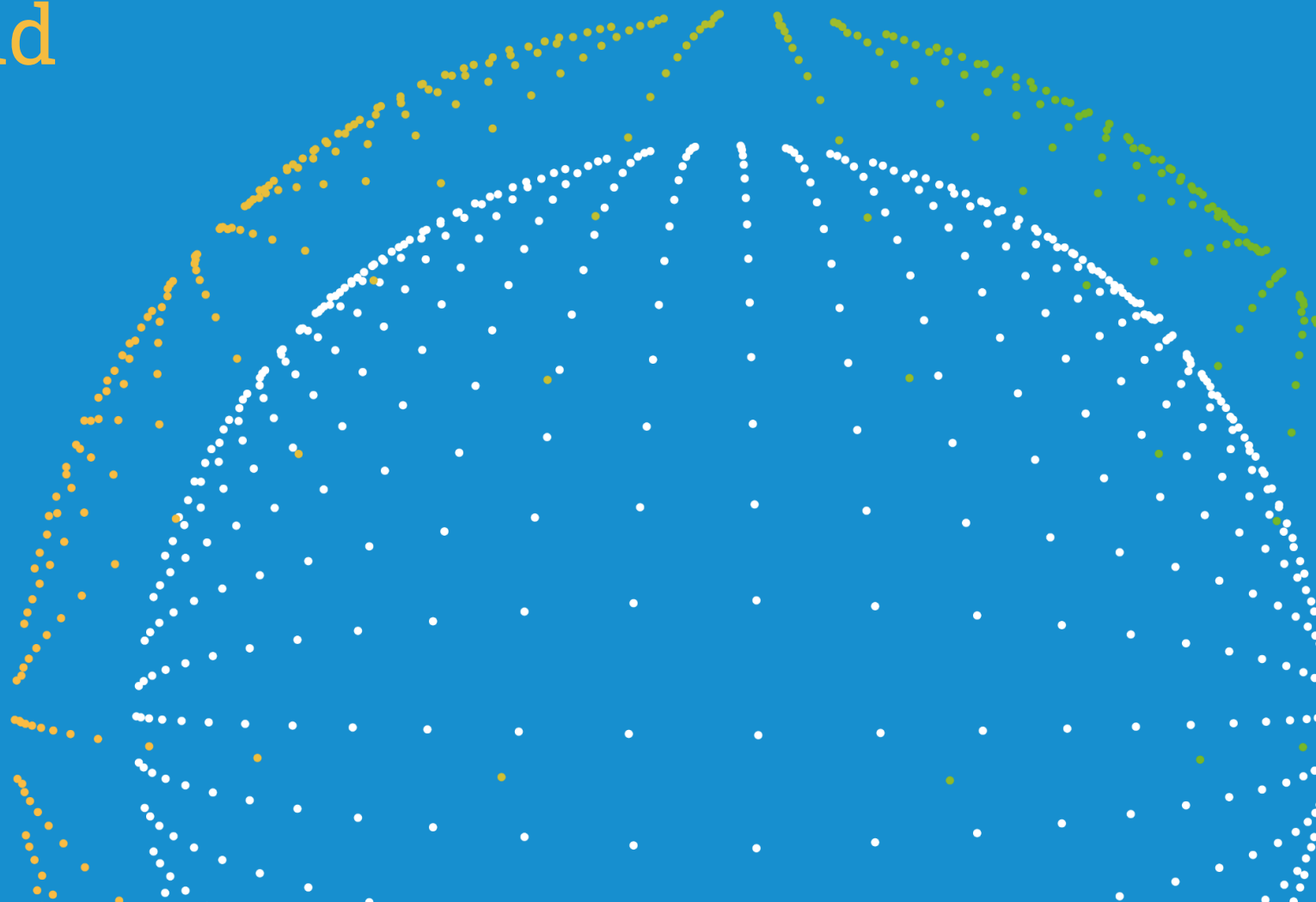
Data Spaces Symposium

Innovation in Data Spaces: lowering barriers and increasing value

Daniel Alonso, Alberto Abella, Antoine Garnier, Jeanette Nilsson, Nuria De Lama, Ed Curry, Georg Rehm, Shane Ó Seasnáin



Objectives of the session

- To discuss on how to **lower barriers** to facilitate the design, development, use and onboarding in data spaces, from different perspectives: **skills and training** needed for **easy access** and onboarding, how to facilitate **technical adoption** of data spaces

- To discuss **value creation on data spaces**, how **AI and other disruptive applications** can benefit from their connection to controlled and trusted environments as data spaces, that can fuel **through data sharing the potential of those applications**. Identify **requirements and benefits**

Data Spaces Symposium

Unite. Innovate. Adopt.

Innovation in Data Spaces: lowering barriers and increasing value

13 March 2024 | 14:00 - 15:15

Daniel Alonso
BDVA

Alberto Abella
FIWARE

Antoine Garnier
IDSA

Jeanette Nilsson
RISE

Ed Currry
Insight

Shane O'Shean
TUE

Georg Rehm
DFKI

Nuria De Lama
IDC

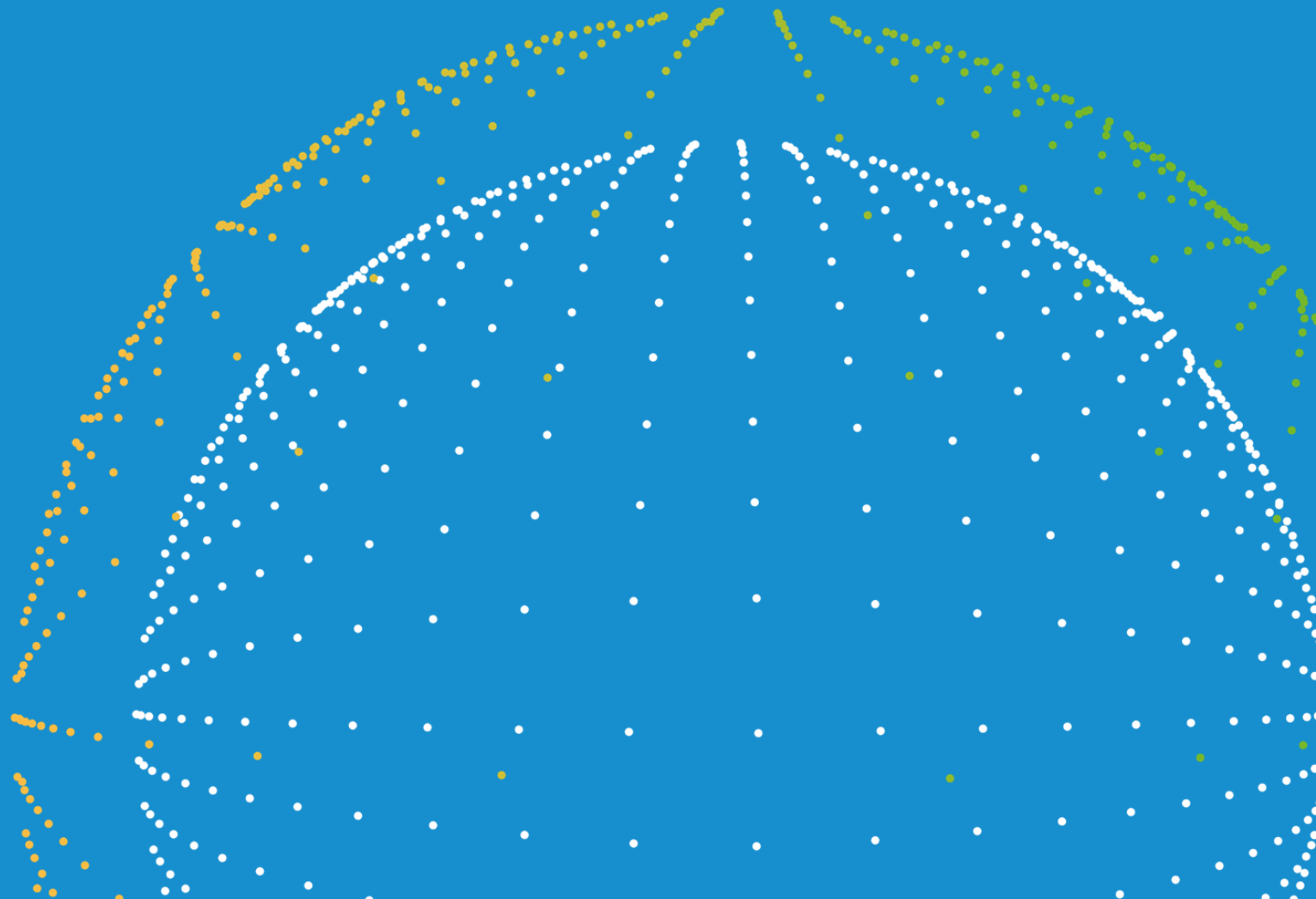
Agenda

Daniel Alonso (BDVA)	Intro and framework
How to lower barriers and boost innovation in DS (initial statement + panel discussion)	
Alberto Abella	Hands-on technical approach to data spaces. FIWARE connector, solutions and data models
Antoine Garnier	Teach skills to build data spaces. How to equip workforce with the necessary knowledge
Jeanette Nilsson	Hubs / BDVA i-Spaces as practical instruments to lower those barriers
Value creation in Data Spaces (initial statement + panel discussion)	
Nuria De Lama	Enablers for providing (business) value
Edward Curry	Data Spaces for Generative AI (and AI for DS). Intelligent Data Spaces
Georg Rehm	Large Language Models and Data Spaces. Observations from the project OpenGPT-X
Shane O'Shean	Digital twin and data spaces

How to lower barriers and boost innovation in data spaces

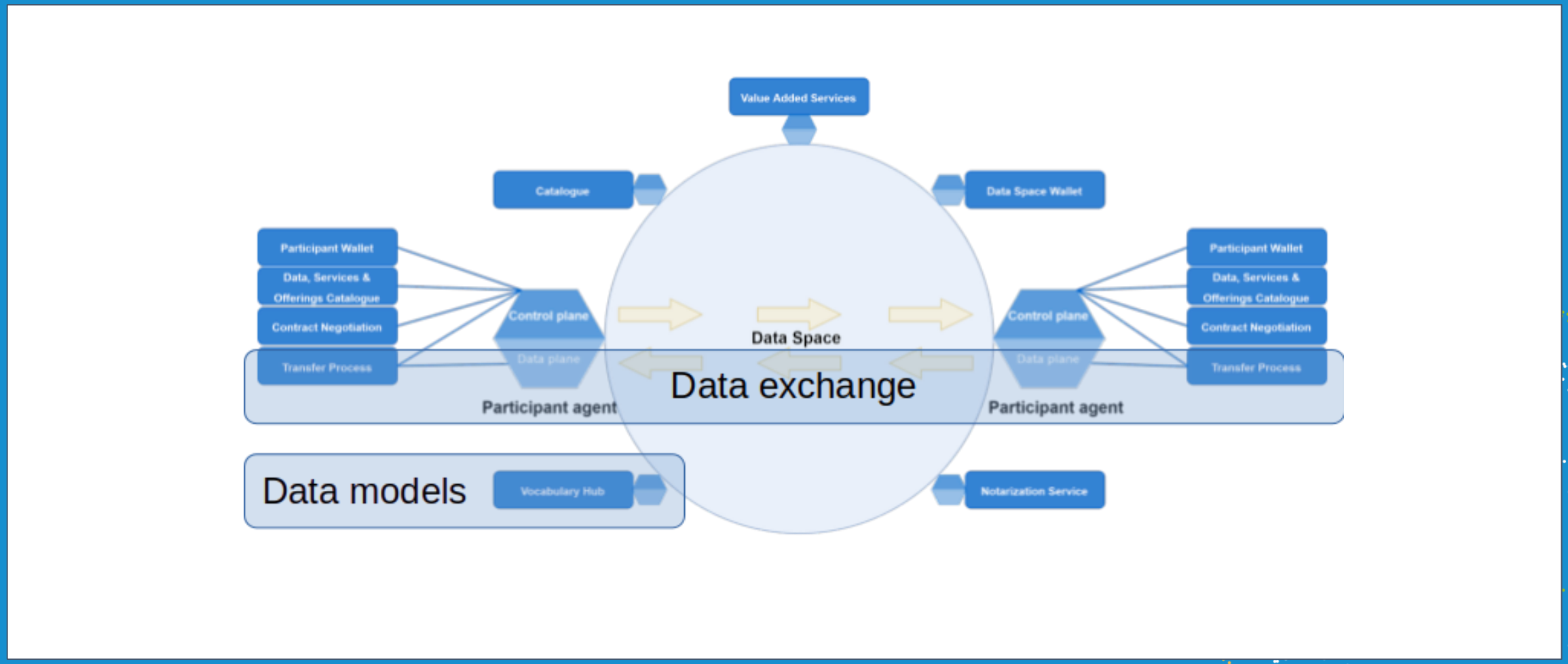
Data Spaces Symposium

FIWARE connector, solutions and data models

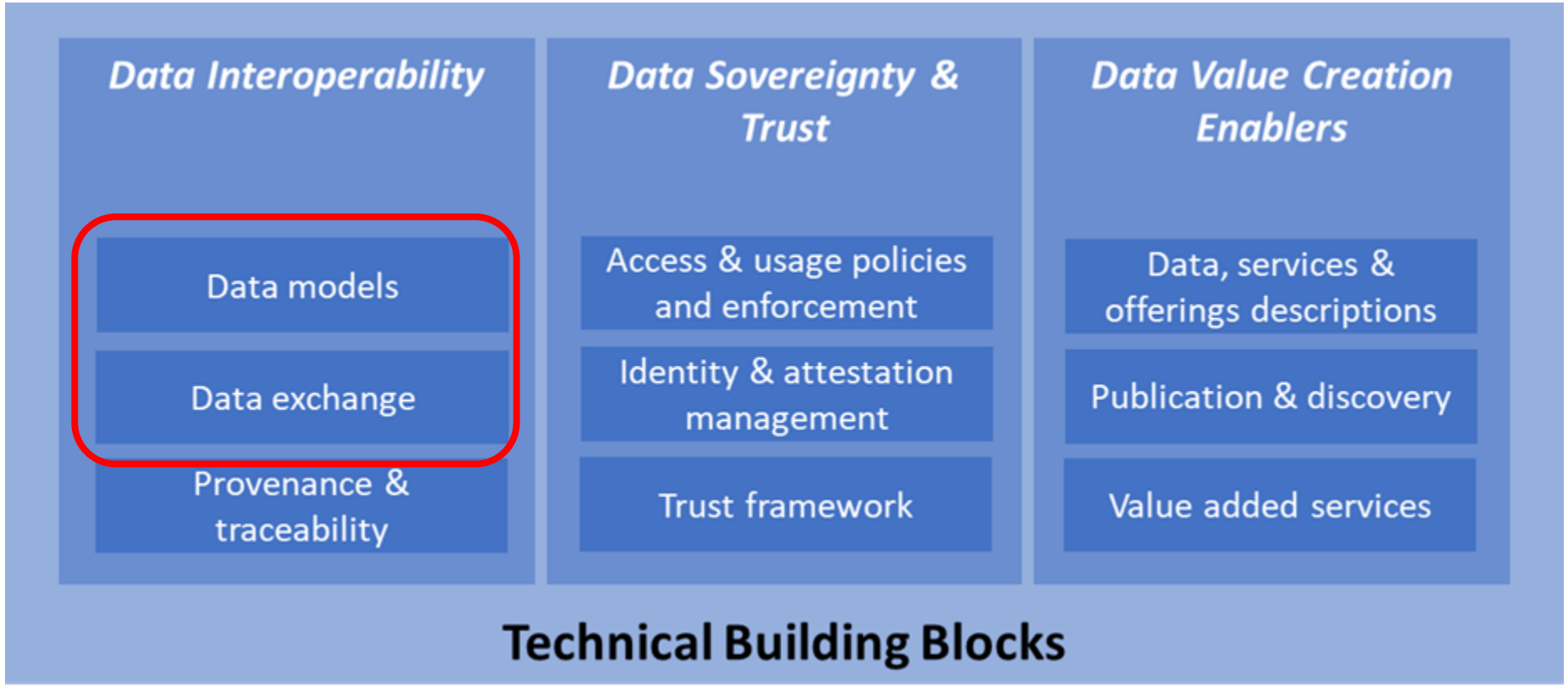


Alberto Abella.
FIWARE Foundation
Data Modelling Expert
@aabella

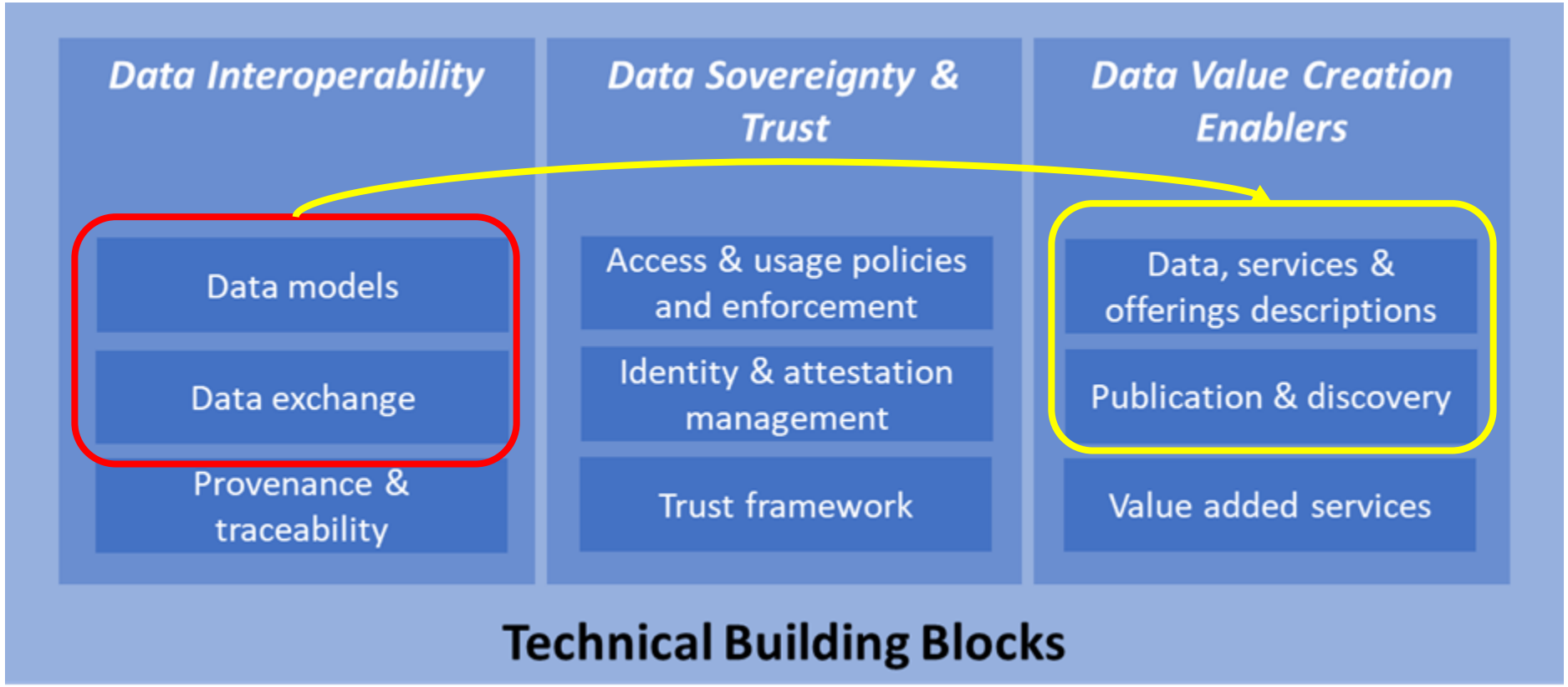
Data spaces by functions



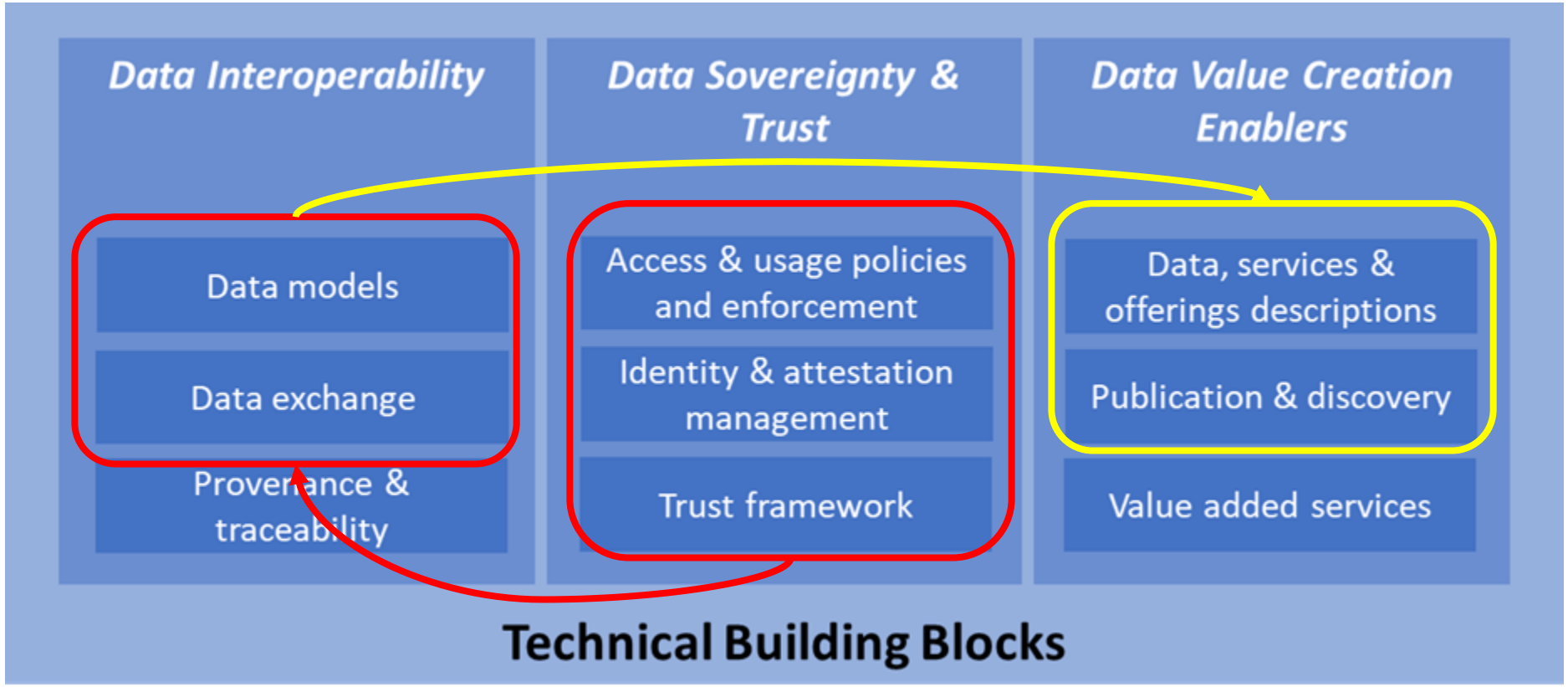
Data spaces by Building Blocks



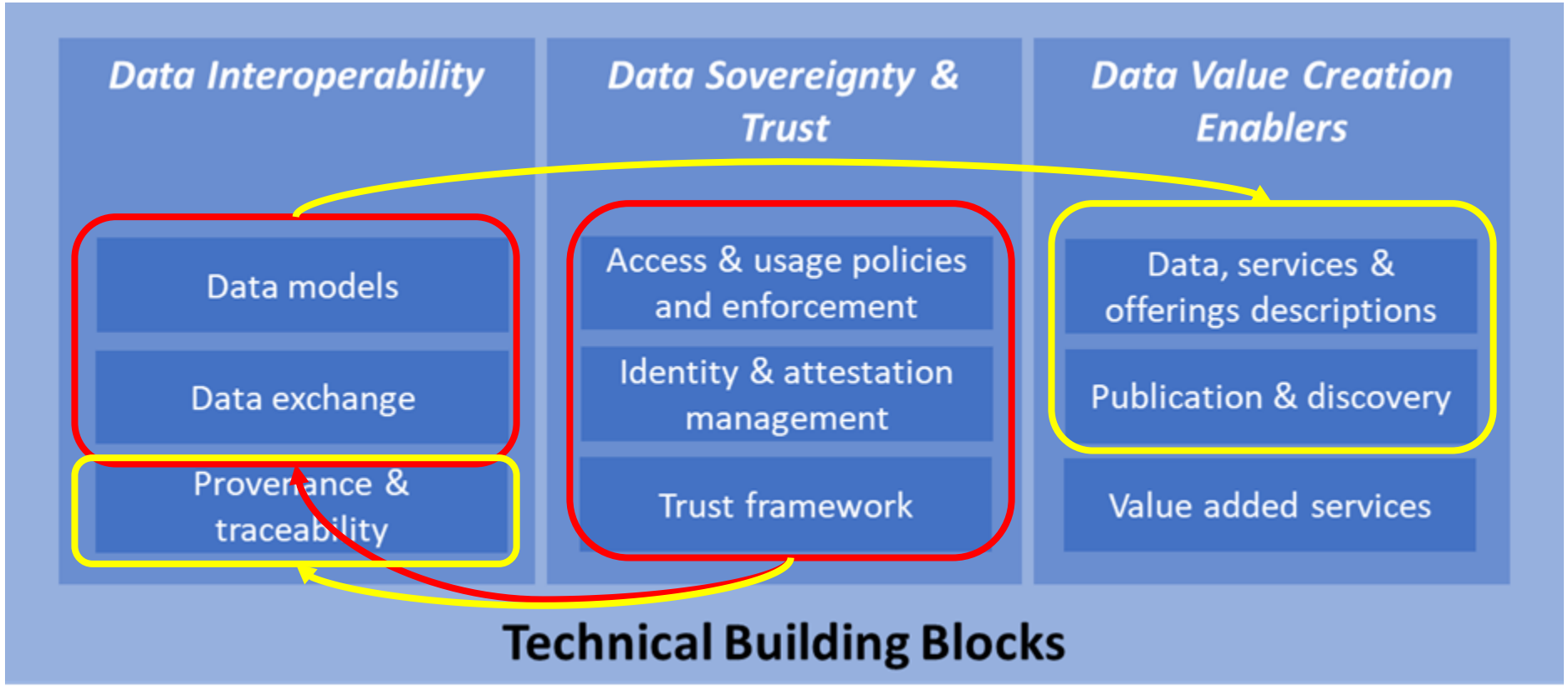
Data spaces by Building Blocks



Data spaces by Building Blocks



Data spaces by Building Blocks



FIWARE Data Space Connector components

FDSC Open source available at:
https://github.com/FIWARE-Ops/data-space-connector
It provides:

- Identity management W3C DID VC/VP
- Compliant with SIOPv2 / OIDC4VP protocols
- Data interchange based on NGSI-LD
- Attribute-based access control (ABAC) following an XACML P*P architecture
- Trust Services aligned EBSI specifications
- Compliant with EU digital identity / Wallet
- Compliant with DSBA recommendations
- Based on existing components

Used in multiple data spaces (i4Trust, DSCaaS, DOME, Citcom.ai, etc)

Smart Data Models initiative

Smart Data Models is a collaborative program to provide data models for digital twins and data spaces

- Free and open-licensed data models for digital market (0€ cost)
- Multisector
- Based on real use cases and adopted open standards. Collaborative.
- At market speed
- Customizable to local needs
- Compatible with linked data

Available at https://smartdatamodels.org
Repository at https://github.com/smart-data-models

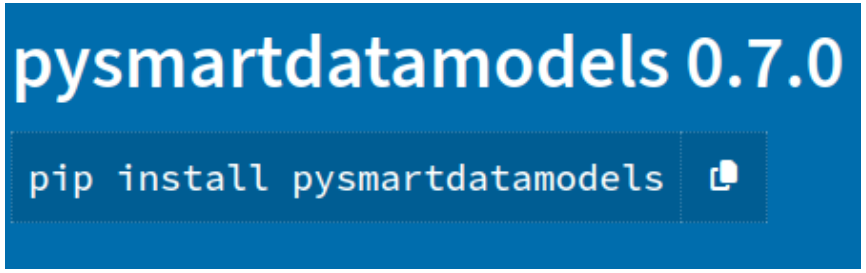
Smart Data Models for Data spaces

pysmartdatamodels python package includes 1.000 DM and 156.000 terms
It manages data models' assets with *23* functions including:

- Inserting compliant data in a context broker
- Listing any asset (names, definitions, types)
- Fuzzy search
- Generation of data model compliant examples
- Update from central repository
- Extends information beyond DCAT catalogue limitations (structure)
- Provides support for creating new DM on the fly

Soon:

- Manage your own data models (private)
- Check compliance with Data Models
- Contribute your data models



- Open source
- Fully documented
- Possible stand alone

- Available https://pypi.org/project/pysmartdatamodels/
- 104 stars
- 52 forks

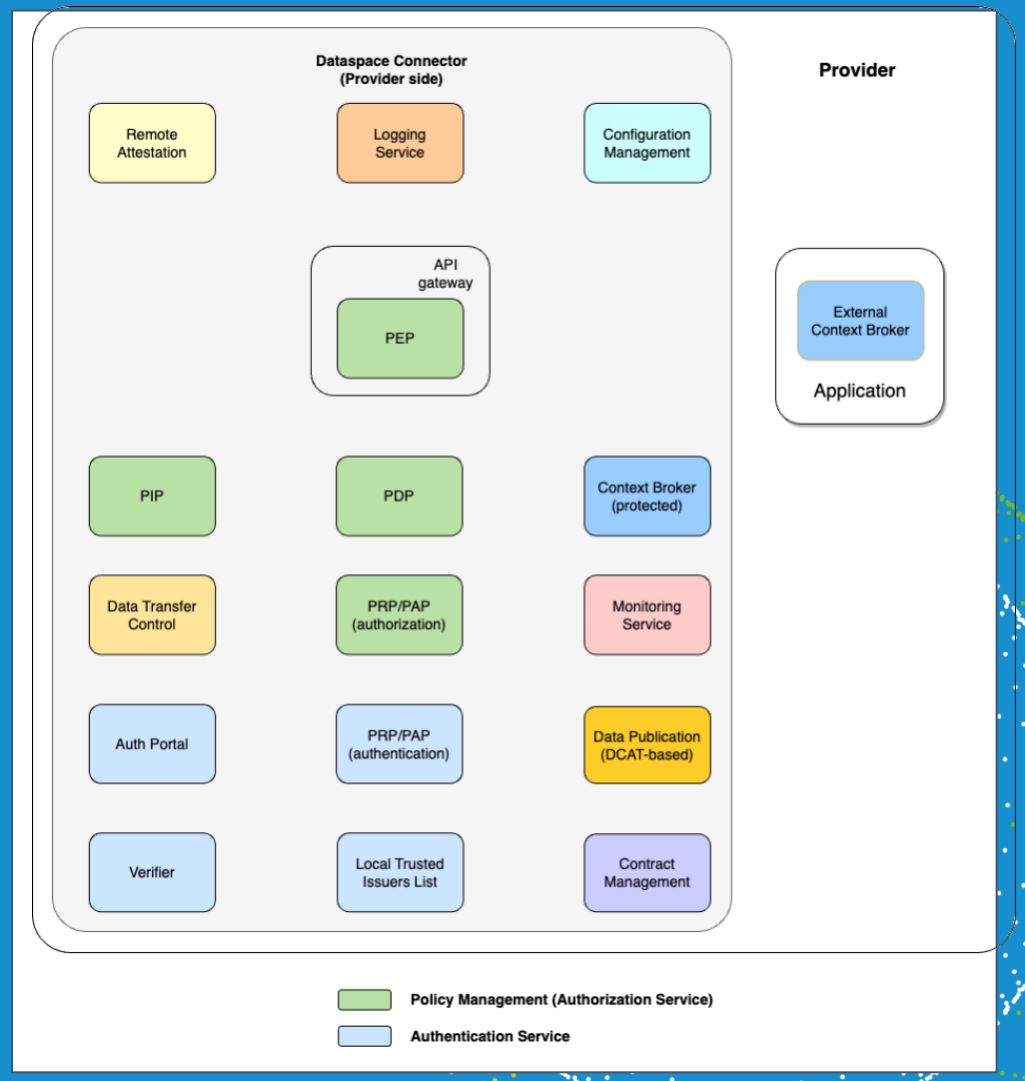
Summary

FIWARE components provides the foundations for running data space:

- Based on open source technology
- Based on already proved components
- Compliant with standards
- Beyond some limitations DCAT

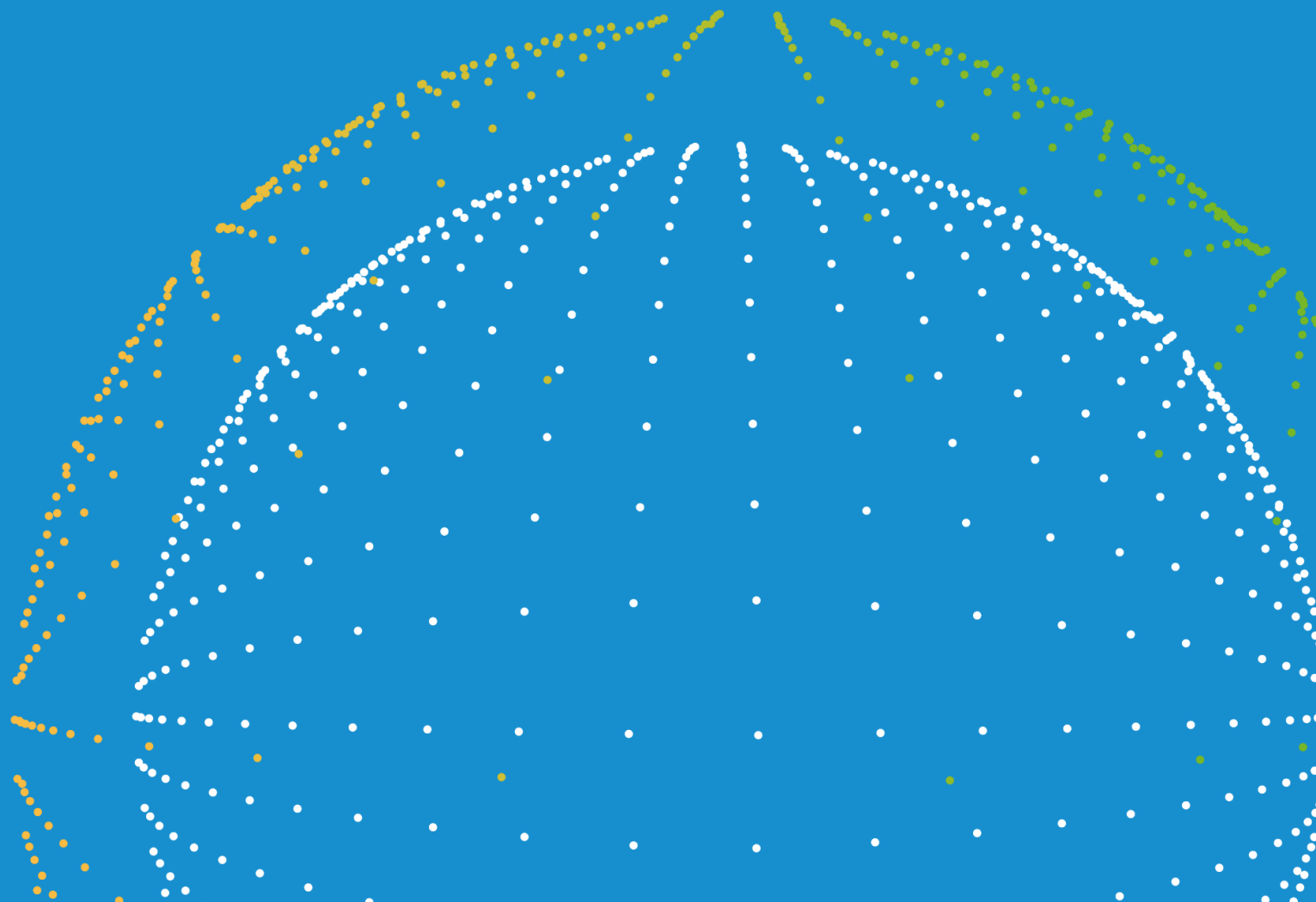
pysmartdatamodels 0.7.0

```
pip install pysmartdatamodels
```



Data Spaces Symposium

Professional Certification for Data Spaces Experts

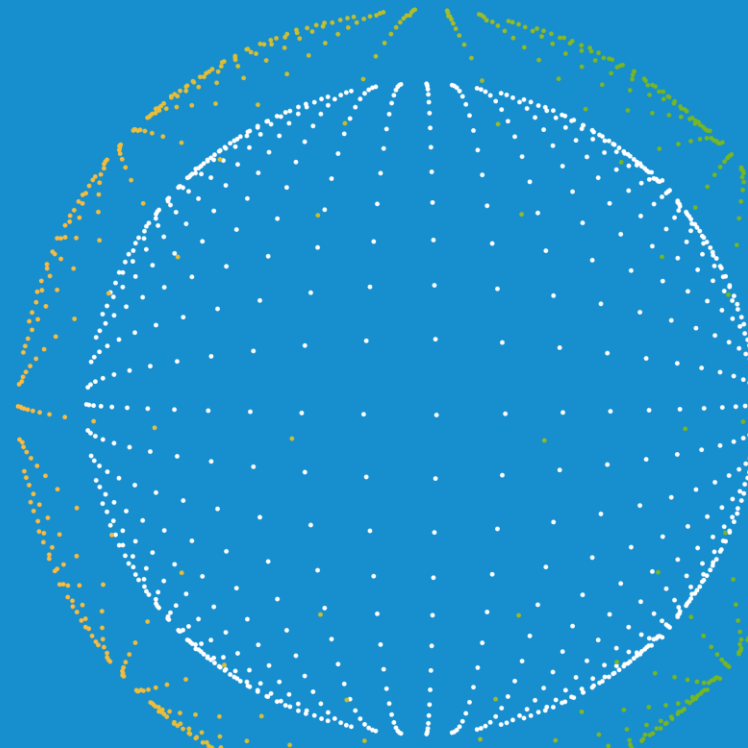


Antoine Garnier

IDSA

Skills Gap

Data spaces cannot be built without knowledge and expertise

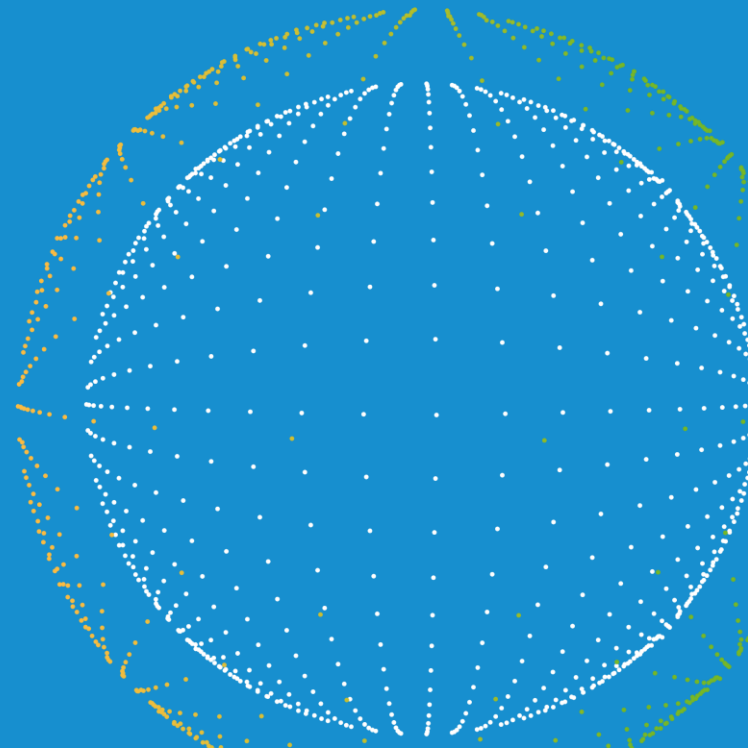


Professional Certification Programme

- IDSA is developing the **Data Spaces Body of Knowledge (DSBOK)** laying down essential information to develop professional training

- New Certifications
  - **Data Spaces Fundamentals**
  - **Data Spaces Business Consultant**
  - **Data Spaces Technical Consultant**

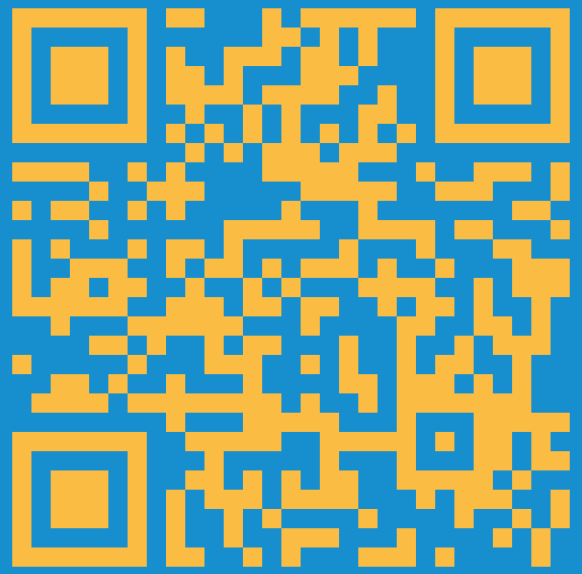
- Open to the community



Join us now!

- Join as training provider for the programme

- Keep me posted about the first training courses to be available



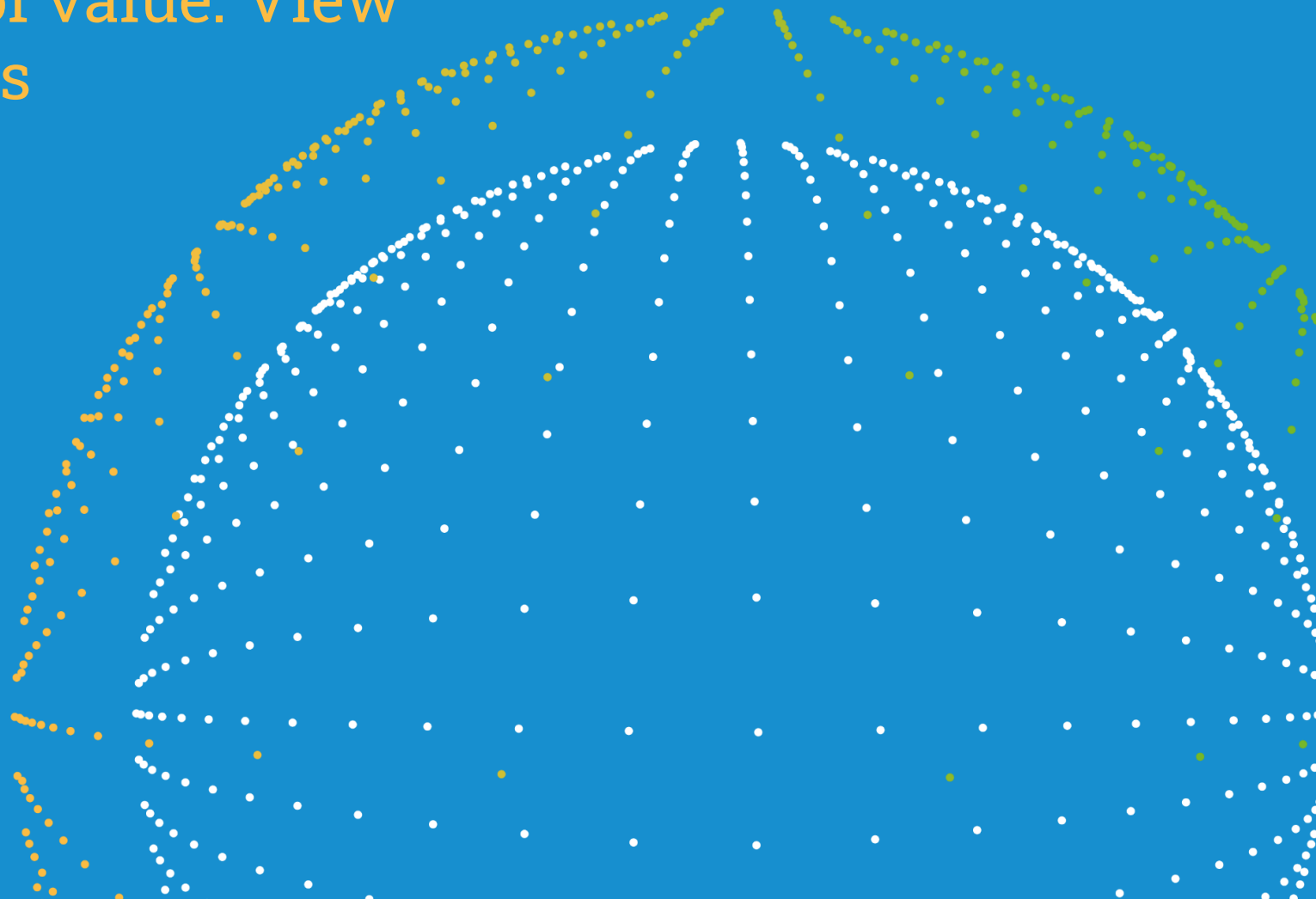
Scan me!

Scan me!

Data Spaces Symposium

How data spaces are providing value to their stakeholders, types of value. View from hubs / BDVA i-Spaces

Jeanette Nilsson RISE

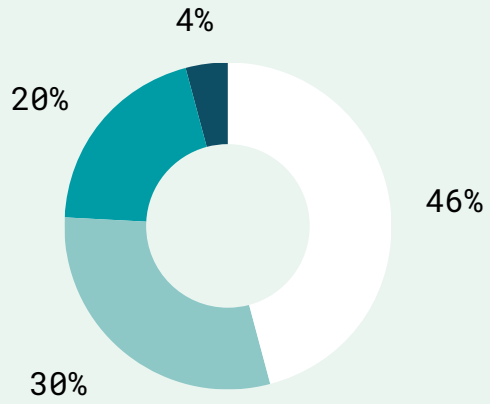


3,993

SEK million, net sales

Operating results: 22 SEK million

Operating margin: 0,6%



Distribution of net sales

Business sector	1,831 MSEK
Public funds	1,179 MSEK
State funds	812 MSEK
EU funds	171 MSEK

Nearly

3,300

employees

130+

Testbeds and demonstration environments

We are represented at

35

locations around Sweden

78

Customer Satisfaction Index

Alla siffror avser 2022.

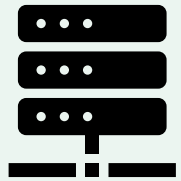
Cybersecurity

- Cyber Range testbed
- Vulnerability testing
- IoT security
- AI & Cyber
- Cyber Node

Internet of Things and 5G

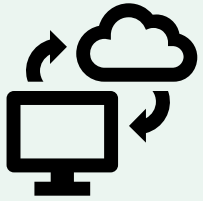
- Battery-free IoT
- Secure IoT transfer
- 6G security

Datacenter



- Datacenter technologies
- Heat reuse & energy efficiency
- Cloud & Edge testbed

Data platforms

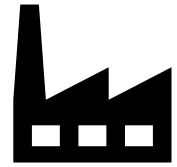


- AI & Earth observation data
- Digital twins
- Edge computing platforms
- High Performance Computing

AI and machine learning

- AI for network automation
- Resource- efficient ML
- Soundscape analysis
- Cross-lingual and Multilingual AI

Industrial data analysis



- Knowledge graphs and reasoning
- Predictive maintenance
- Causal inference
- Compilers

AI at RISE – 150+ projects, 450+ network

Center for applied AI	Computer vision	Advanced data analysis	Cyber security	Digital innovation	Transport and mobility	Energy systems	Materials and production
Deep learning	Natural language processing	Planning and scheduling	Digital ethics		Telecom	Healthcare	ICE data center
AI platforms and edge	Automated reasoning	Knowledge representa-tion	Education	Industrial automation	AI agenda	Built environment	ToD: Cyber Range
Policy innovation	Interaction design	Software engineering	Climate change	Food and agriculture	Chemistry and pharma	ToD: AstaZero	ToD: 5G/Edge

- Technology areas
- Synergistic areas
- Application areas
- Test & Demo
- Special projects

RI. SE

BDVA labelled i-Spaces 2023

BDV i-Space BRONZE

BDV i-Space SILVER

BDV i-Space GOLD

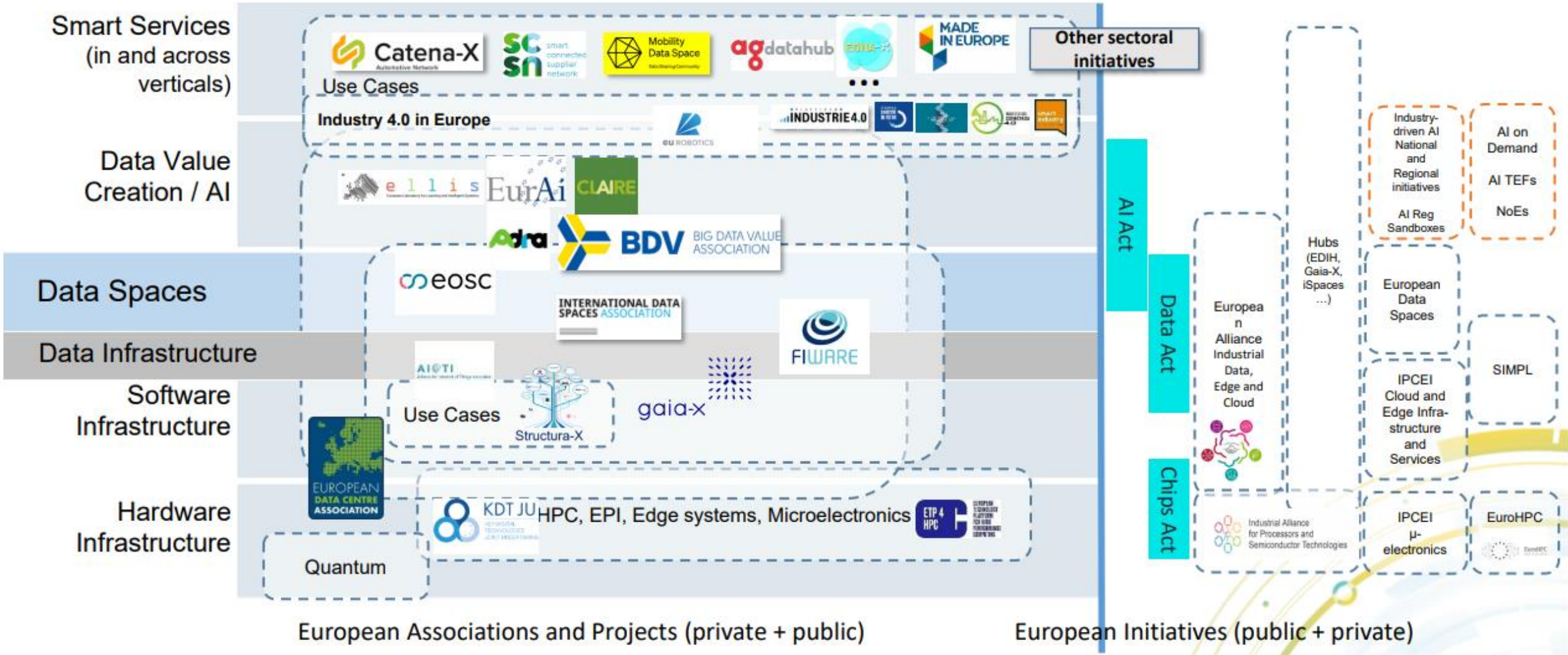
BDV i-Space PLATINUM

EUHUB

EUROPEAN FEDERATION OF DATA DRIVEN INNOVATION HUBS

BDV BIG DATA VALUE ASSOCIATION i-Spaces 2023

Strategic Digital/Data/AI related initiatives for Europe (showing technology coverage)



How to lower barriers and boost innovation in data spaces

Alberto Abella
FIWARE

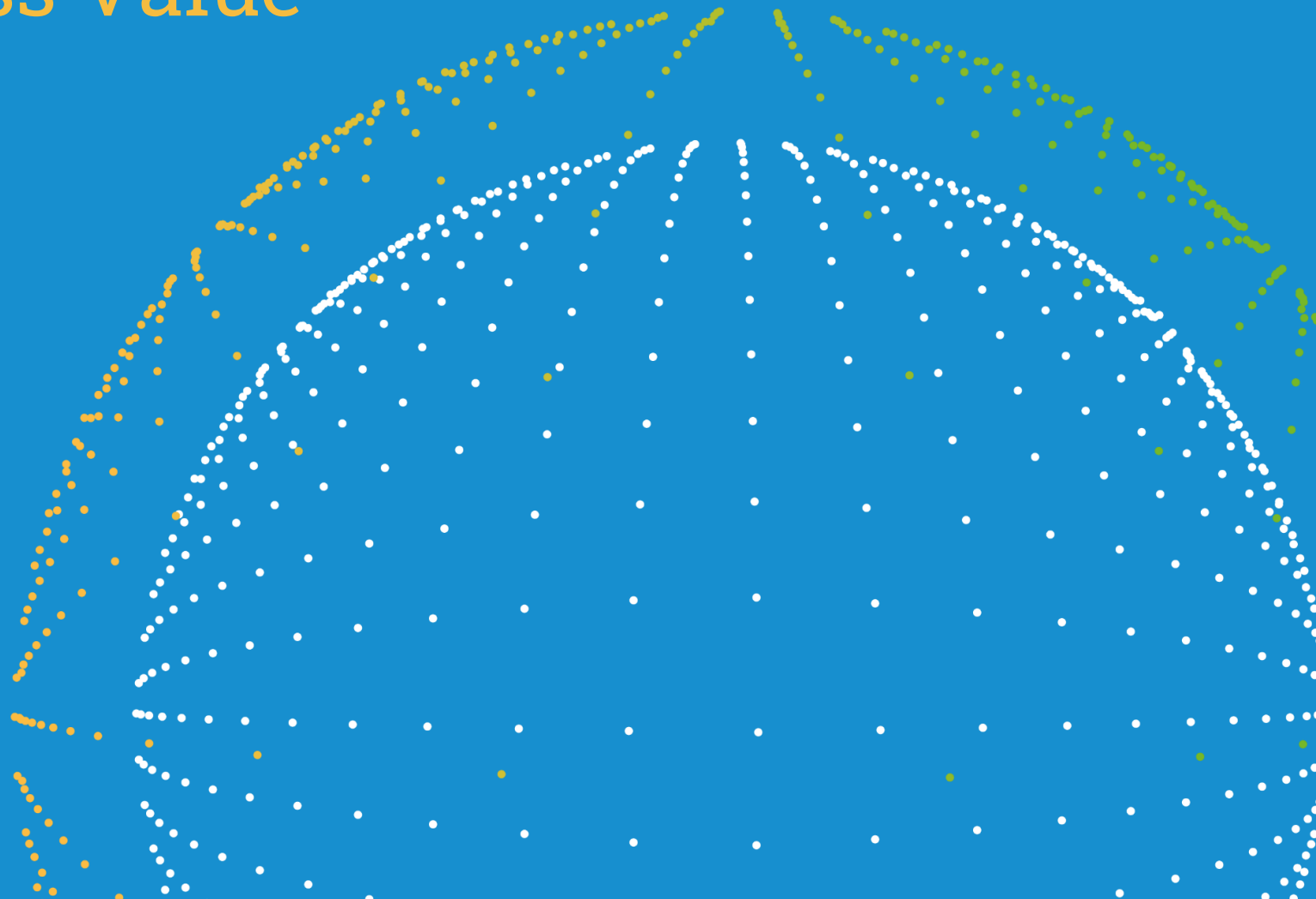
Antoine Garnier
IDSA

Jeanette Nilsson
RISE

Value creation in Data Spaces

Data Spaces Symposium

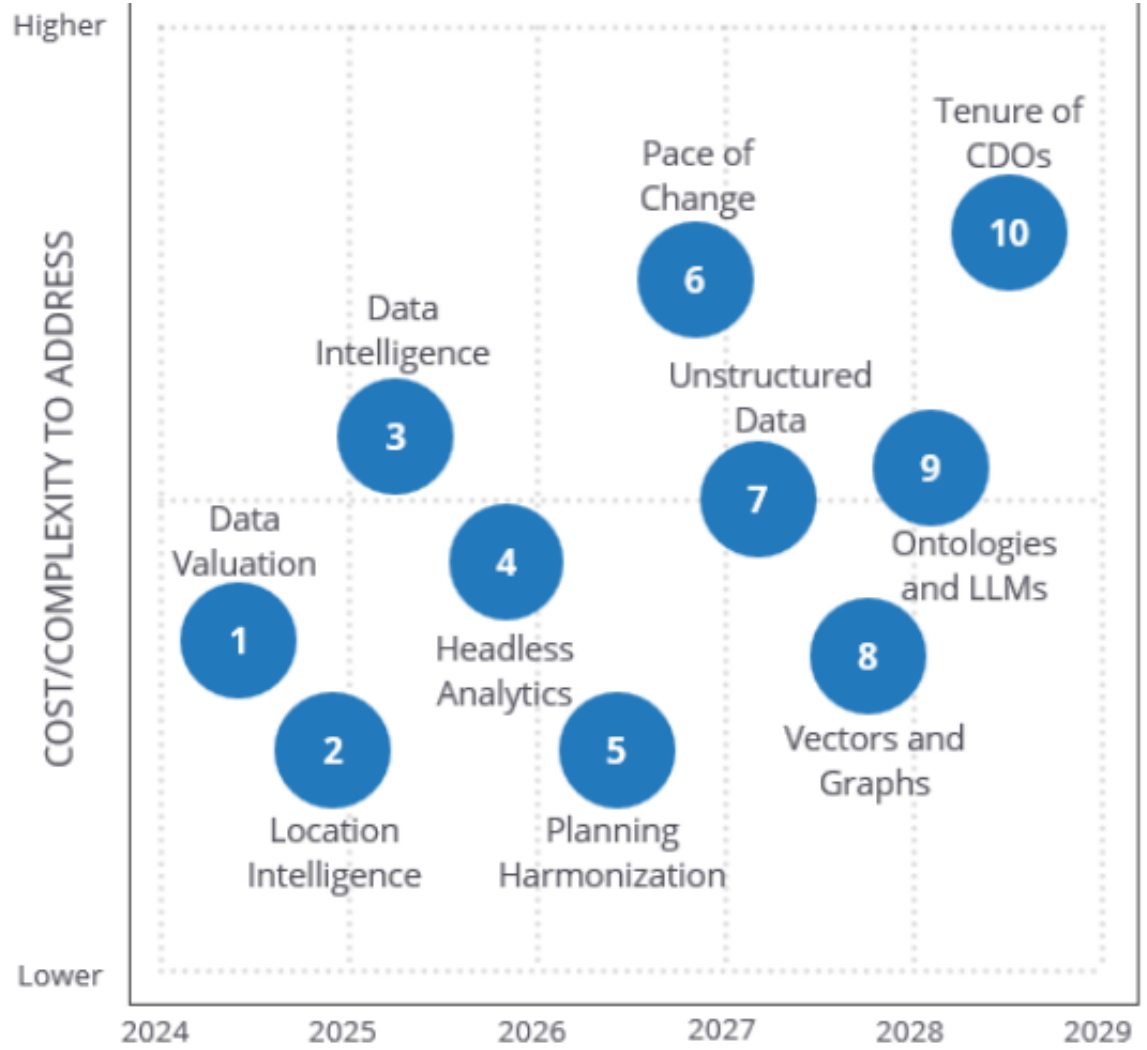
Enablers for Business Value



Nuria de Lama
Consulting Director
IDC Government Consulting

Investments to derive value from data: Worldwide Data and Analytics 2024 Predictions (IDC)

IDC forecasts that **data and analytics software spending will grow at a CAGR of 16%** through 2027 to reach close to $340 billion for both analytics and operational workloads



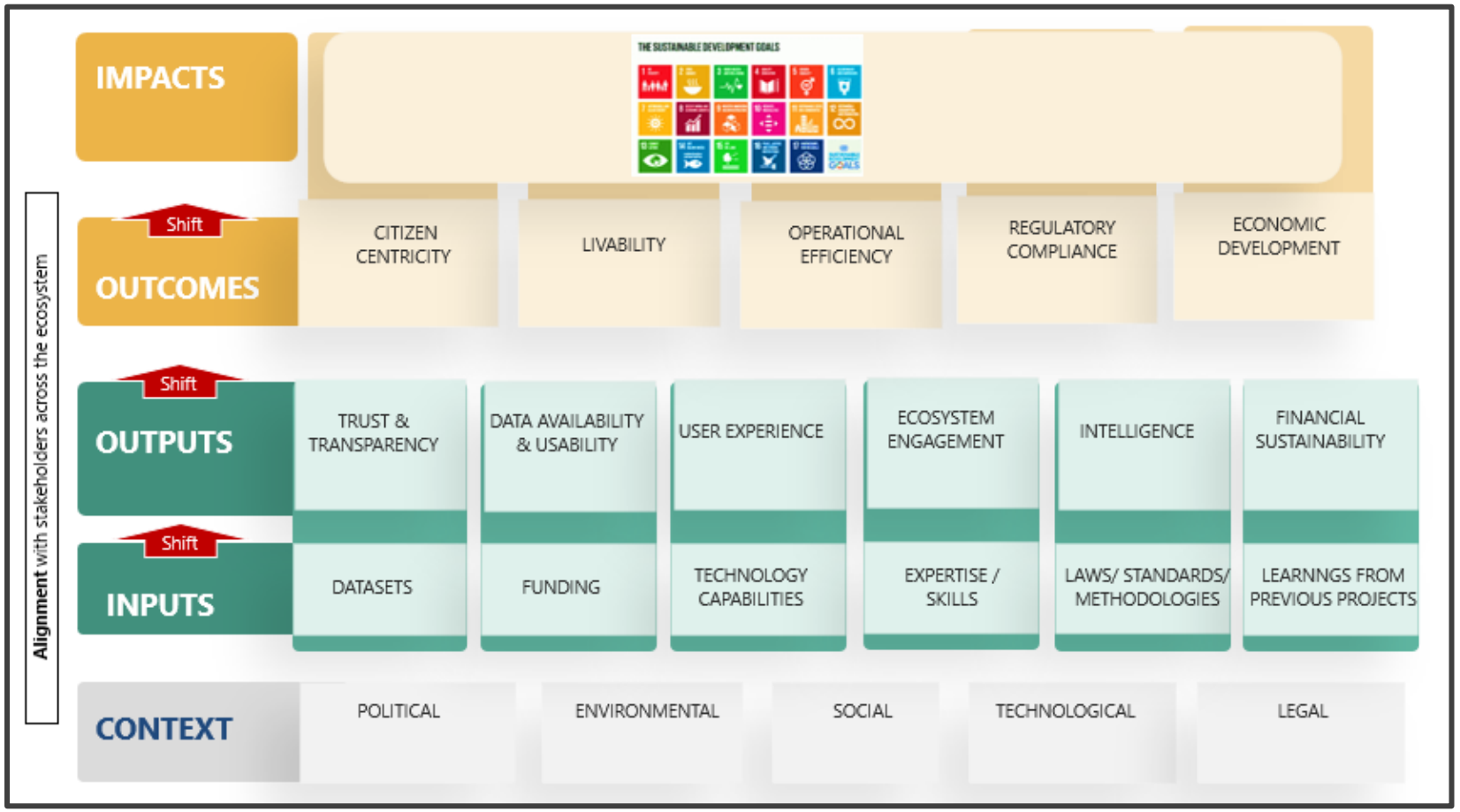
- **Prediction 1:** By 2024, data valuation initiatives will become standard in quantifying internal data, AI, and analytics project ROI and in acquisition valuations but will be hampered by inconsistent methodologies.
- **Prediction 2:** By 2025, the combination of geolocation and business analytics will be ubiquitously used by all G2000 companies, leading to greater precision and personalization of AI-enabled solutions.
- **Prediction 3:** By 2025, adoption of GenAI-driven data intelligence and integration software will result in a new automated data control plane, resulting in at least a 25% increase in data engineers' productivity.
- **Prediction 4:** By 2025, 66% of G2000 will adopt AI-driven headless BI and analytics with chat, Q&A, and proactive notification functionality, quadrupling the number of users with access to contextual information.
- **Prediction 5:** By 2026, GenAI will be deployed to spot inconsistencies across internal planning models and external economic forecasts, resulting in doubling of new cross-functional enterprise planning initiatives.
- **Prediction 6:** By 2026, the differential in the velocity of tech vendors' releases and tech users' adoption of AI-driven data and analytics software will double spending on reskilling and change management.
- **Prediction 7:** By 2027, GenAI will help equalize spending on unstructured and structured data processing and analysis software, doubling unstructured data's productive use.
- **Prediction 8:** By 2027, the need to combine dual representation of enterprise knowledge will lead 50% of G2000 to combine vector embeddings stored in vector databases with graph databases for AI model training.
- **Prediction 9:** By 2028, 75% of G2000 will use LLMs to speed development of ontologies, which in turn will guide firm-specific LLM training to enable knowledge management and decision intelligence.
- **Prediction 10:** By 2028, the tenure of the average CDO will at least double, reflecting business executives evolving understanding of the path to greater enterprise intelligence, data, and AI value creation.

Technology Capabilities and Organizational Competencies for Data Spaces

	Technology capabilities	Organizational Competencies and Capacity
Data space regulator		Data privacy Data ethics Cybersecurity Intellectual property Competition law Licensing, certification, and auditing Data platform market economic and technology analysis Orchestration of national and international communities that influence/ define data space policies and standards
Data space operator	Data ingestion Data lineage Data security Metadata management Master data management Data integration and interoperability Data exchange, data sharing Open and linked data API management User access security Digital rights management and billing	Business expertise to identity new valuable data sources and to drive innovative models on how to monetize the value of the data that they intermediate. Data quality End-user experience Data privacy Cybersecurity Data ethics Participation in national and international communities that influence/ define data space policies and standards
Data space enabler	Cloud computing IoT and edge computing Connectivity infrastructure Cybersecurity	Cybersecurity IT operations management Digital sovereignty
Data provider	Sensoring and measuring Data and event capture Data classification Device and data transfer security Data interoperability Open data	Data engineering Data quality Data privacy Data ethics
Data user	Data ingestion Data discovery Data lineage Data analysis, AI, and visualization	Data quality Data analysis and AI Data reuse regulatory compliance

Source: IDC, 2023

A view on Impact Assessment



Impacts: long-term effects produced as a result of an intervention

An **outcome** is a likely or achieved short-medium term effect of an intervention's outputs

Outputs are the products, goods and services which result from implemented change activities including new policies, legislation, technological solutions, infrastructure.

Inputs are the required set of resources needed for a transformation process

The **context** refers to specific features and dimensions to drive transformation and impact through data space use cases and initiatives

DATAWEEK 24
JOIN.LEARN.SHARE.GET VALUE

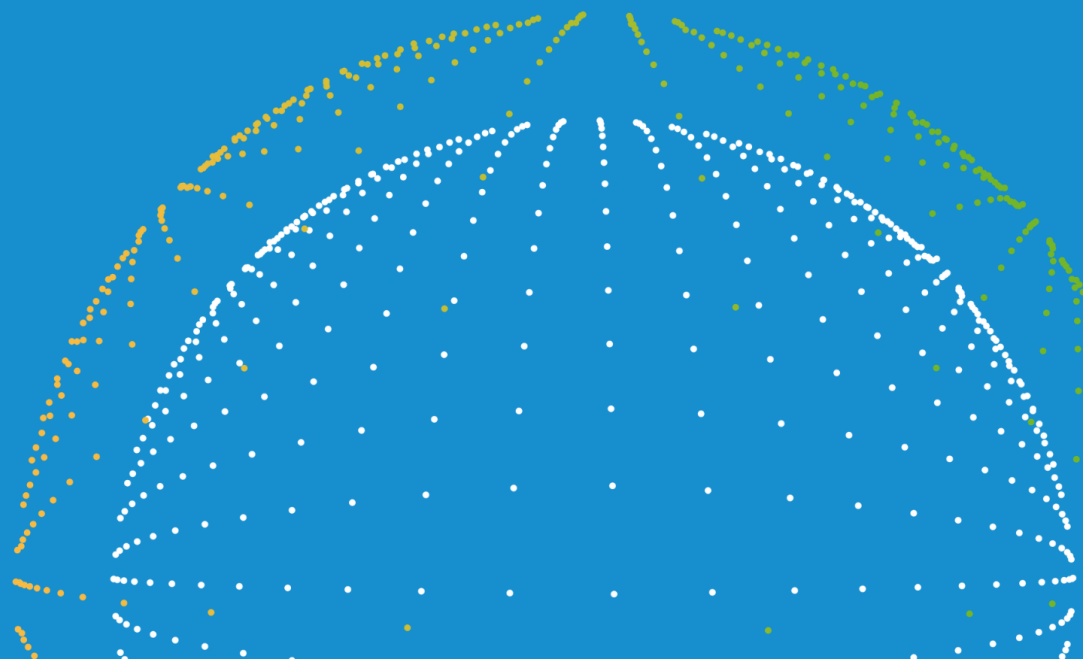
DW 24 under the umbrella of:
Data Spaces Symposium
Unite. Innovate. Adopt.
Darmstadtium | Frankfurt region

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

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

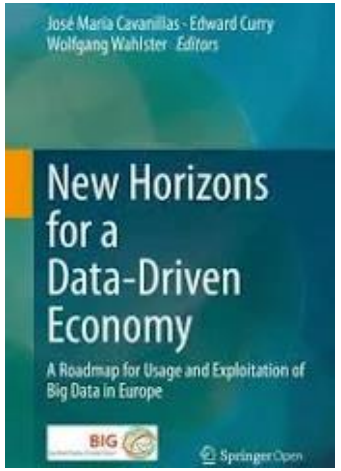
Data Spaces for GenAI & GenAI for Data Spaces

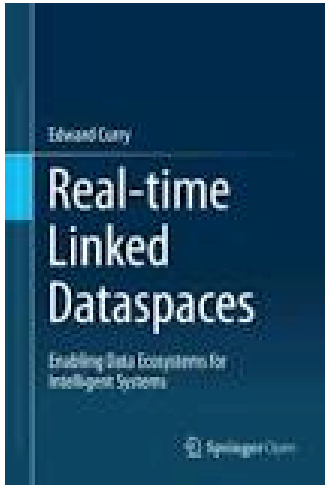
Prof. Edward Curry
University of Galway
Insight SFI research Centre for Data Analytics

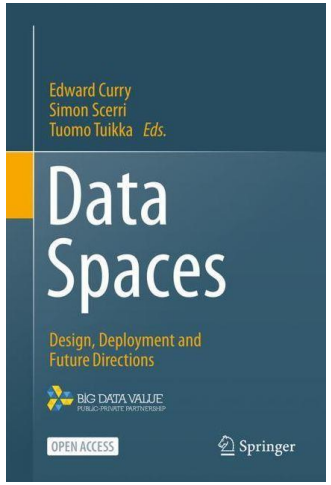


Edward Curry

I have been researching the underlying technology for data spaces for the last decade…







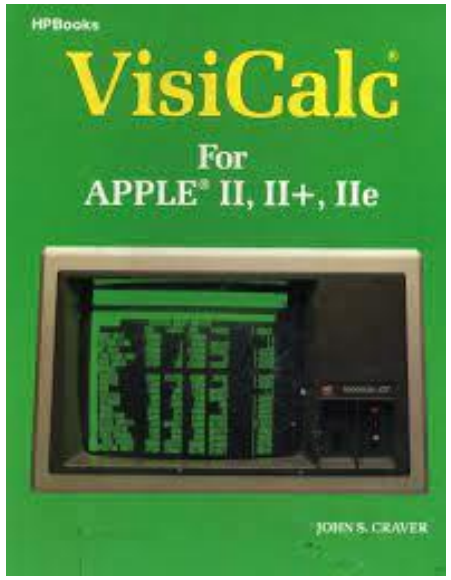
Ollscoil na Gaillimhe
University of Galway

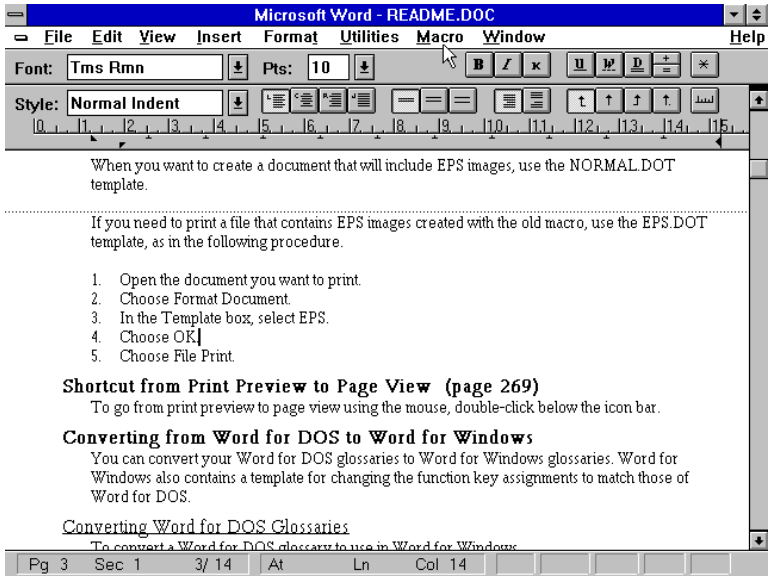
Insight
SFI Research Centre for Data Analytics

BDV
Big Data Value Association

The AI Data Robotics Association

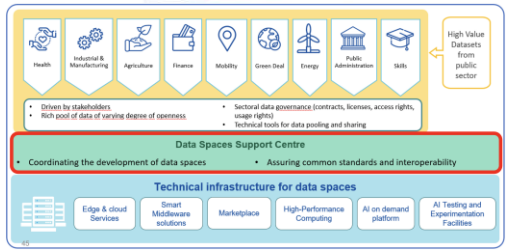
Generative AI and Foundation models will be the Killer App for Data Spaces….

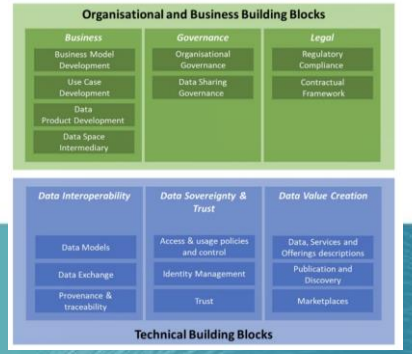




Insight

Symbiotic Relationship between Data Spaces and AI….





High Quality Data
Community of Data Users/Owners
Data Goverance/compliance

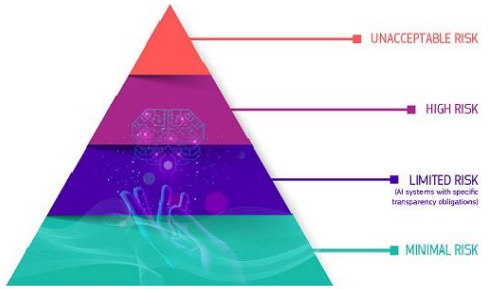
Provides training data
(Data Space 4 GenAI)

Data Space

Foundation / Generative AI

Provides models to enable
(GenAI 4 Data Space)

Extract Value from Data
Faster innovation cycles
Synthetic Data

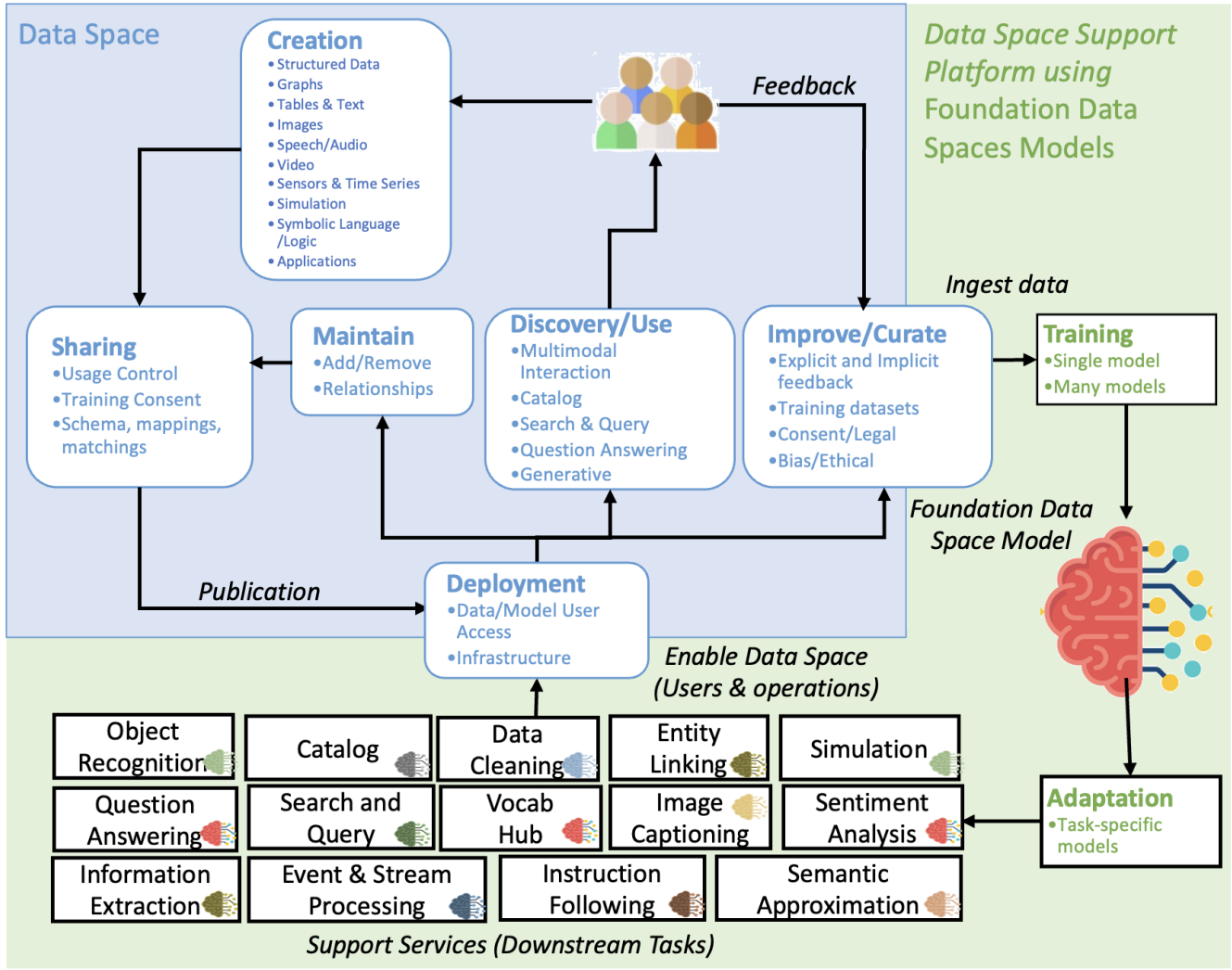


The AI Data Robotics Association

Insight

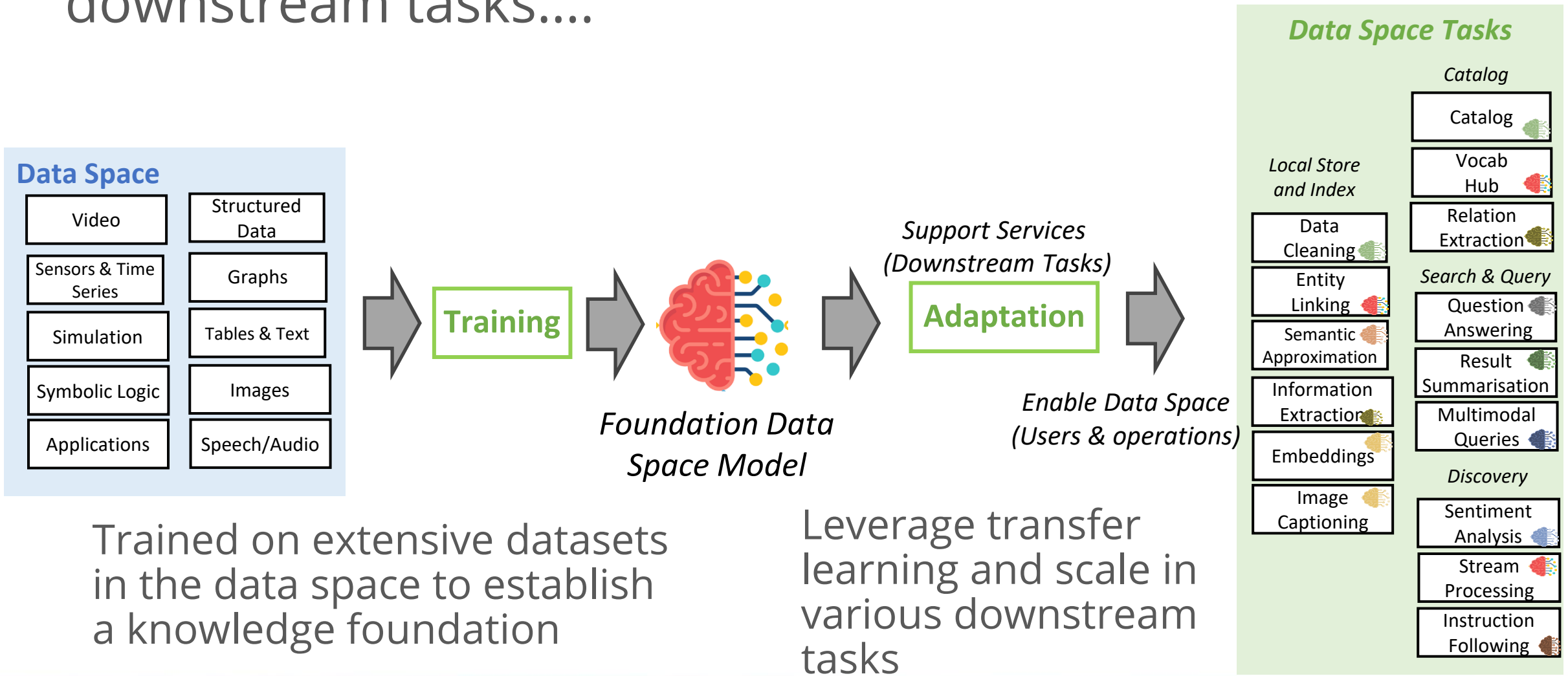
Foundation Data Space Models……

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

Data Spaces for GenAI: Centralise information from the data space, then adapt to a wide range of downstream tasks….



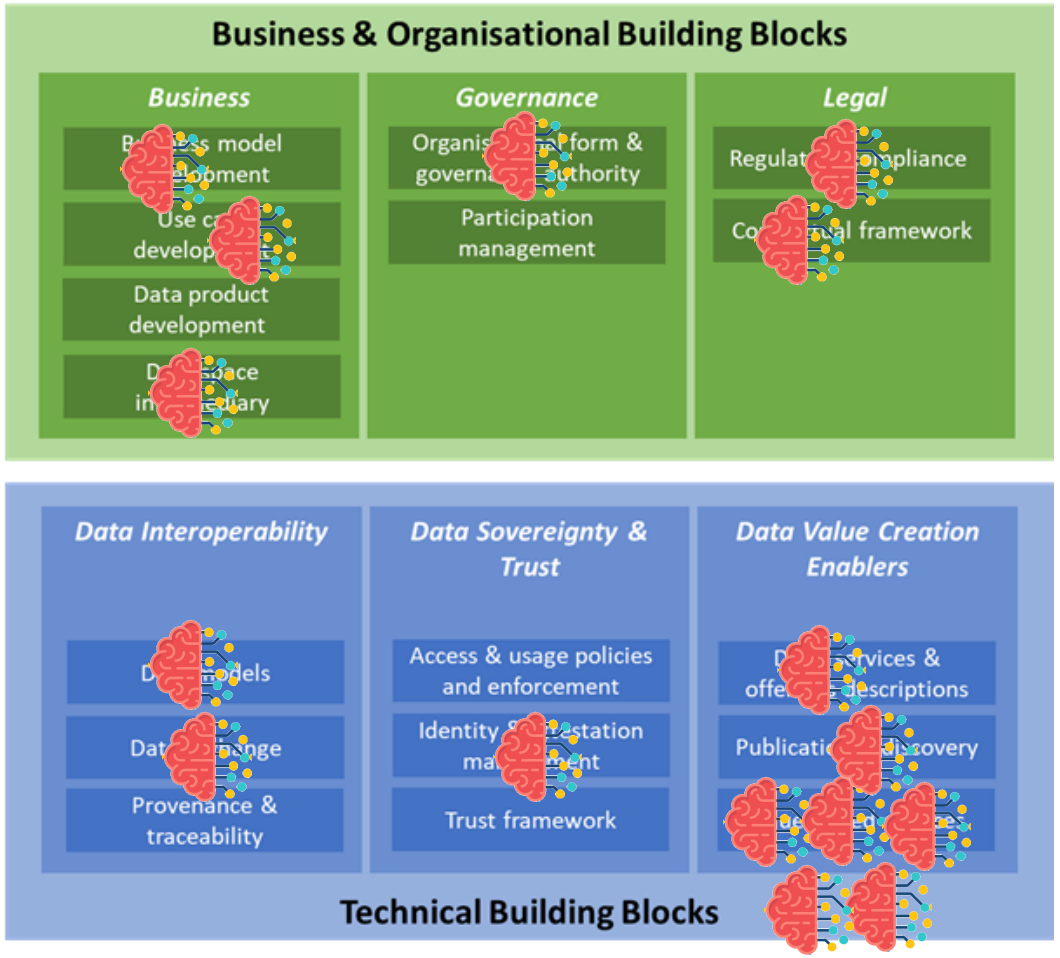
Trained on extensive datasets in the data space to establish a knowledge foundation

Leverage transfer learning and scale in various downstream tasks

Data Space Task coverage by Foundation Models

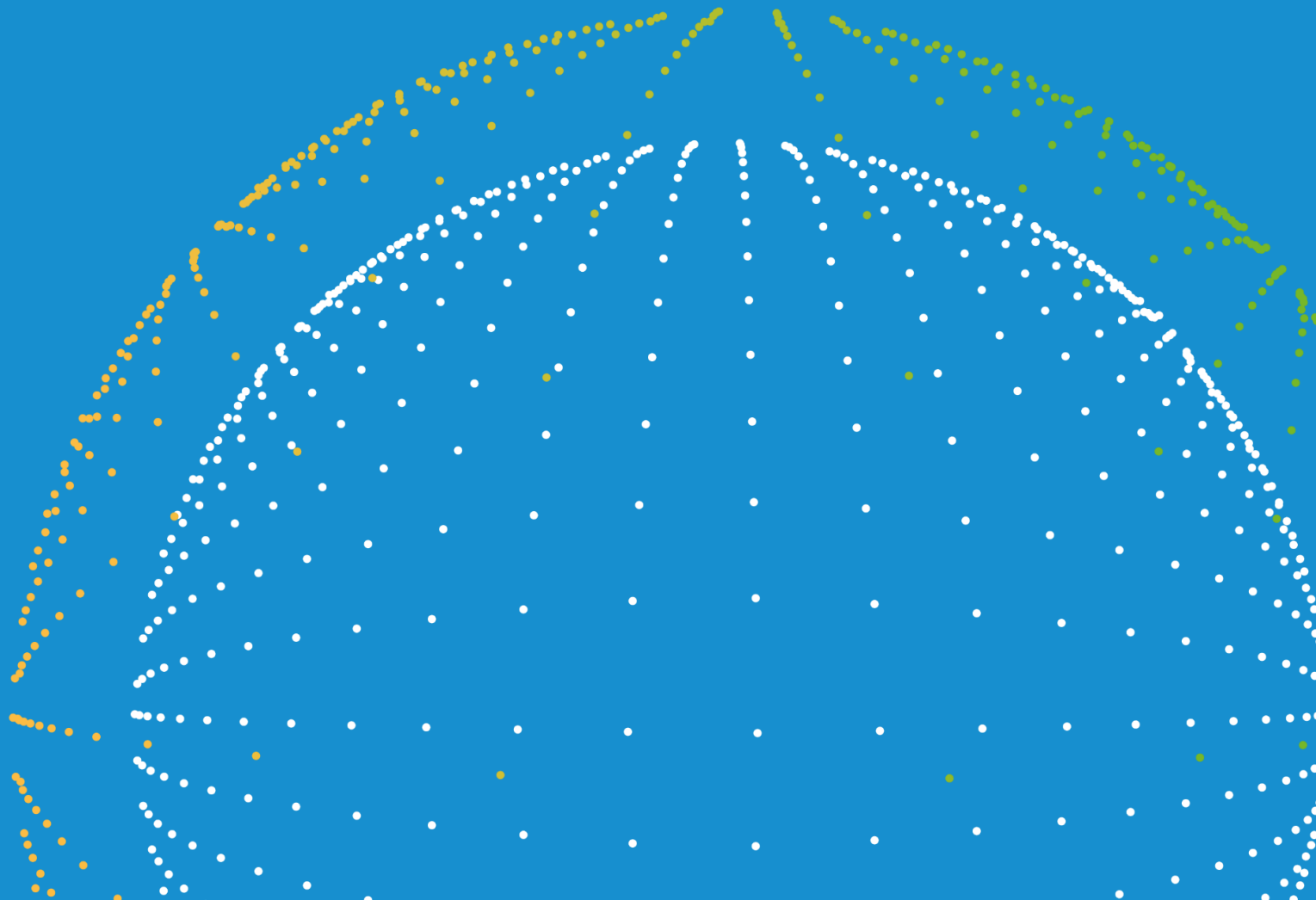
Service	DSSP Task	Adaptation to Downstream Tasks
Catalog	Metadata management	Automated metadata generation, metadata tagging and classification, multimodal metadata enrichment, Named Entity Recognition (NER), content summarization, semantic enrichment, and data relationship mapping.
	Search and Browsing	Keyword search enhancement, natural language query processing, semantic search, similarity search, multimodal search, content summaries, and question answering.
Search and query	Keyword Search	Semantic understanding, query expansion, natural language query processing, Named Entity Recognition, Contextual awareness and personalization
	Metadata queries	Metadata attribute recommendation, complex query parsing, semantic matching, metadata summaries, relation extraction, Named Entity Recognition
	Structured and complex queries	Semantic understanding, Query completion, translation, reformulation, or rewriting, complex query parsing, query planning and optimization, multimodal queries, query result summarization, question answering
	Approximate query processing	Semantic understanding, approximate nearest neighbour search, textual similarity and semantic matching
Local Store and index	Schema mapping	Word Embeddings, textual similarity and semantic matching, relation extraction, Named Entity Recognition
	Data integration	Data quality assessment, data cleaning, data transformation
	Caching and Monitoring	Caching recommendations, cache invalidation, caching content generation (e.g., summaries, responses, suggestions), event detection
Discovery	Relationship Identification	Knowledge Graph Embeddings, relation extraction, link prediction, clustering, and entity resolution and matching
	Approximate Semantic Matching	Word embeddings, query expansion, approximate nearest neighbour search, and textual similarity

GenAI for Data Spaces: *Supports tasks across the life cycle of the data space…*



Insight

Data Spaces Symposium



Georg Rehm

DFKI

openGPT-X

LLMs and Data Spaces: Observations from OpenGPT-X

Georg Rehm and Martin Courtois (DFKI)

gaia-x

Data Spaces Symposium Darmstadt,
Innovation in Data Spaces Session
13 March 2024

Gefördert durch:

Bundesministerium
für Wirtschaft
und Klimaschutz

aufgrund eines Beschlusses
des Deutschen Bundestages

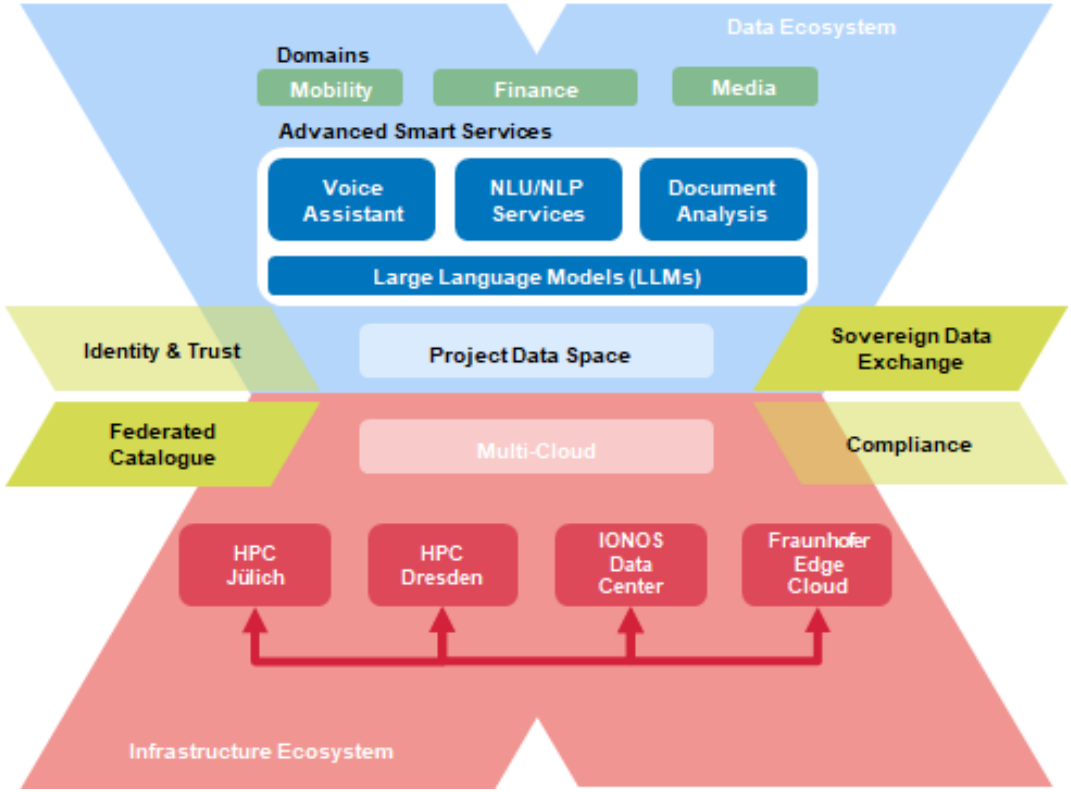
openGPT-X

multilingual – versatile – trustworthy – open source

Develop open LLMs for enterprises

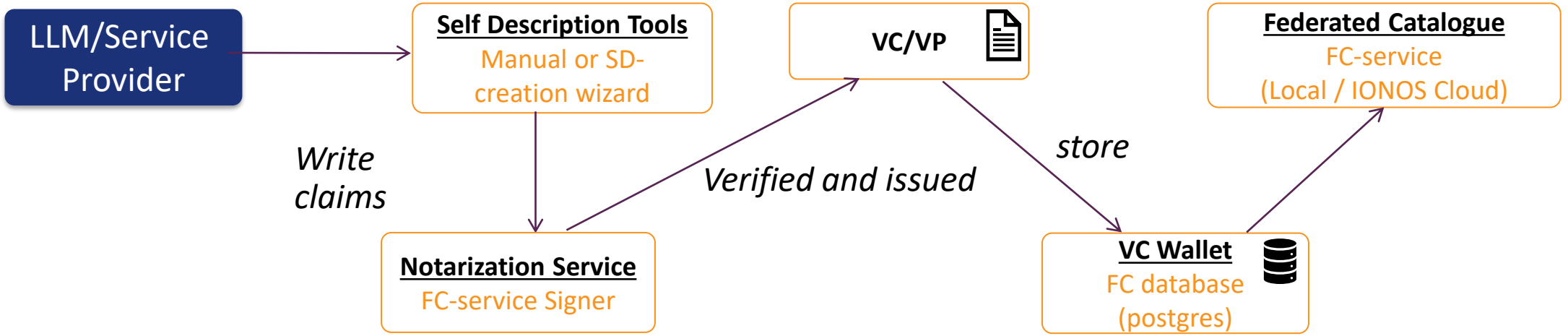
- Gaia-X project funded by the Federal Ministry for Economic Affairs and Climate Action (BMWK): 19 million euros total budget (2022 to 2024)
- Nine project partners.
- Coordinator: Fraunhofer IAIS & IIS.
- Addresses concrete business needs.
- German and European (EN, FR, ES, IT) languages

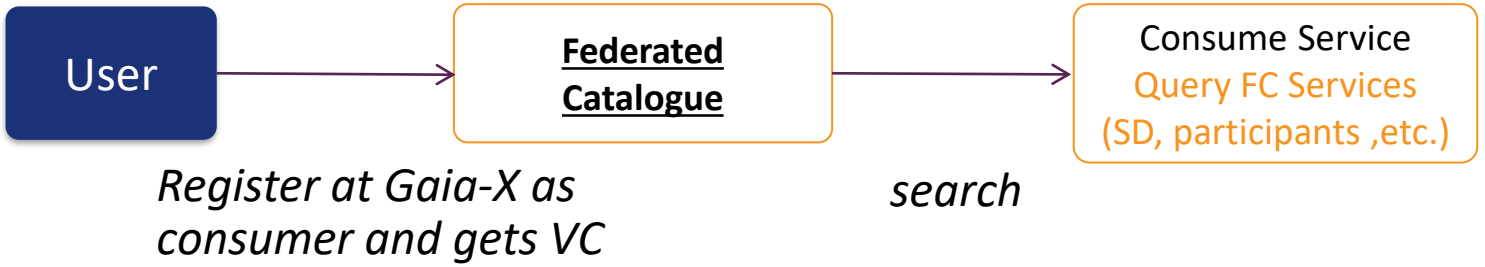
➢ Improve freedom of choice for enterprises
➢ Strengthen Germany's digital sovereignty



Fraunhofer IAIS · aleph alpha · Fraunhofer IIS · ControlExpert · ZIH Center for Information Services & High Performance Computing · JÜLICH Forschungszentrum · IONOS · WDR · dfki ai · KI BUNDESVERBAND

Technical Architecture and Components





- **VC** – Verifiable Credentials
- **VP** – Verifiable Presentation
- **LLM** – Large Language Model
- **FC** – Federated Catalog
- **SD** – Self Description

Technical Architecture and Components

Step 1: Register LLM Service Provider as a Participant

Create and ingest Self-Description to Gaia-X Federated Catalogue

- Develop Self-Description using SD-generator
- Ensure Self-Description conforms with the respective Gaia-X schema
- Validate using SHACL shapes
- Generate public-private key pair

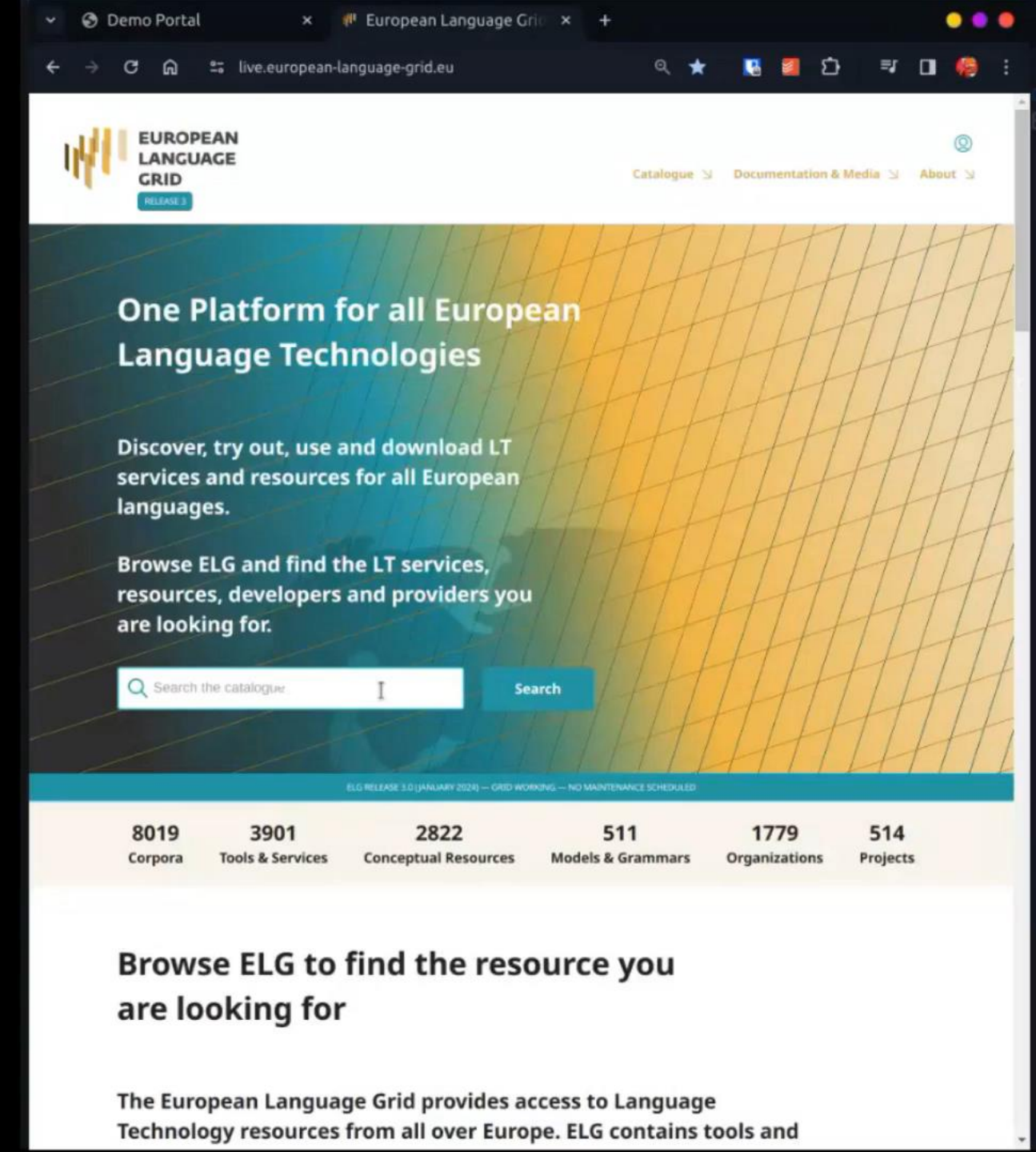
Technical Architecture and Components

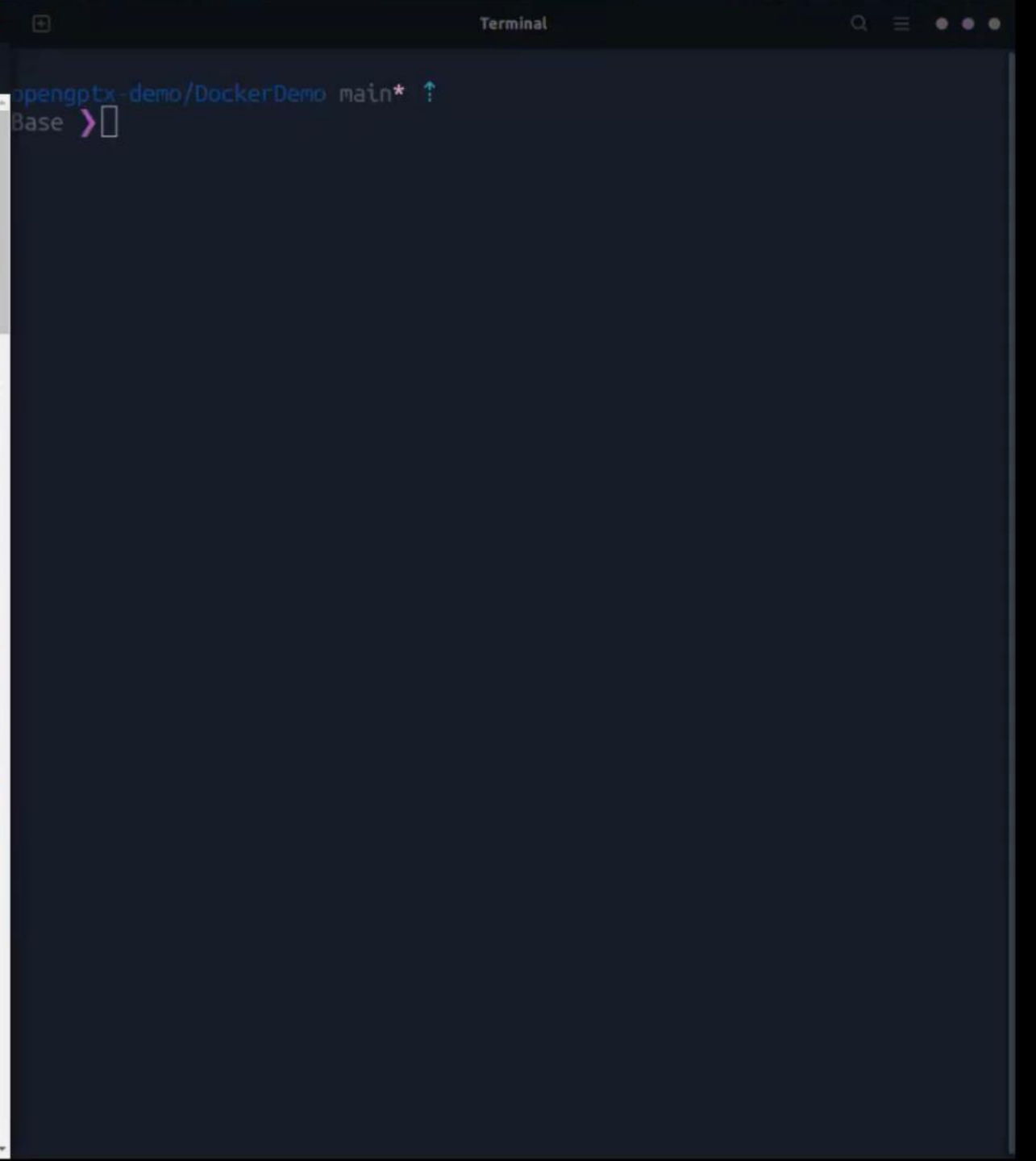
Step 2: Publish LLM Service Offering

- Create Self-Description to describe the offering
- Then verify, sign and publish SD to the Federated Catalogue to be accessed by the consumer

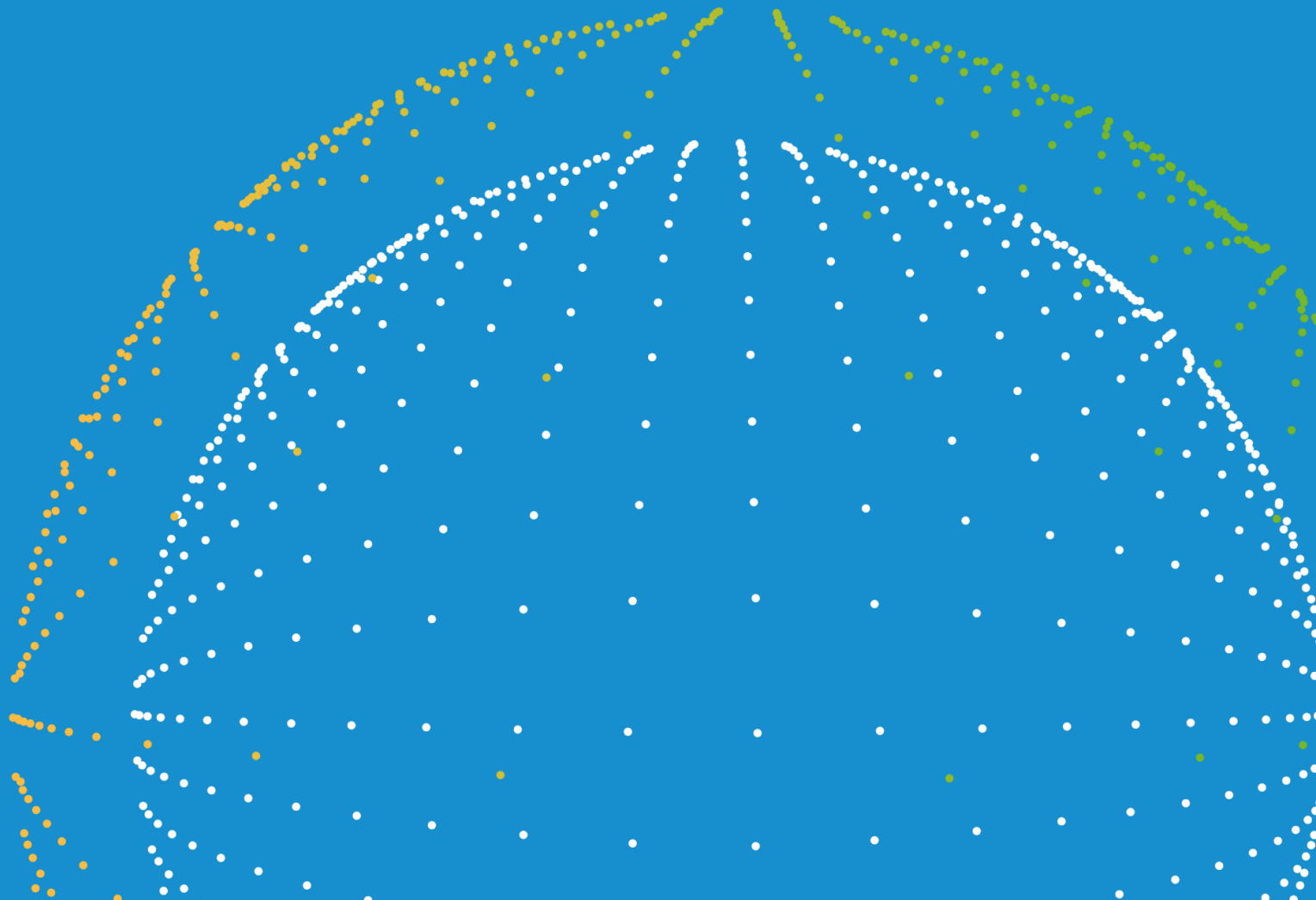
The example in the demo shows the existing Language Technology platform **European Language Grid** developed in an EU project (2019-2022), which we further extend in OpenGPT-X so that it adheres to Gaia-X.

In the example, we provision a data set in Gaia-X through ELG.


Demo Portal
European Language Gri
live.european-language-grid.eu
EUROPEAN LANGUAGE GRID
RELEASE 3
Catalogue
Documentation & Media
About
One Platform for all European Language Technologies
Discover, try out, use and download LT services and resources for all European languages.
Browse ELG and find the LT services, resources, developers and providers you are looking for.
Search the catalogue
Search
ELG RELEASE 3.0 (JANUARY 2024) — GRID WORKING — NO MAINTENANCE SCHEDULED
8019 Corpora
3901 Tools & Services
2822 Conceptual Resources
511 Models & Grammars
1779 Organizations
514 Projects
Browse ELG to find the resource you are looking for
The European Language Grid provides access to Language Technology resources from all over Europe. ELG contains tools and


Terminal
opengptx-demo/DockerDemo main*
Base

Data Spaces Symposium

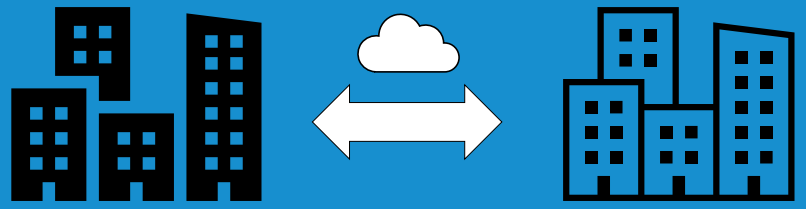


Shane Ó Seasnáin

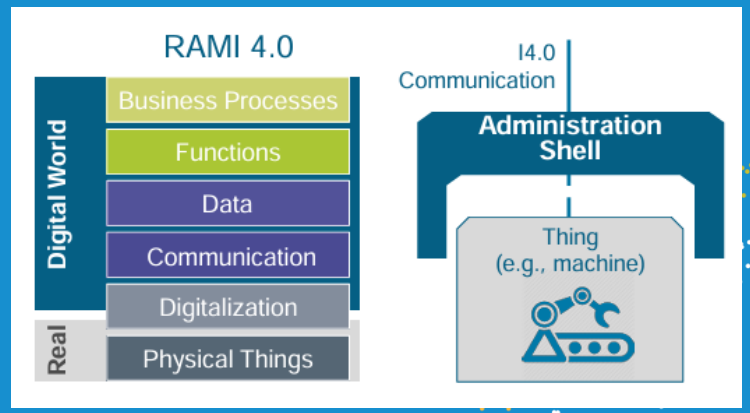
TUE

European Contribution to Technology

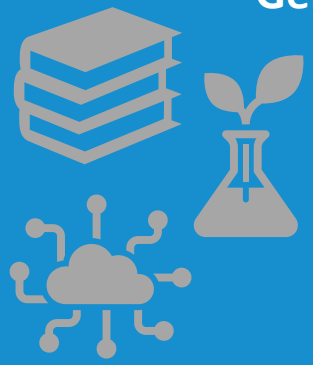
Local Digital Twins



Data Spaces

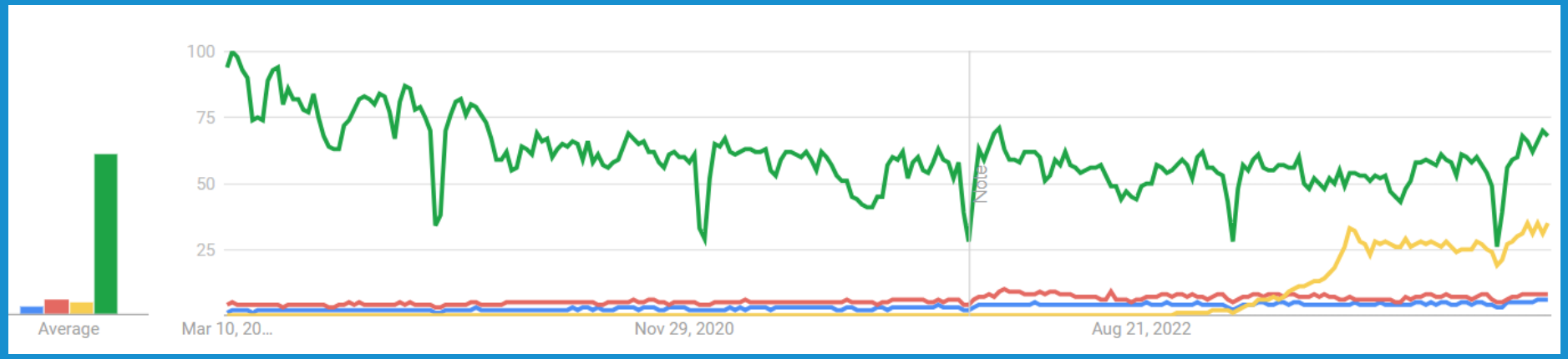


Generative AI

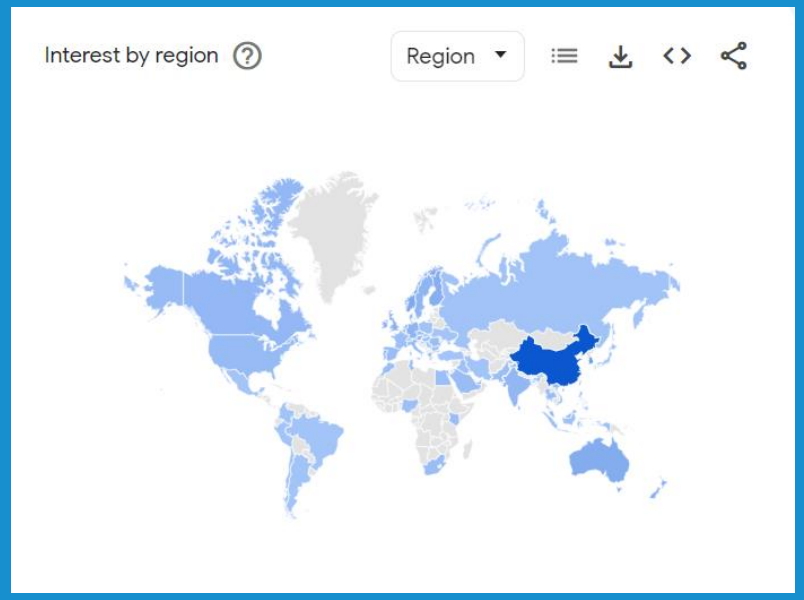


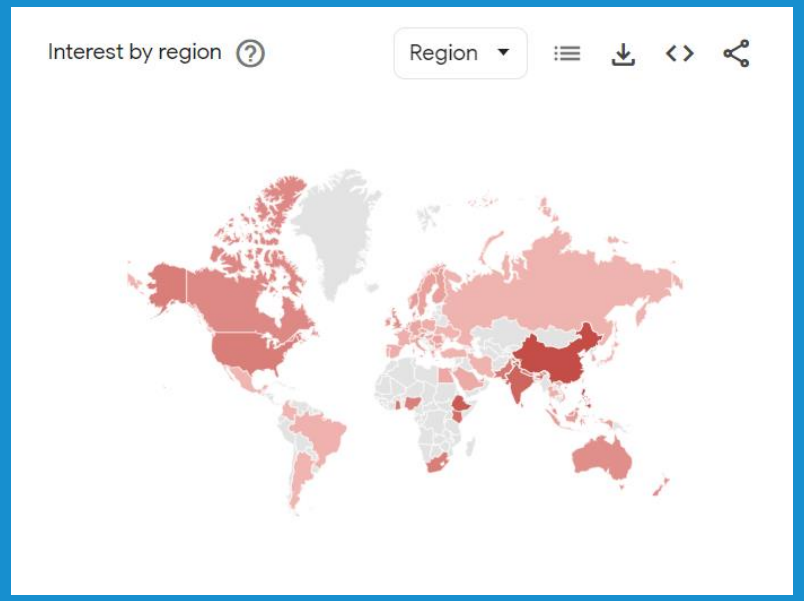
e.g. Aleph Alpha, Mistral AI

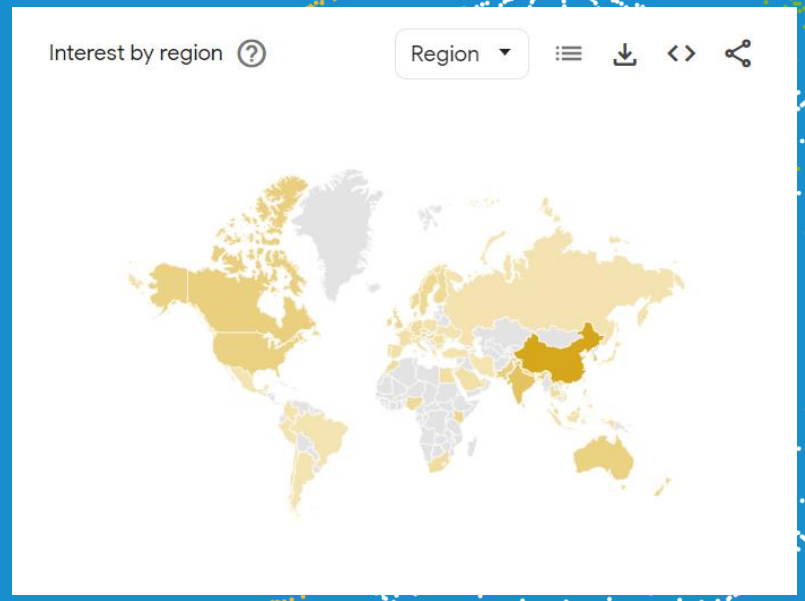
A Crisis of Awareness



- Digital Twin
- Data Space
- Generative AI
- GDPR







Major Societal Problems

Global Warming

Housing Shortages

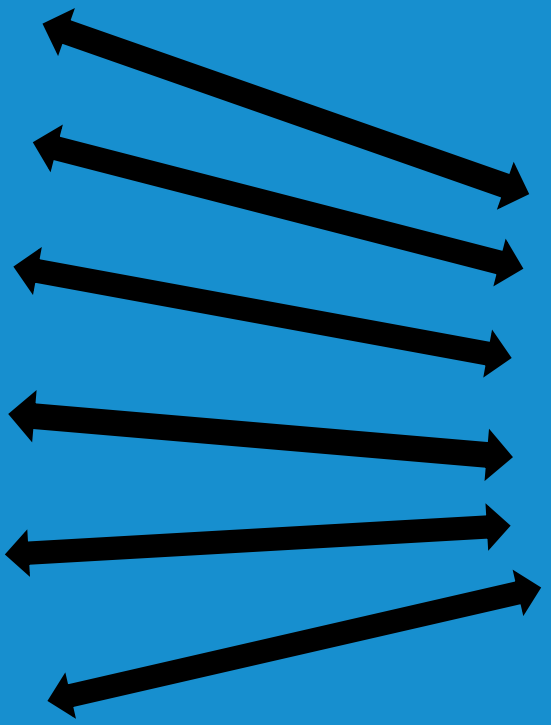
Nitrate Air Pollution

Energy Transition

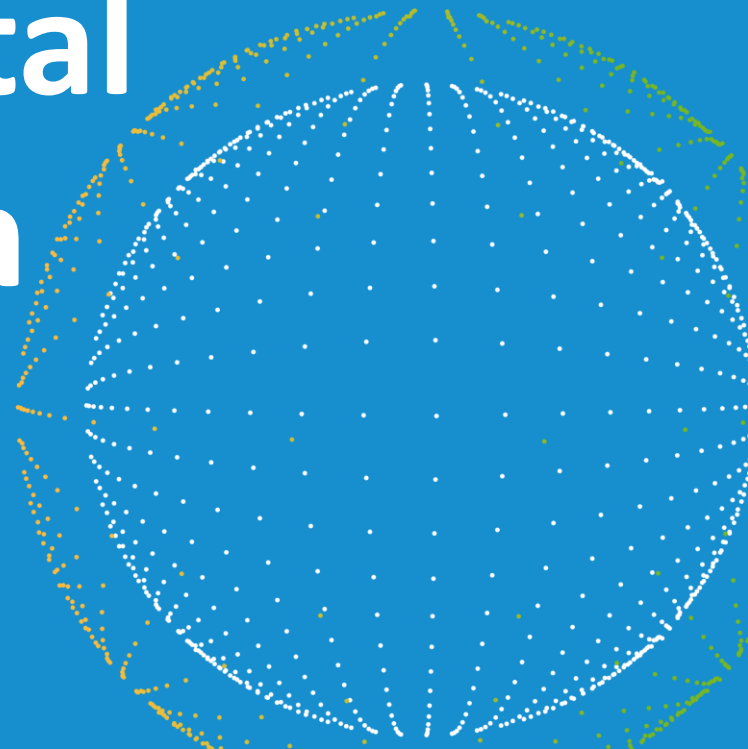
Labor Shortages

Strategic Security

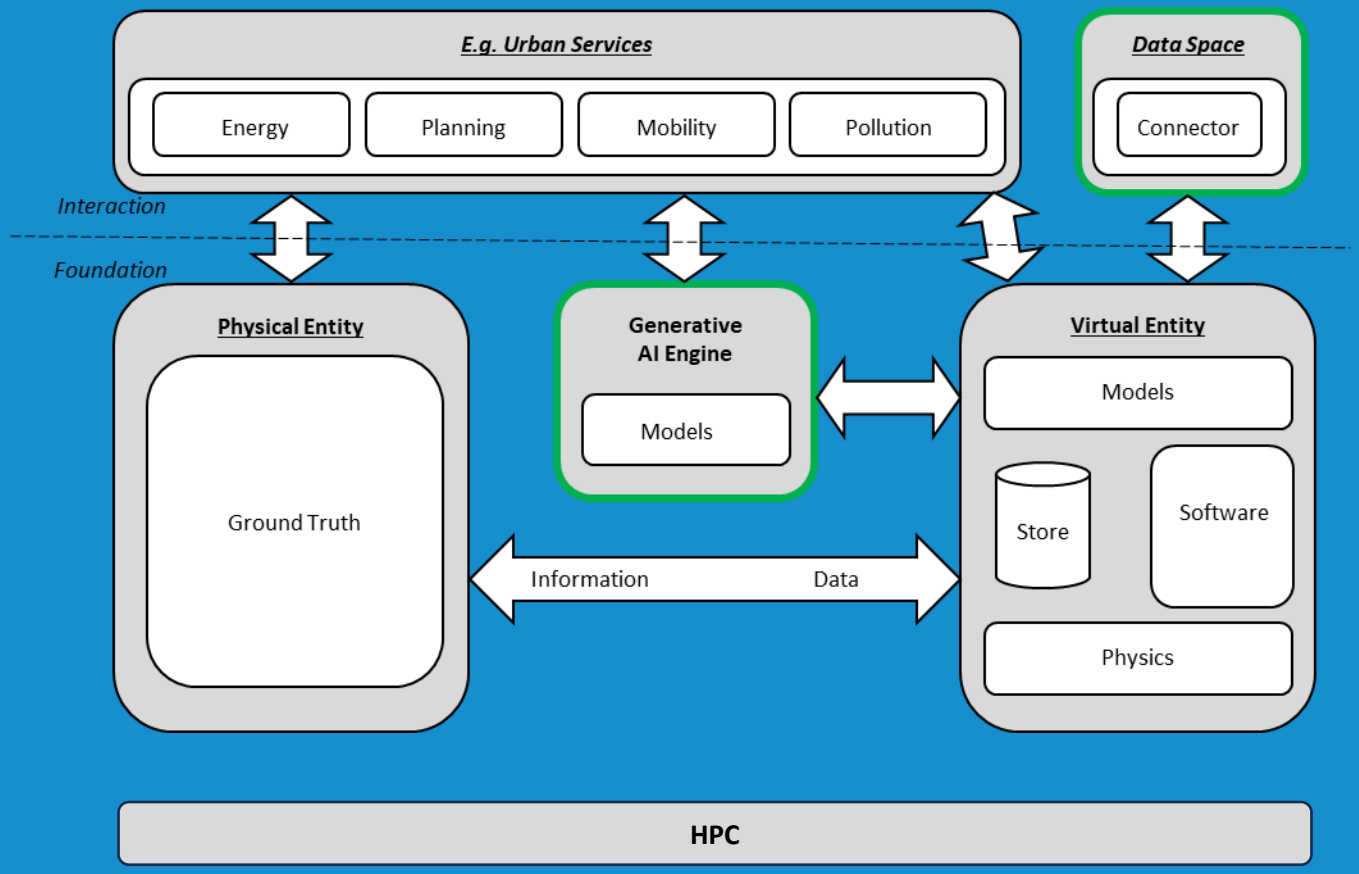
...



Digital Twin

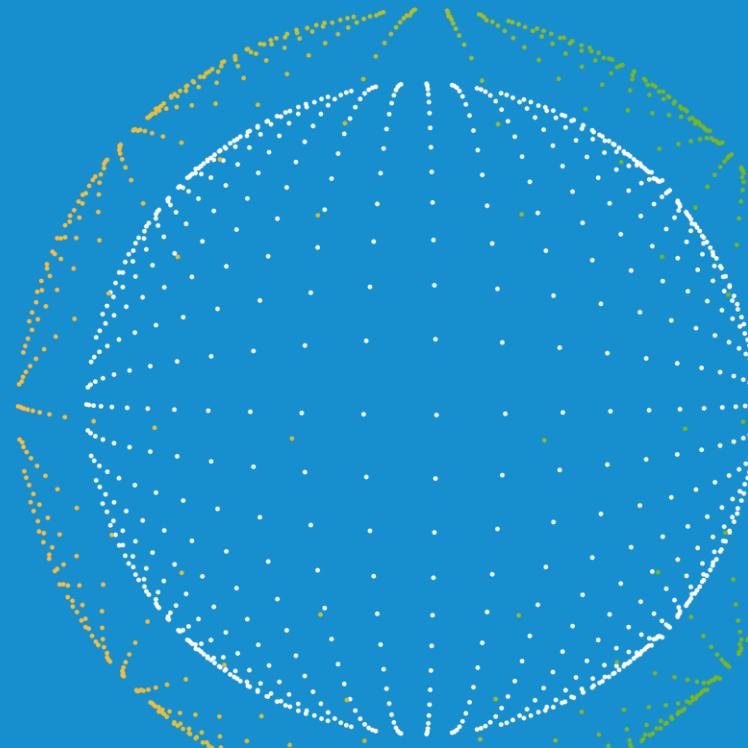


Making Twins Simpler Through Data Spaces and Gen AI



Data Space as a Service:
- Feature Extraction
- Version Management
- Public Data Spaces

Generative AI Engine:
- Generated (Surrogate) Twins
- User Flexibility
- Thinking with the User



Value creation in Data Spaces

Shane Ó Seasnáin
TUE

Nuria De Lama
IDC

Ed Currry
Insight

Georg Rehm
DFKI

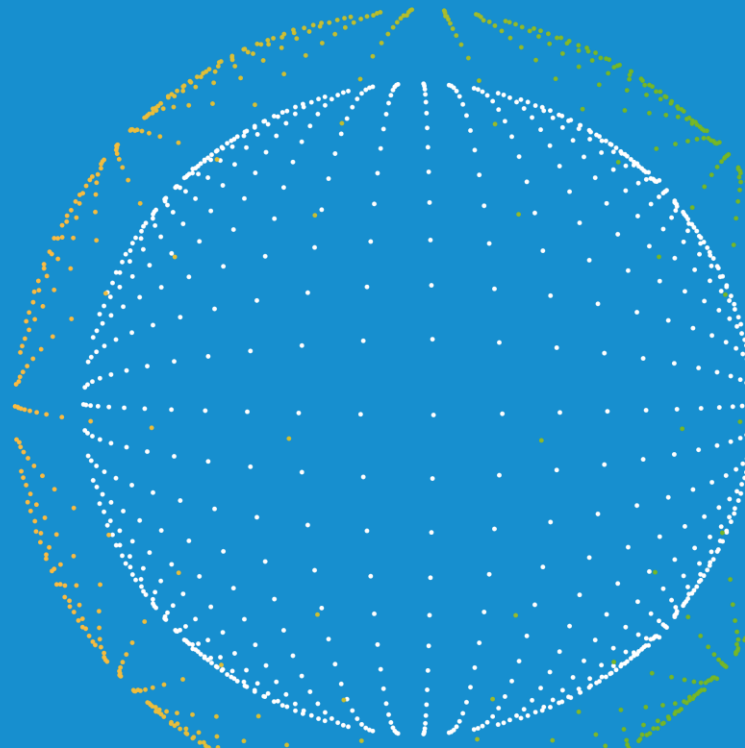
Data Spaces Symposium

Unite. Innovate. Adopt.

Thank you!

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