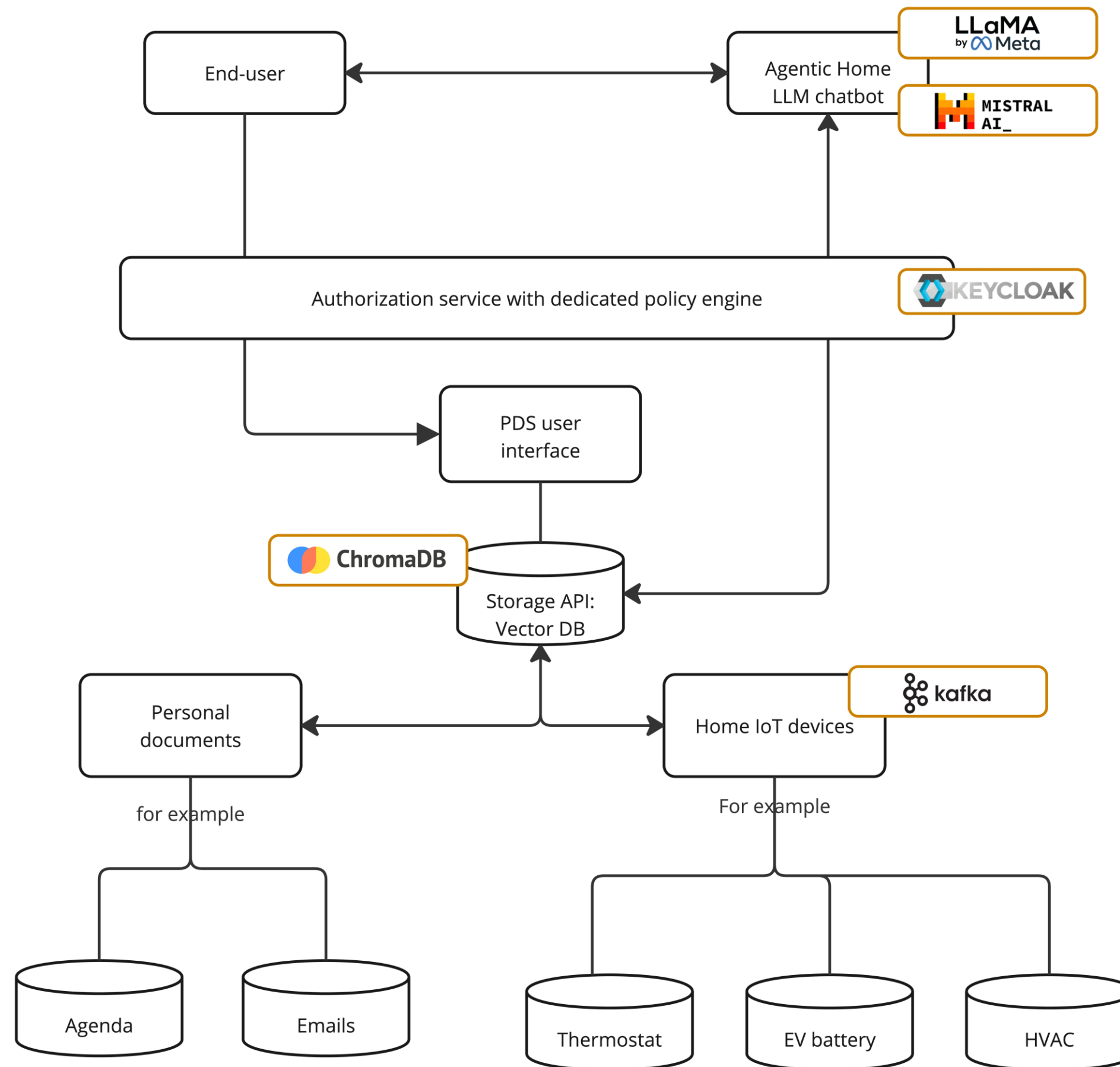
Agent
Solar panels

Agent
HVAC



Agent
Smart speaker

Agent
Docs





# Stan

the personal home assistant

Adam Grant



Hi Stan, it's starting to get chilly outside, what's the temperature in our kitchen like?



The current temperature in your kitchen is 20.4 degrees Celsius.

Can I help you with something related to the temperature, such as adjusting the thermostat or finding a cozy spot in the house?

Access to sensor data

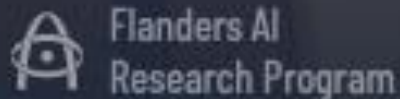
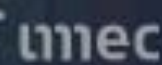
Yes, adjust the thermostat to 22 degrees please



I set the setpoint in the living room to 22 degrees.

Actuation through method calling

Ask me about your personal data, smart home devices, or anything else



SolidLab Flanders



Universiteit Antwerpen



IDLab  
INTERNET & DATA LAB

It sounds like you're asking about a specific person's financial information. I can help you with that.

Check Blake financial info, how much debt is he in?



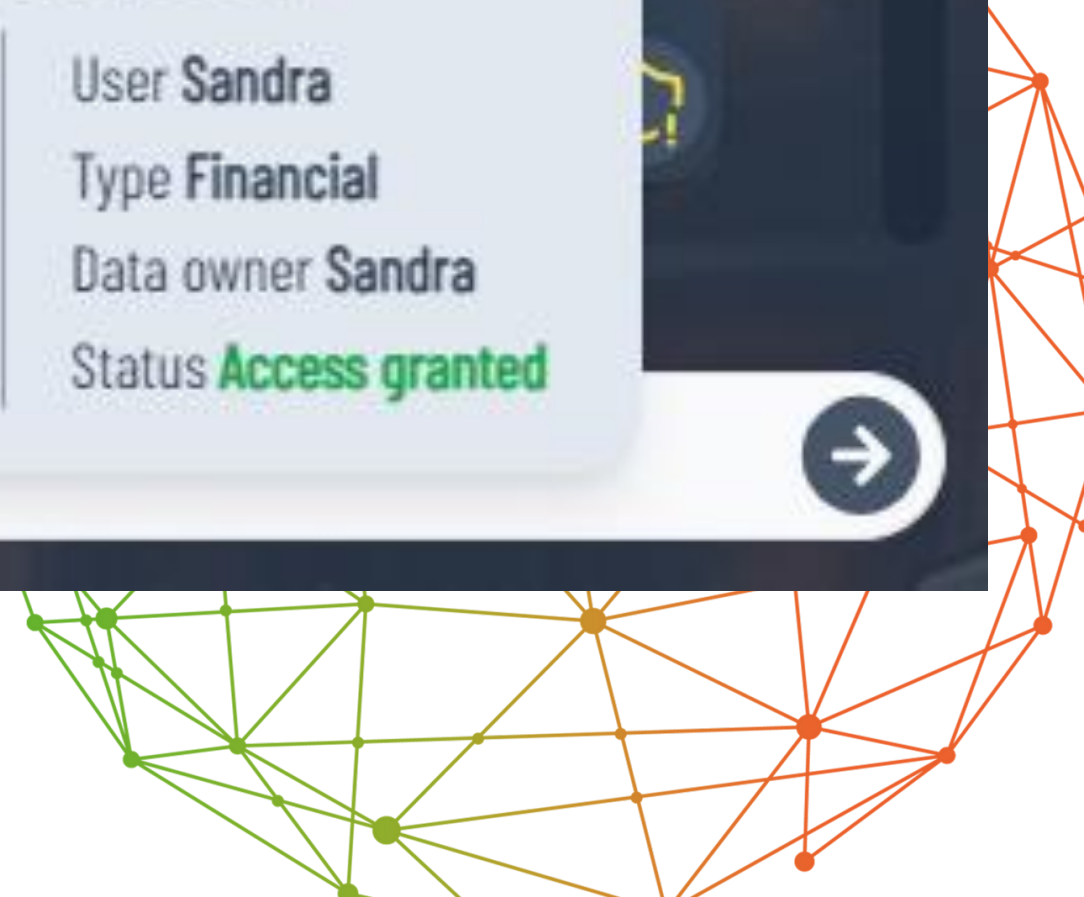
Sorry, I was unable to get the required information to answer the question. I can help you with something related, like checking Blake's calendar or schedule if you'd like.

#### PERSONAL DATA VAULT ACCESS

User **Sandra**  
Type **Financial**  
Data owner **Blake**  
Status **Access denied**

User **Sandra**  
Type **Financial**  
Data owner **Sandra**  
Status **Access granted**

Ask me about your personal data, smart home devices, or anything else



Browser tabs: kvasir / kvasir-server · GitLab

Address bar: gitlab.ilabt.imec.be/kvasir/kvasir-server

Header: IDLab Explore Sign in

Search: Search or go to...

Project: kvasir / kvasir-server

Project navigation: Manage, Plan, Code, Build, Deploy, Operate, Monitor, Analyze

Repository: kvasir-server (2 stars)

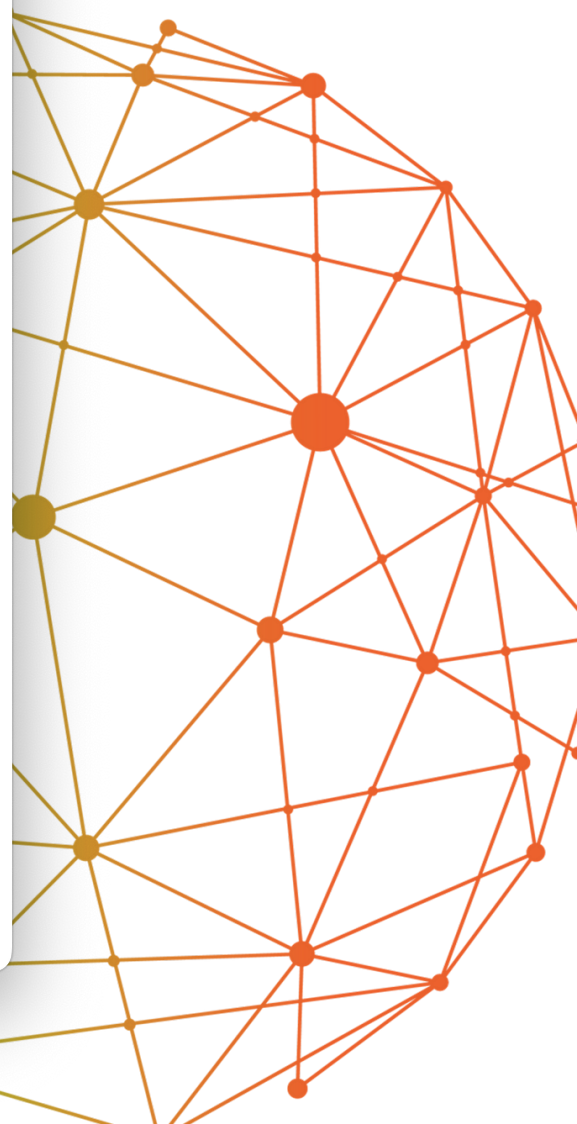
Current commit: fix: Fixed provided context not being used for querying (always...)  
Wannes Kerckhove authored 19 hours ago  
057ab868

Name	Last commit	Last update
.deployment/docke...	chore: use port 8081 for the...	1 week ago
.mvn/wrapper	chore: added maven wrapper	1 week ago
Writerside	build: monolith module can ...	1 week ago
libs	fix: Fixed provided context ...	19 hours ago
plugins	feat: added support for filte...	2 days ago
services	fix: Fixed provided context ...	19 hours ago
.gitignore	chore: updated .gitignore	1 week ago
.gitlab-ci.yml	ci: remove static set ALGOLI...	1 week ago
LICENSE	chore: added license	2 weeks ago
README.MD	build: monolith module can ...	1 week ago
SWAPI-WD-data.ttl	feat: ChangeRequests can n...	2 months ago
docker-compose.yml	fix: change kvasir-ui port in ...	1 week ago

Project information:

- 161 Commits
- 1 Branch
- 10 Tags
- README
- Apache License 2.0
- GitLab Pages

Created on: November 19, 2024



imec-int/solid-rag-toolbox

github.com/imec-int/solid-rag-toolbox

imec-int / solid-rag-toolbox

Code Issues Pull requests Actions Projects Security Insights

Unwatch 5 Fork 0 Star 0

main 1 Branch 0 Tags

Go to file Add file Code

taeymma Merge pull request #7 from imec-int/feat/update-readme 1cc54ba · 3 days ago 28 Commits

documentation	feat: update docusaurus + remove blog	3 days ago
keycloak-config	fix: small port fix chunkapi	last week
psx-chunk-api	chore: remove commented gitignore files	last week
psx-keycloak-auth-middleware	chore: remove commented gitignore files	last week
psx-vector-api	chore: remove commented gitignore files	last week
.gitignore	Initial commit	2 months ago
LICENSE	feat: add license	last week
README.md	chore: cleanup readme	3 days ago
docker-compose.yaml	feat: add all services	3 weeks ago

README MIT license

## SOLID/RAG Toolbox

**NOTE** This setup is created for a demo project for PSX. There are still some references present to have a basic setup of policies and user roles in keycloak. There are 3 personae: Adam, Blake and Sandra which have certain access rights on each others data. To add data for these personae see "Adding Data" in the documentation, but make sure to add a metadata field 'type' with one of the following: calendar, financial, medical. For these 3 types and the personae policies are created on startup.

https://github.com/imec-int/solid-rag-toolbox/network/dependencies

**About**  
No description, website, or topics provided.

- Readme
- MIT license
- Activity
- Custom properties

0 stars  
5 watching  
0 forks

**Releases**  
No releases published

**Packages**  
No packages published

**Contributors 2**

- taeymma Maarten Taeymans
- ismakutl Ismail Kutlu

**Languages**

- Python 86.7%
- Dockerfile 9.9%
- PKI 3.4%





# Our ambition...

... is to build an integrated personal data store stack that is compliant by design, AI ready and data space compatible.

Reach out if you want to connect:

[tanguy.coenen@imec.be](mailto:tanguy.coenen@imec.be)

[www.linkedin.com/in/tanguycoenen/](http://www.linkedin.com/in/tanguycoenen/)

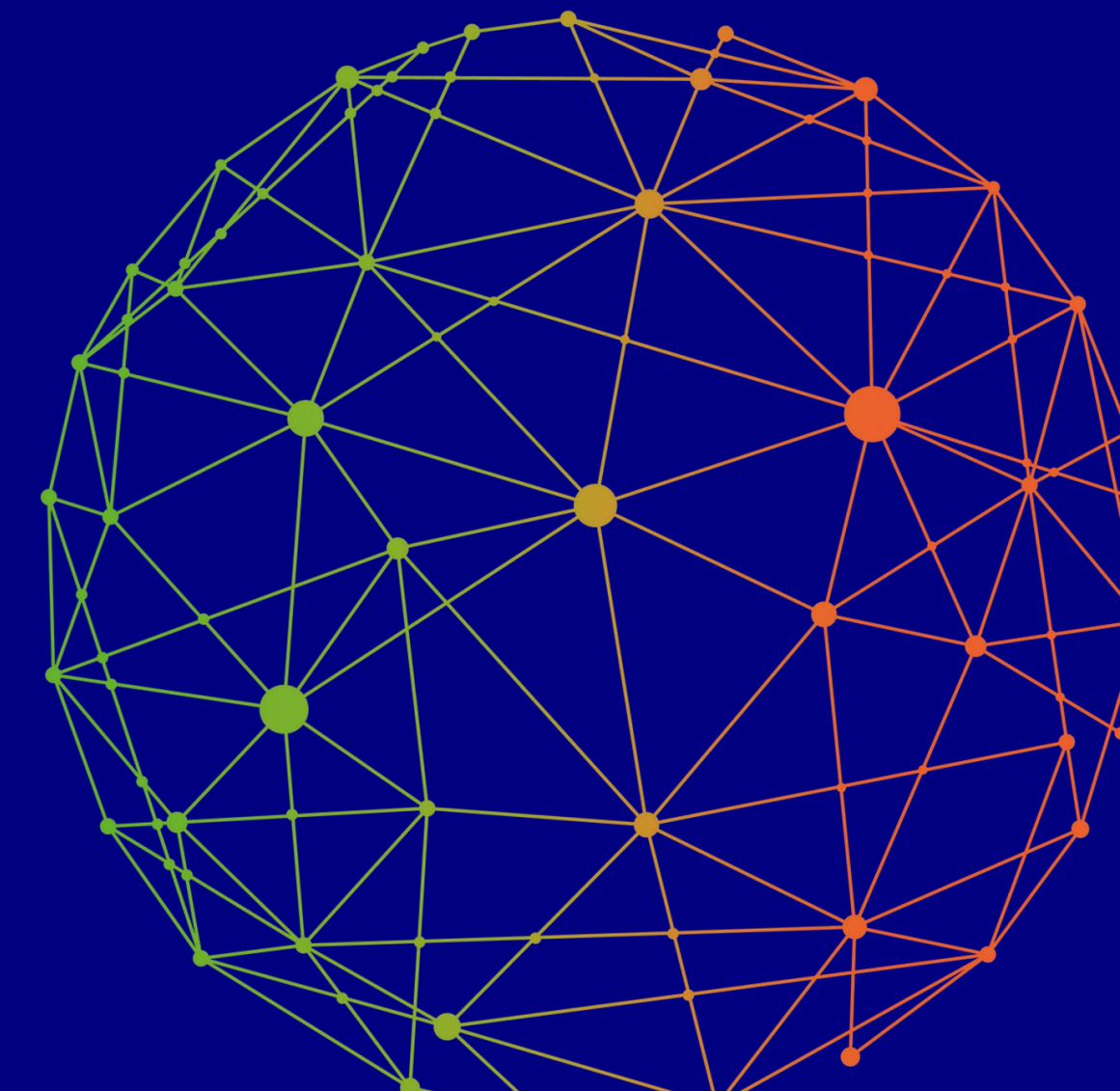


# Leveraging data spaces for the use of AI – Illustrated by practice-oriented software demonstrators

## Focus session AI

DSS Spaces Symposium 2025

Marius Bensley (acatech/MISSION KI)



DSBA



BDV  
BIG DATA VALUE  
ASSOCIATION

FIWARE  
FOUNDATION

gaia-x



INTERNATIONAL DATA  
SPACES ASSOCIATION



DATA SPACES  
SUPPORT CENTRE

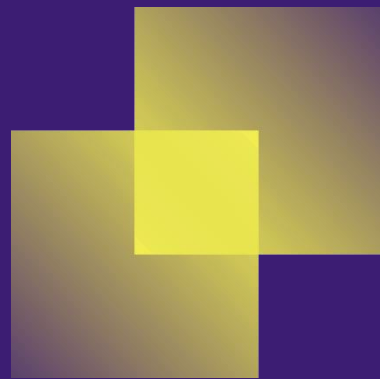


Funded by  
the European Union

The Data Spaces Support Centre receives funding from the European Union Digital Europe Programme under grant agreement n° 101083412

# MISSION KI strengthens Germany's competitiveness in AI along three pillars

1



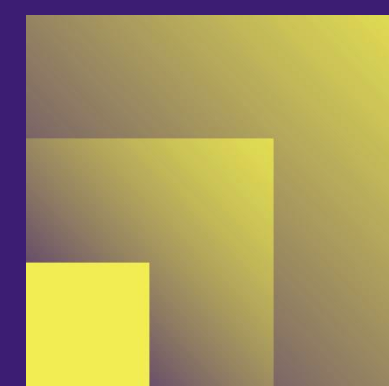
Improving the data  
**foundation** for AI

2



Strengthening the  
**development** of  
trustworthy AI

3



Supporting the **growth**  
of AI innovations

## Data spaces can support AI training while also using AI to enhance data quality and compliance

AI models require **domain-specific training data** to make reliable predictions



Data spaces provide **diverse and high-quality datasets** for AI while AI improves data spaces through **validation and enrichment**

To be usable by AI, data produced and provided in Germany must be in a **directly usable and quality-verified format**



Data spaces enable **curated data sharing**. AI increases **usability** of data spaces by structuring and optimizing data


To share suitable data for AI without **legal violations**, compliance with data provision regulations (e.g. GDPR, Data Act) must be ensured



Data spaces ensure **legal compliance with data law** - AI supports this by detecting **risks** and automating governance

## MISSION KI is developing software demonstrators in real-world use cases and will offer them for free use


MISSION KI focuses on projects that demonstrate how AI benefits from high-quality, interoperable and legally compliant data

1 

**FAIR Digital Objects (FDOs) for making data findable, accessible, interoperable and reusable**

**What it does:** Standardizes and connects data descriptions for seamless discovery


**AI Benefit:** Enables AI agents to autonomously interpret and use datasets through metadata-driven functionality

2 

**Dataset search engine for publishing and discovering domain-specific data**

**What it does:** Enables cross-domain search for relevant datasets

**AI Benefit:** Improves training data amount and quality, leading to better AI outputs across sectors

3 

**AI-based compliance monitoring for enabling legally-compliant data transactions**

**What it does:** Uses AI to ensure compliance with legal frameworks

**AI Benefit:** Large Language Models (LLMs) actively monitor and enforce regulatory compliance, ensuring trustworthy AI usage.

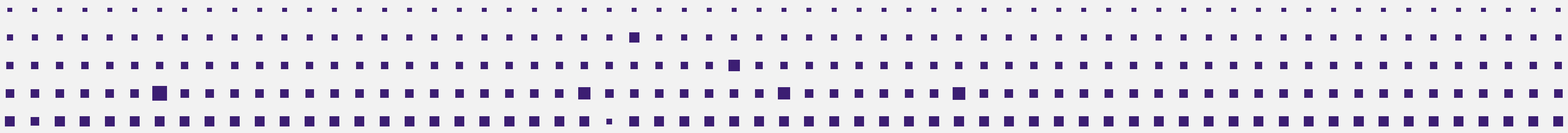
# AI-based Compliance-Monitoring (MIT Licenced)

Run compliance checks on your data transactions with our free-to-use Legal LLM

Gefördert durch:

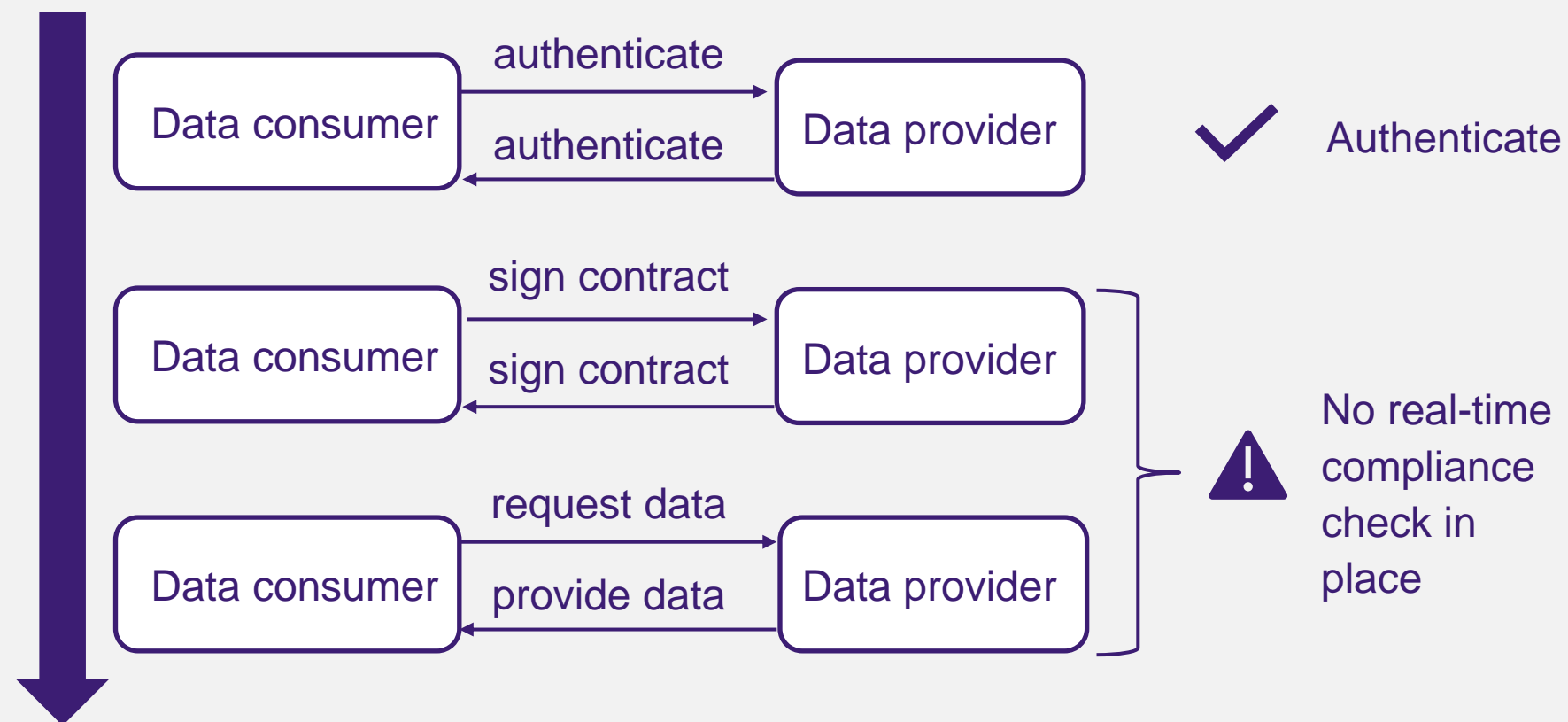


aufgrund eines Beschlusses  
des Deutschen Bundestages



## Today, transactions between data providers and consumers cannot be automatically checked for legal compliance

### Today's transaction process



### Different architectures for exchanging data...

- Peer-to-peer exchange
- Connector-based data spaces
- Federated data spaces
- Data marketplaces

### ...rely on different approaches for securing data exchange

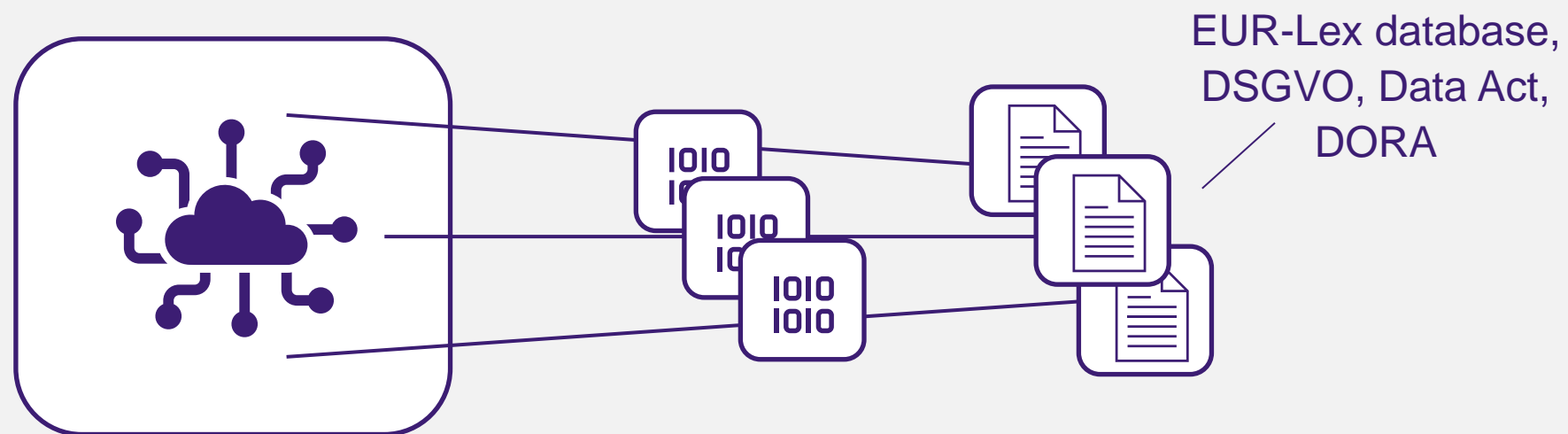
- Direct authentication
- Self-sovereign identity (SSI)
- Central access

### Ensuring compliance is currently a tedious effort

- No dynamic assessment of context-specific legal requirements
- No verification of legal compliance in real-time

## We trained a legal LLM based on the LLaMA 3.1 model that allows users to interactively monitor compliance

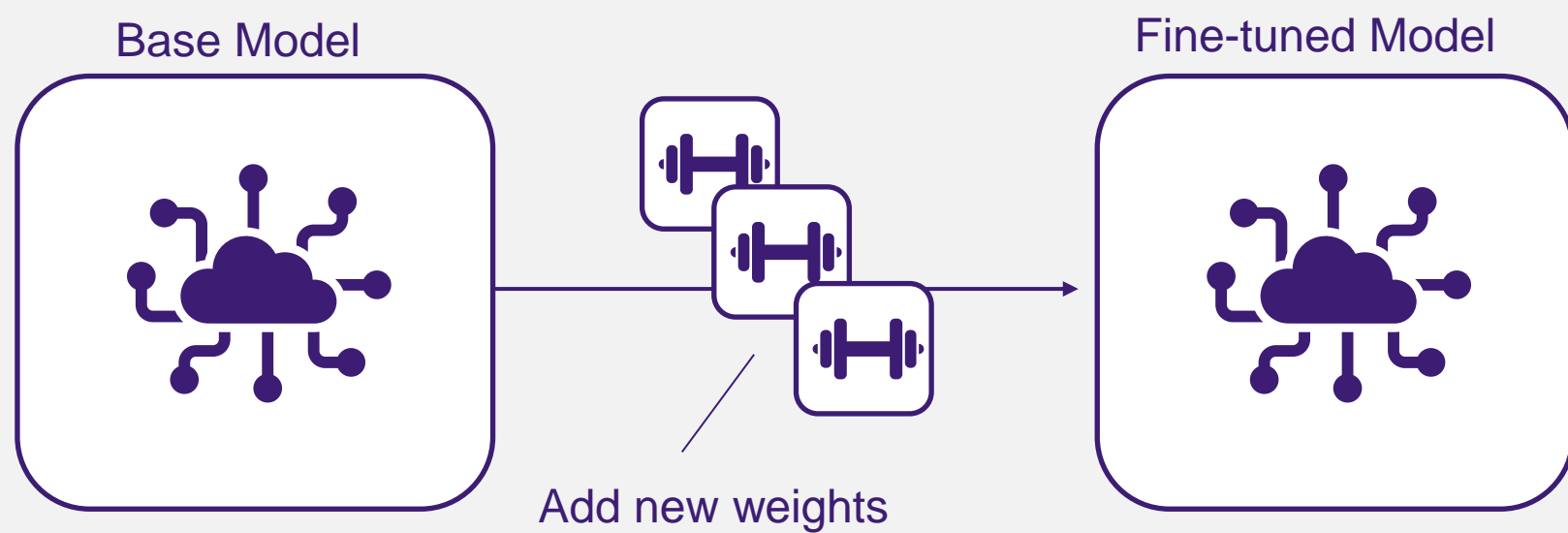
### 1 Training the LLAMA 3.1 model with relevant legal input texts



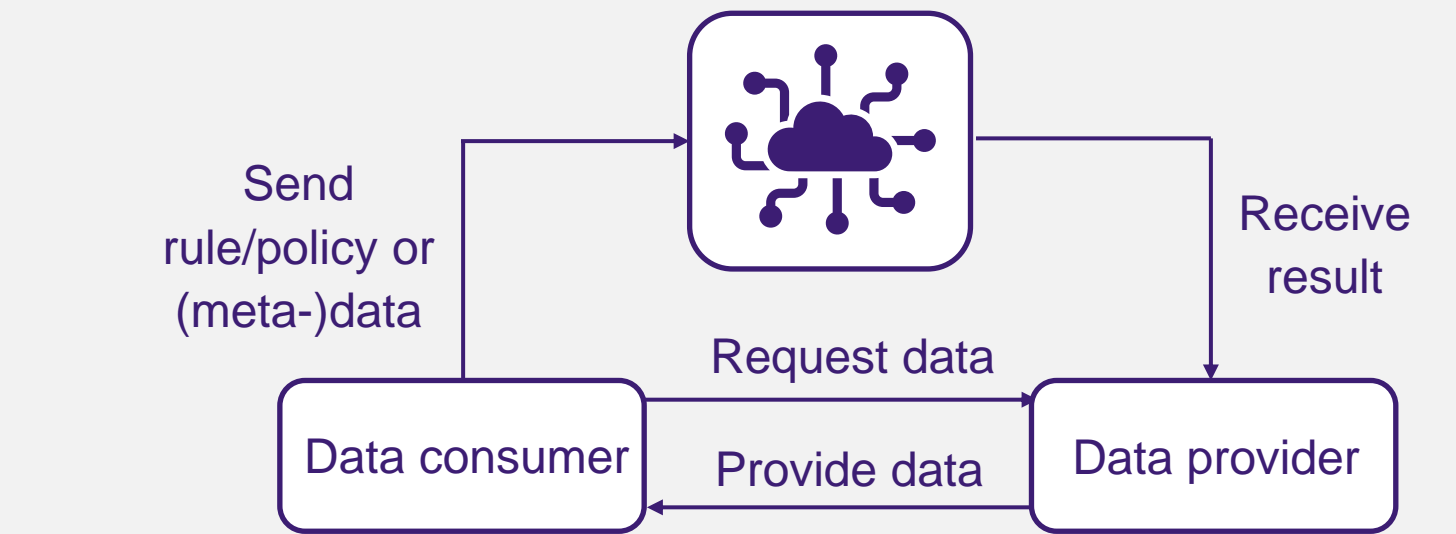
### 3 Adding a chat interface for interacting with the LLM



### 2 Using the re-adapt method for fine-tuning the LLM



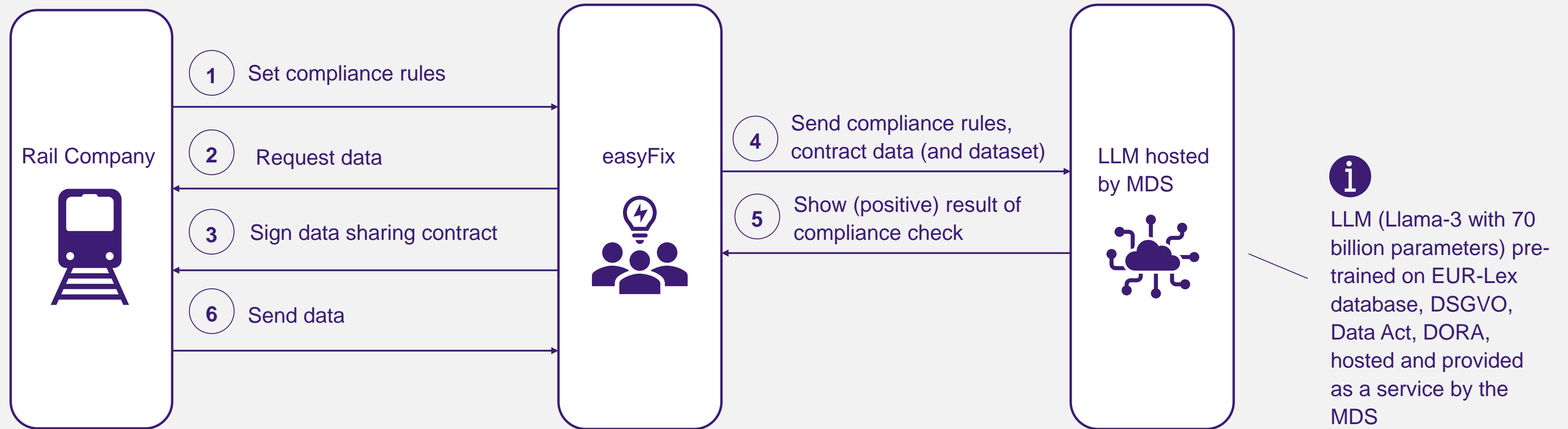
### 4 Using the LLM for checking contracts, answering questions etc.





## Our legal LLM allows data-driven users to automate their due diligence of data sharing contracts

Let's see how two companies (Rail Company and easyFix) from the Mobility Data Space can check the compliance of a planned data transaction



## Meet the team behind MISSION KI's AI-based Compliance Monitoring project

### Project lead

**MISSION KI**

**acatech**

DEUTSCHE AKADEMIE DER  
TECHNIKWISSENSCHAFTEN



Marius Bensley  
+49 (0)170 2024428  
Bensley@acatech.de

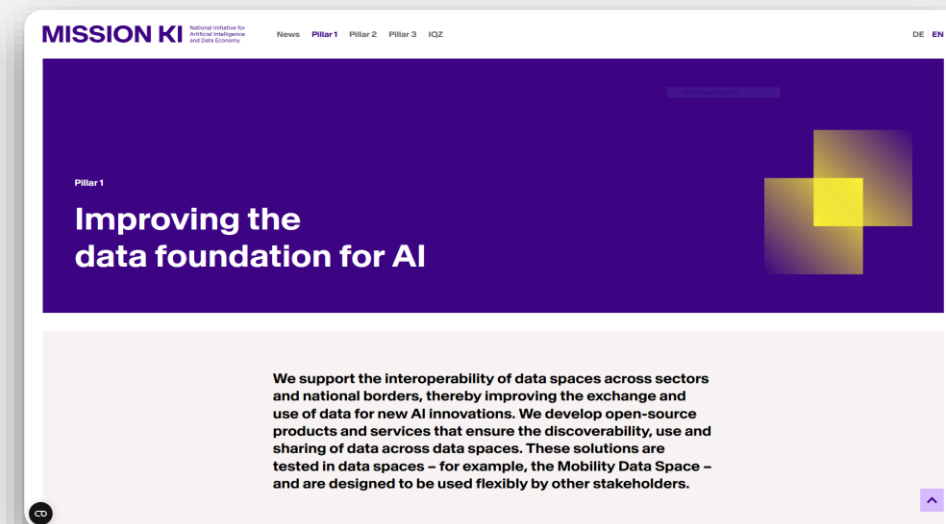
### Project partners



Project results to  
be published in  
April 2025



### Check out our MISSION KI website



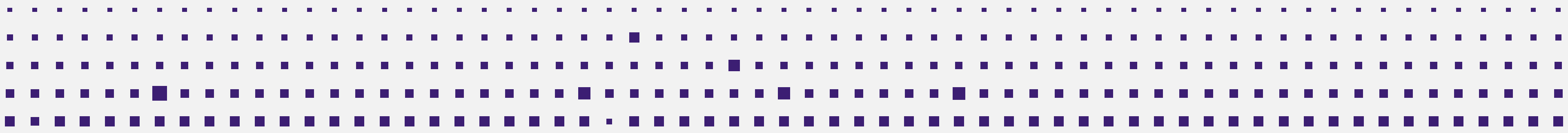
## Fair Digital Objects – A standardized format for data exchange

Providing you the tool(s) for building the global data space of FAIR Digital Objects

Gefördert durch:



aufgrund eines Beschlusses  
des Deutschen Bundestages



## Diverse data sources and standards create barriers for organizations that aim to exchange their data

### Untapped data

**76%** of German companies report that **data sinks keep them from exchanging data**<sup>1</sup>

### Unshared data

**81%** of German companies **do not share data** with other companies<sup>2</sup>

### Heterogeneous data sources

**> 255.000** data companies<sup>3</sup> in the EU along with 14 sector-specific data spaces and several marketplaces are contributing to the **growing data economy across various domains**

### Proprietary metadata standards

**> 100** metadata standards across various domains, including those promoted by large tech companies, **impede interoperability and data integration**<sup>4</sup>

<sup>1</sup> XSphere Industriestudie (2023)  
<sup>2</sup> IW-Trends Studie (2024)

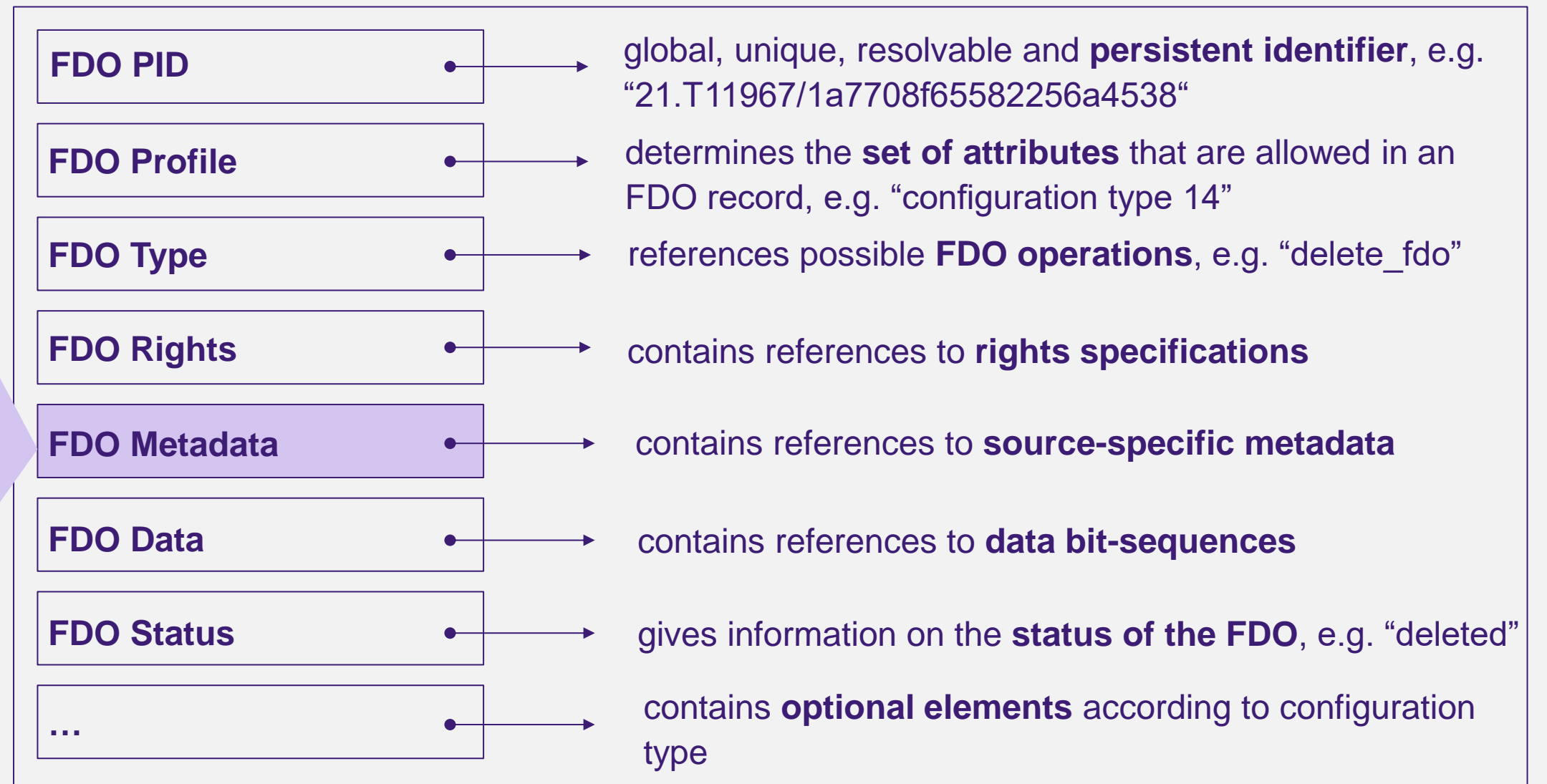
<sup>3</sup> Helmholtz Metadata Collaboration (2025), Metadata-Standards-Catalog  
<sup>4</sup> data.europa.eu (2017), The size and trends of the EU data economy

# FAIR Digital Objects (FDOs) let us integrate these isolated standards by offering an interoperability container

## Source-specific metadata schema (example)

<b>Name:</b>	Public Transport Realtime Data Baden-Württemberg
<b>Publisher:</b>	NVBW - Nahverkehrsgesellschaft Baden-Württemberg mbH
<b>Structure:</b>	Text (CSV)
<b>Size:</b>	1.1 MB
<b>Licence:</b>	dl-de/by-2-0
<b>Data types:</b>	Date (0), String (8), Numeric (6)

## Structure of a FAIR Digital Object (example FDO record according to configuration type 14)

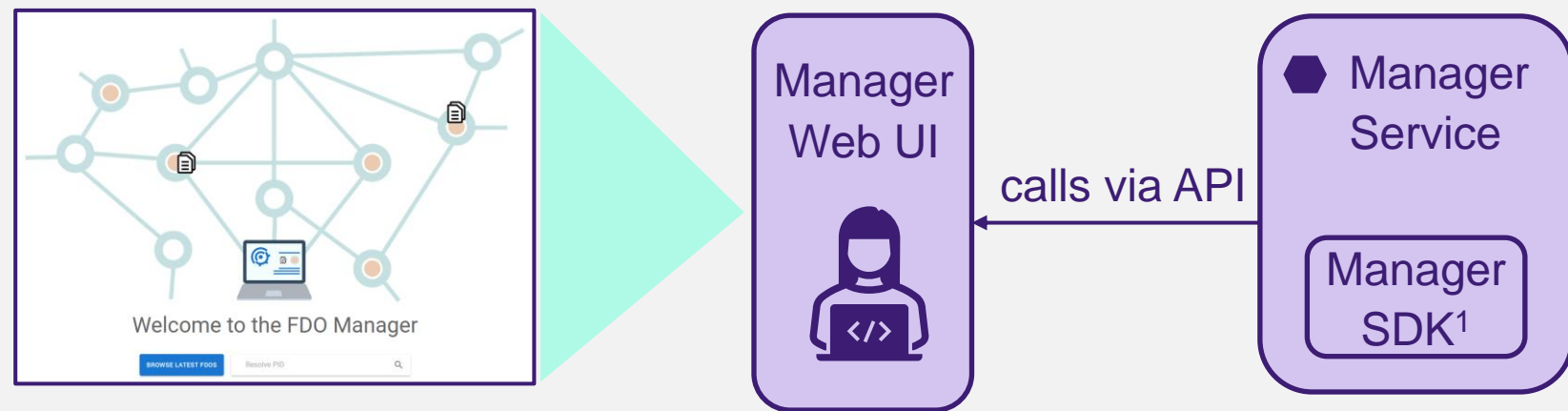


FDOs go beyond simple metadata schemes **by bundling data with additional components** - persistent IDs, rich metadata, and machine-actionable structures - into secure, machine-actionable, and reusable digital objects that allow **FAIR (findable, accessible, interoperable and reuseable)** data use

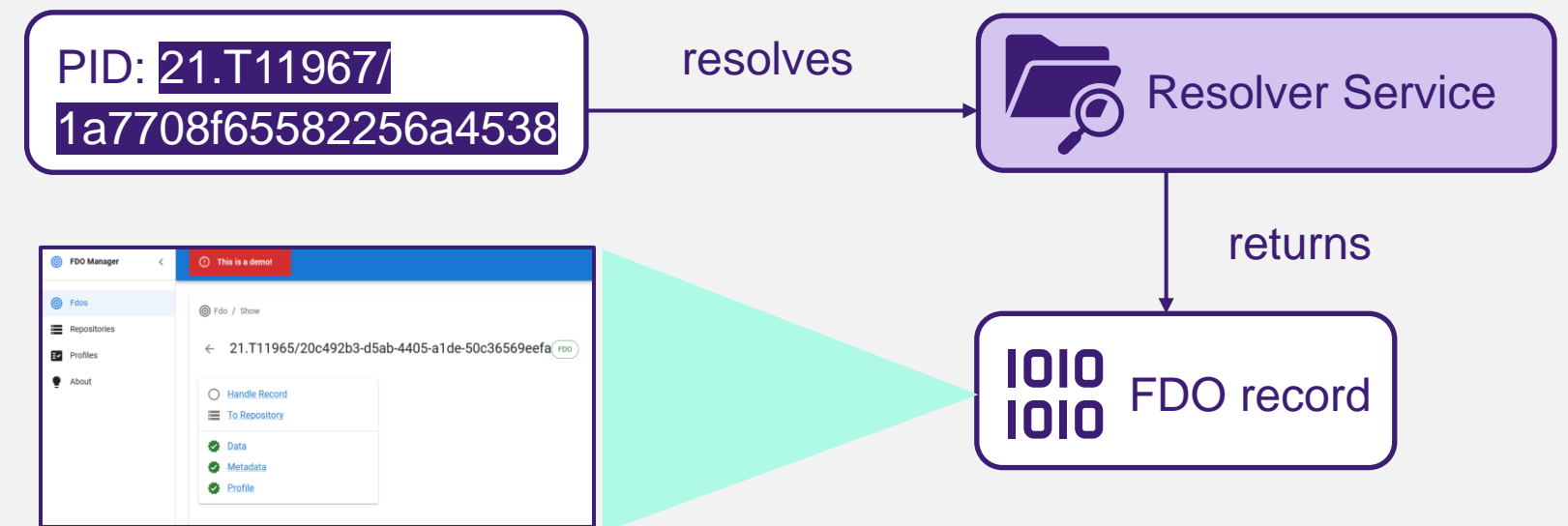
## We developed a suite of software tools that allow users to publish and read FDOs from various data sources

◆ = Business Logic   
   = MISSION KI developments   
   = Legacy

1 We built a **manager service** that allows the manipulation of FDOs via a **web-based UI**

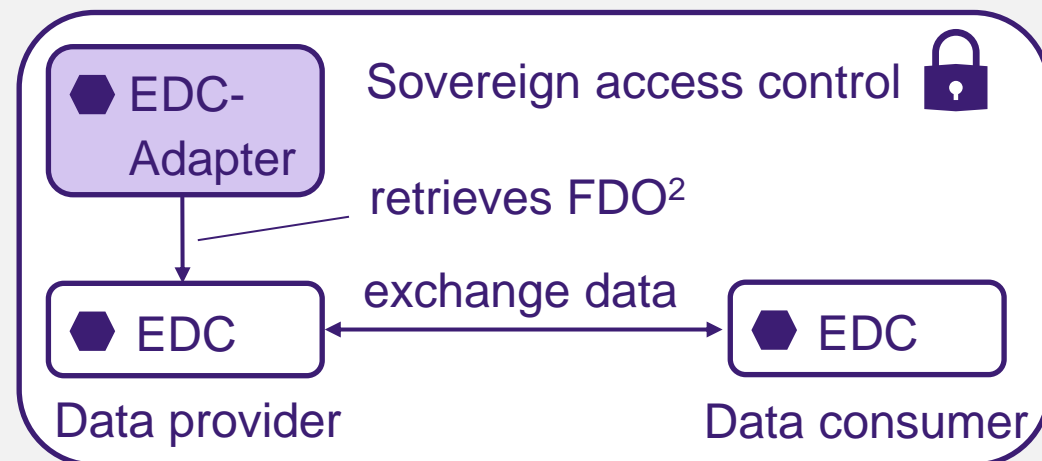


3 A single FDO can be resolved by a **web-based resolver service** via its PID (e.g. handle)...

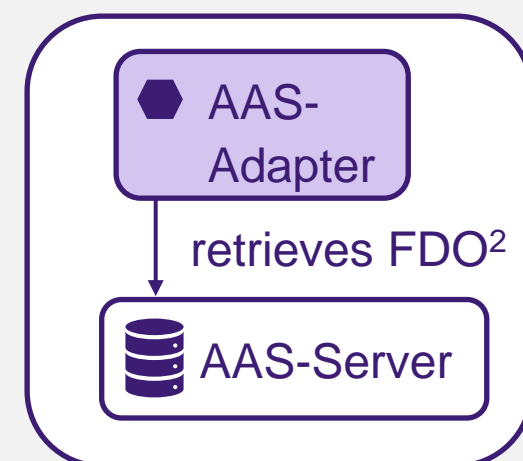


2 Our **EDC-/AAS-Adapters** can generate and retrieve FDOs from sovereign data spaces or AAS repositories in companies and several other sources

Example: Data space



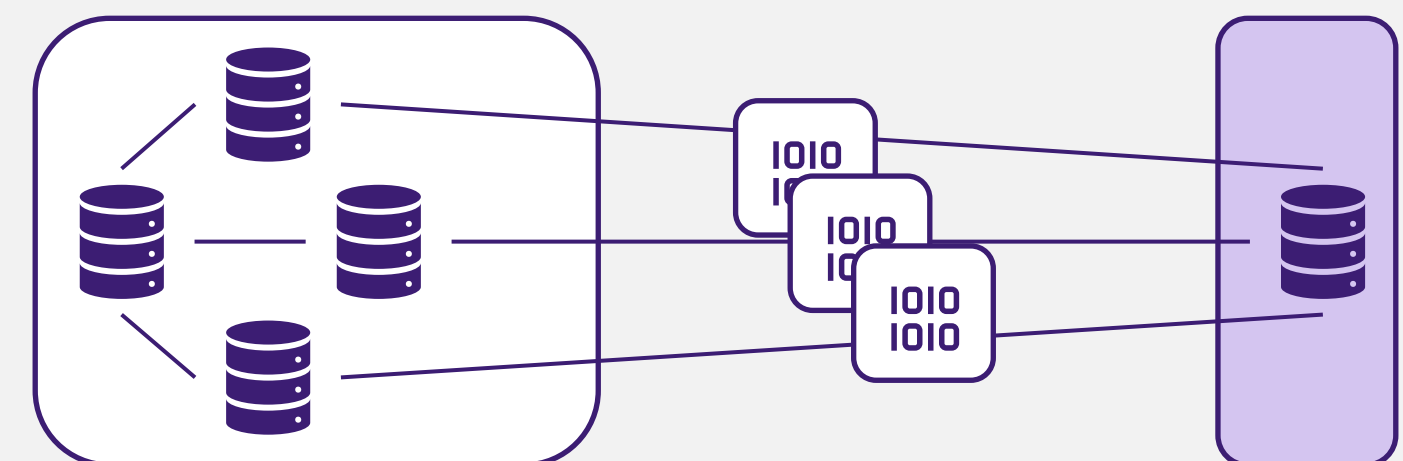
Example: Single Company



4 ...which is part of a **distributed FDO registry** (e.g. the Handle.net network) of which we provide a private, testable instance

Global FDO registry (Handle.net)

Private registry instance



¹ SDK = Software Development Kit    ² Analogously for FDO generation

## Meet the team behind MISSION KI's Fair Digital Objects project

### Project lead

**MISSION KI**

**acatech**

DEUTSCHE AKADEMIE DER  
TECHNIKWISSENSCHAFTEN

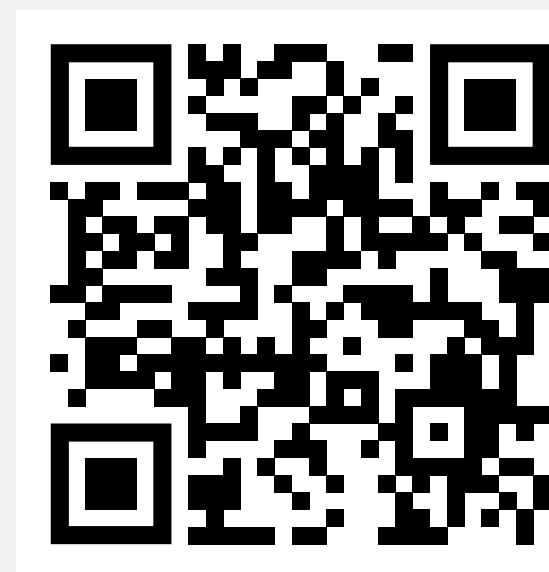
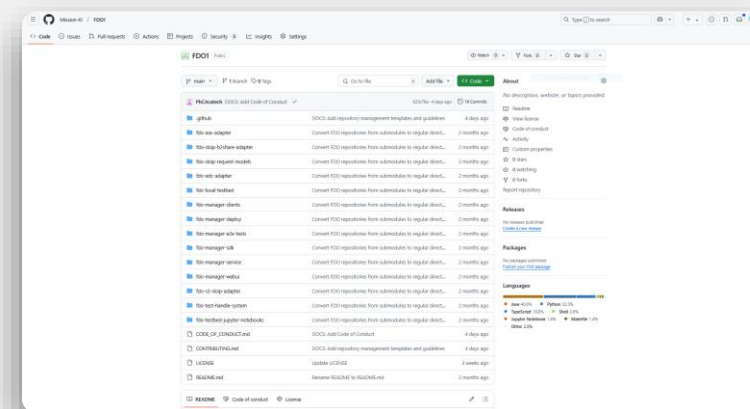


Marius Bensley  
+49 (0)170 2024428  
Bensley@acatech.de

### Project partners



### Check out our website on Github



## Dataset search engine (MIT Licenced)

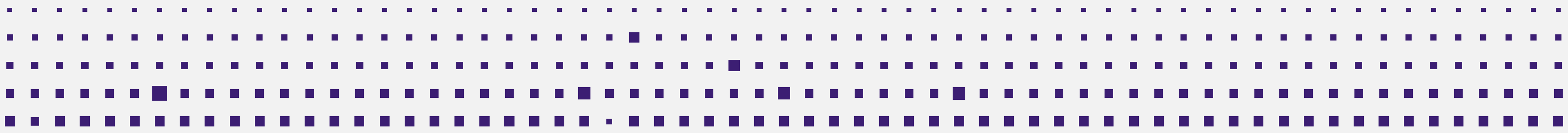
Publish and discover relevant information on  
datasets from various sources



Gefördert durch:



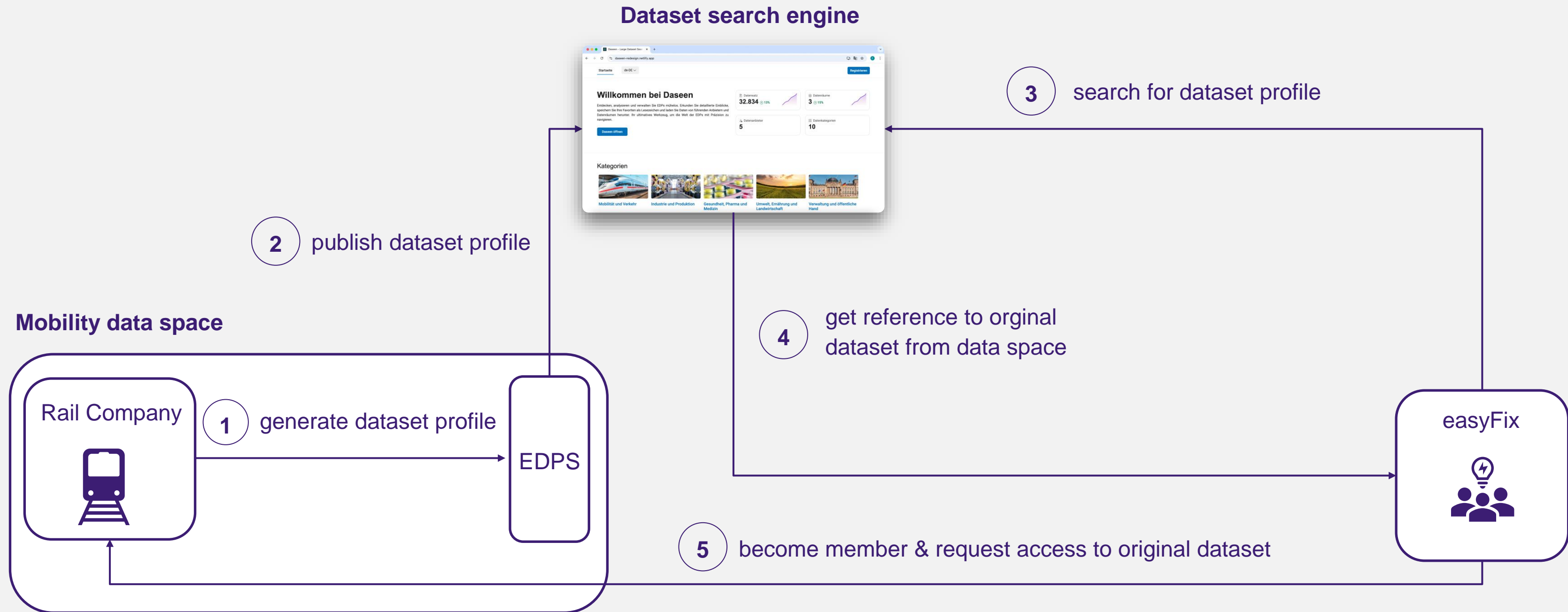
aufgrund eines Beschlusses  
des Deutschen Bundestages





## The dataset search engine makes it easy to publish and discover datasets from multiple sources with one search

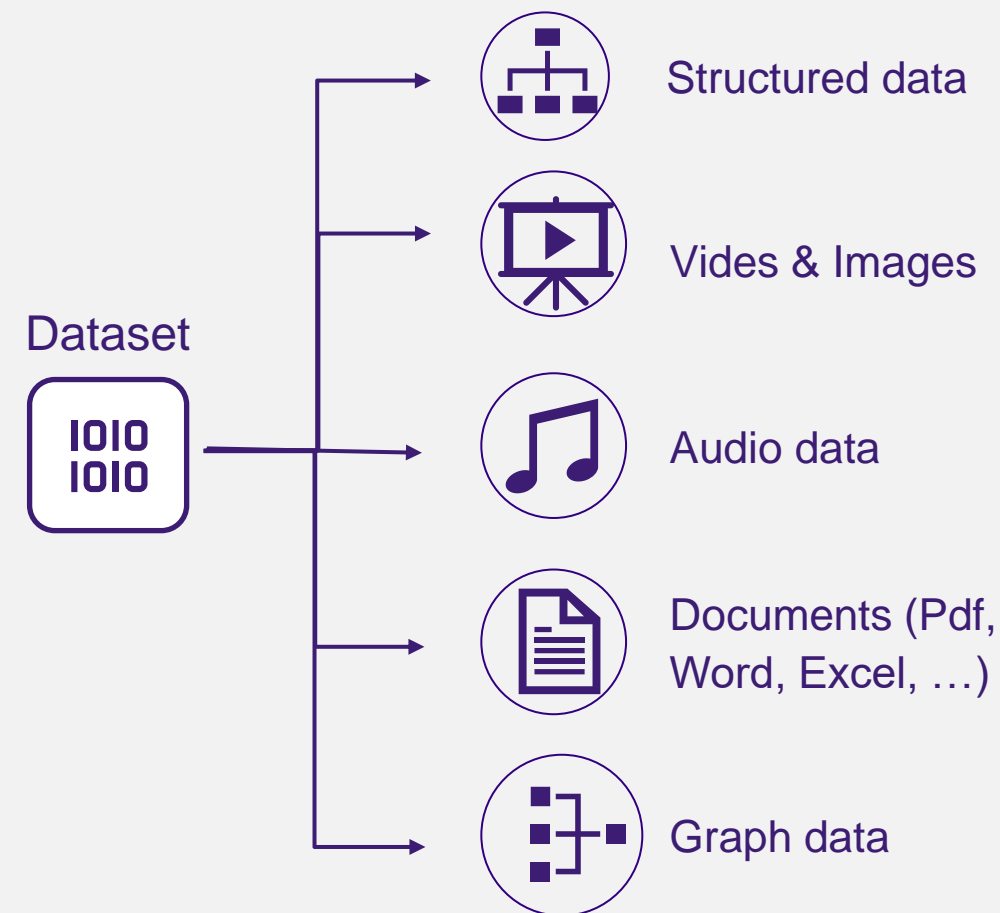
Let's see how two companies (Rail Company and easyFix) can publish and discover information on a dataset via the dataset search engine



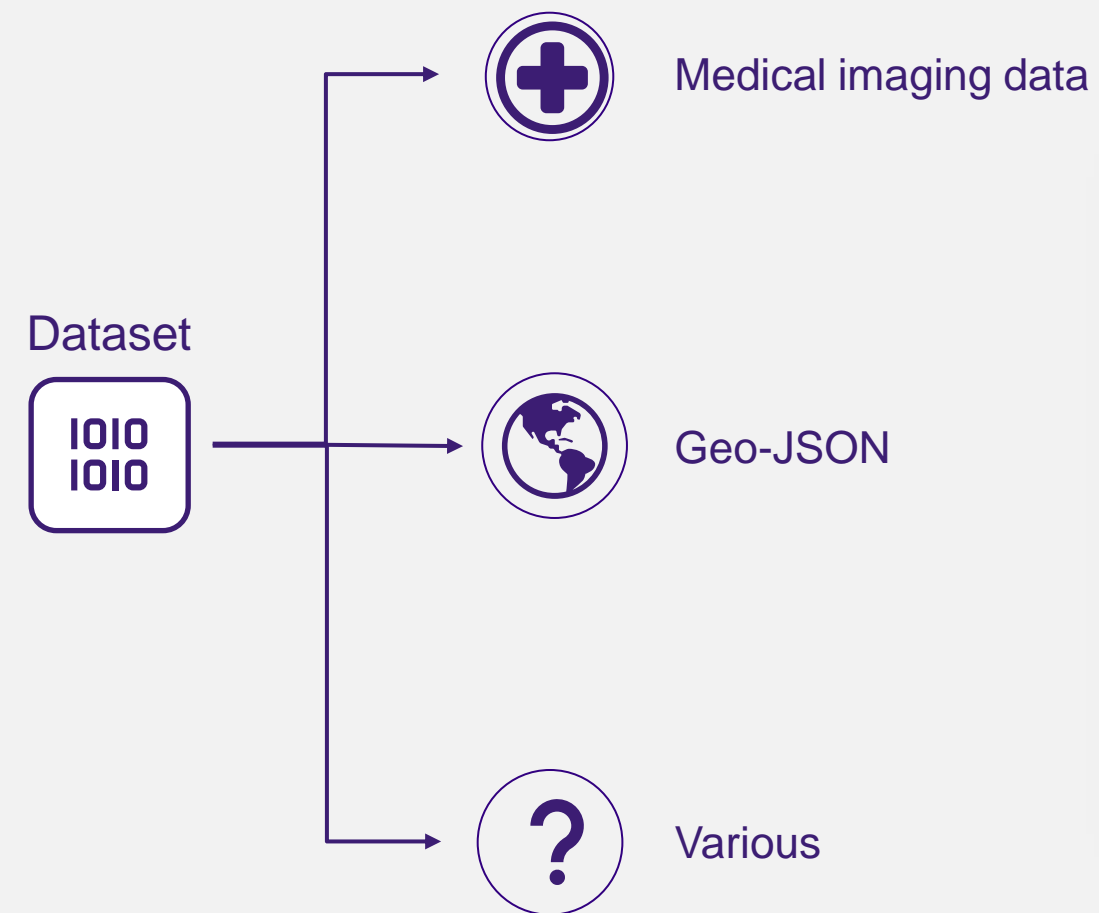
## The EDPS auto-generates metadata for diverse datasets, enabling filtering and custom analyses

Domain-specific users can enhance dataset profiles by integrating custom analysis methods into the EDPS

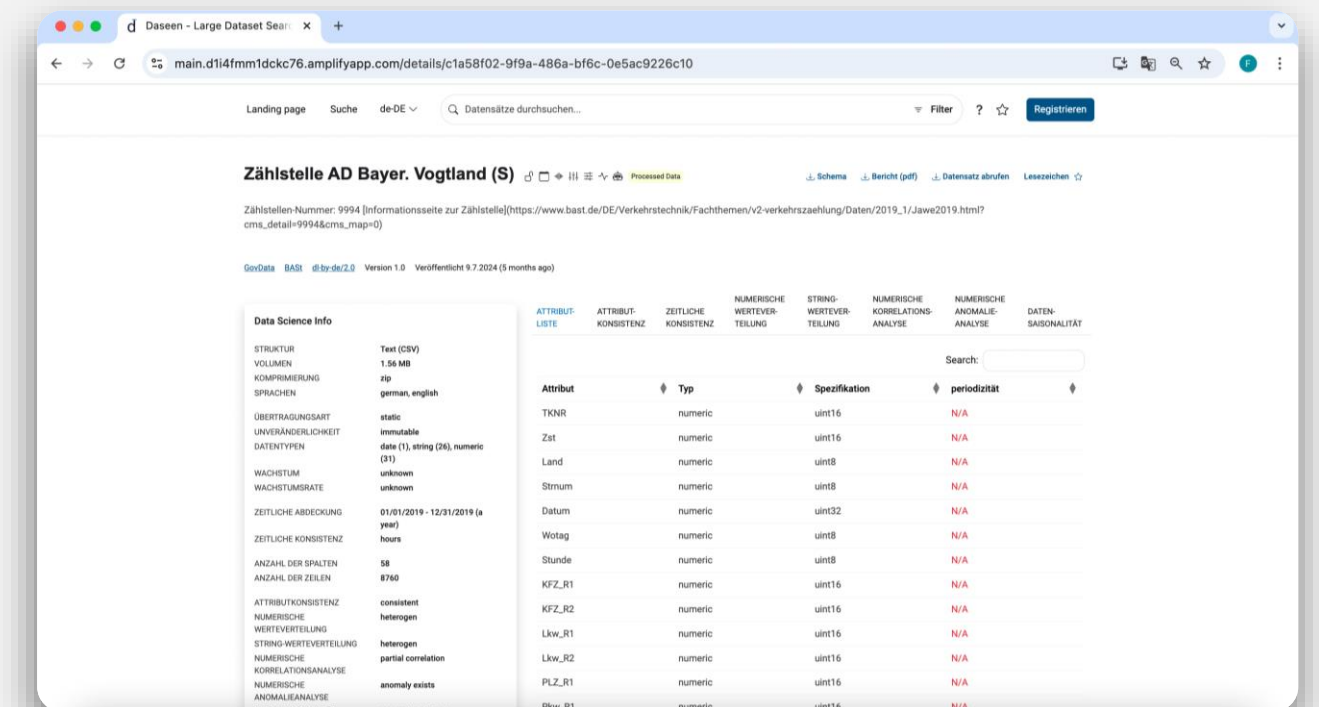
### 1 EDPs for common data types



### 2 EDPs for domain-specific data formats



### 3 Example of an EDP for a structured dataset with adjacent metadata



## Meet the team behind MISSION KI's dataset search engine project

### Project lead

**MISSION KI**

**acatech**

DEUTSCHE AKADEMIE DER  
TECHNIKWISSENSCHAFTEN

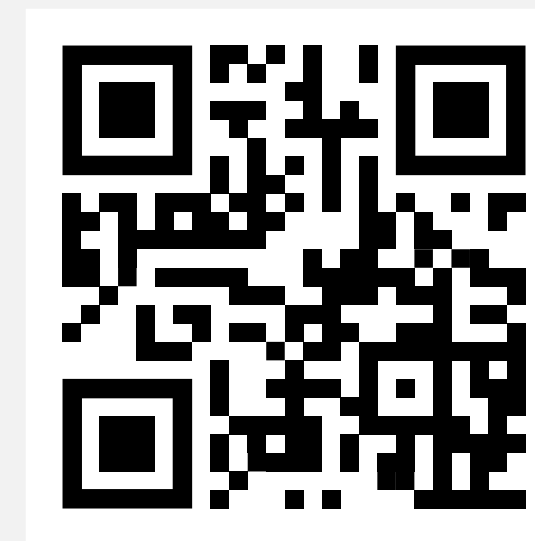
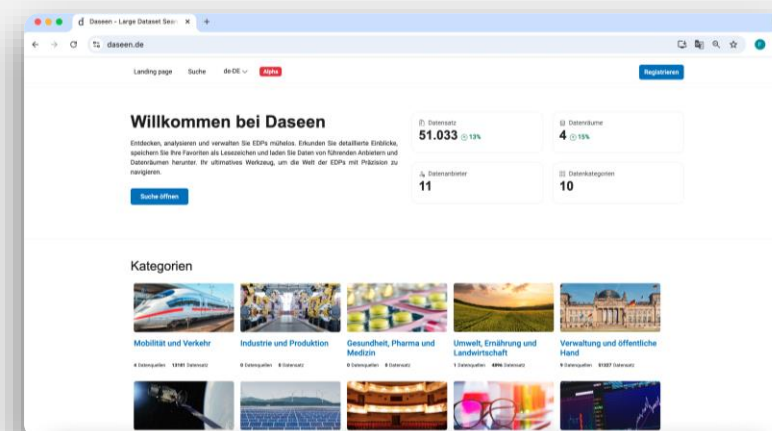


Marius Bensley  
+49 (0)170 2024428  
Bensley@acatech.de

### Project partners



### Check out our website on Github



**Thank you!**



Marius Bensley  
Strategic Project Manager  
Acatech/MISSION KI  
+49 (0)170 2024428  
Bensley@acatech.de



**Follow us on LinkedIn**



# Data Spaces Symposium

12:30

## Lunch & networking

---

Lunch is served – see you back for the next session!



Patronat polskiej prezydencji w Radzie UE  
Patronage of the Polish presidency of the Council of the EU  
Patronage de la présidence polonaise du Conseil de l'UE