
DATA FOR AI

The EC's Horizon Europe/Digital Europe Programmes and BDVA's standpoint



LAMBDA Big Data Analytics Summer School
<https://project-lambda.org/BDA-Summer-School-2020>



16.06.2020

DATA FOR AI

■ Why is it important?

- Data access bottleneck: Risky Sharing
- The European Commission's (EC) Data Strategy

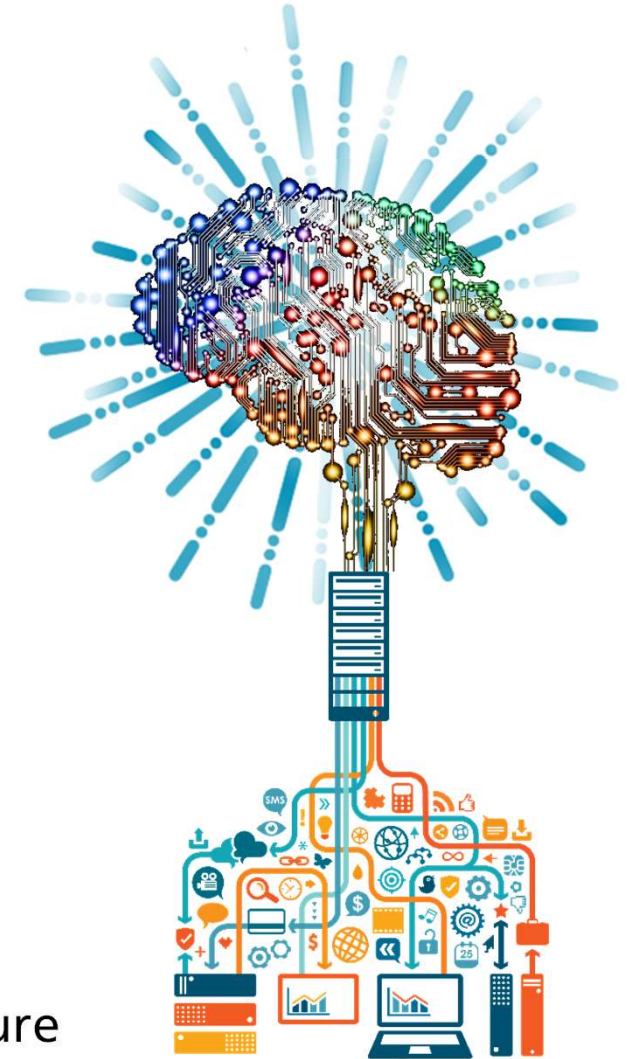
■ What is the vision and which actions must be taken?

- Actions that need to be taken to realise vision
- The Big Data Value Association's (BDVA) Recommendations

■ How could it be technically achieved?

- Secure and Trusted Data Sharing Space
- The International Data Spaces Association's (IDSA) Reference Architecture

■ Realising the Vision – Research Focus



WHY IS IT IMPORTANT?

WHY?

- Greater access to Data
- Democratic AI: Europe's Global Position
- The European Commission's (EC) Data Strategy



Turbo-powering AI: Greater access to Data

■ The Need

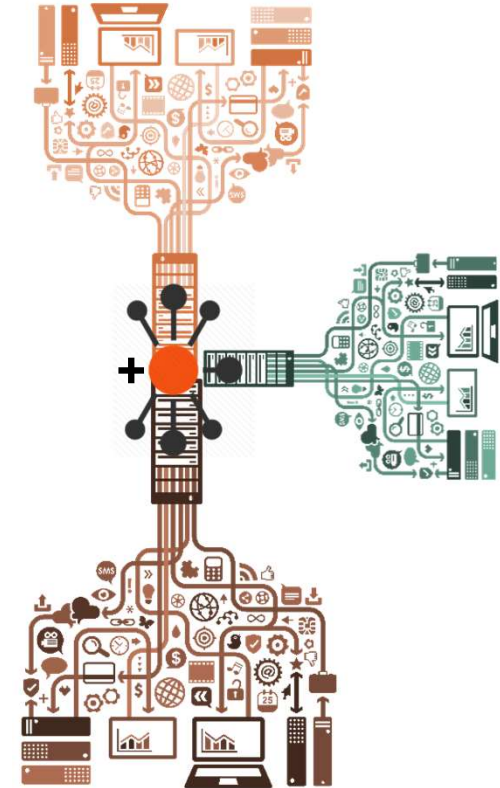
- Limited (Access to) Data = Limited AI
- Data exchange amongst multiple parties to enable more powerful AI
- Mutual benefit and new applications across regions, sectors

■ The Difficulties

- Widespread Reluctance to Share Data
- Lack of Safe and Trustworthy ***Data Sharing Space***

■ Issues to Address

- **Of Major Concern:** Trust, Legal Compliance, Data Governance, Ownership, Workforce Skills
- Technical Maturity not the major concern, but needs further investment: Provenance, Interoperability, Decentralisation and Scalability, Secure Access, etc.



Democratic AI: Europe's Global Position

■ Major Weaknesses in view of Global Competition

- Smaller industrial players, limited data processing & storage solutions
- Safeguards and Restrictions: *"Europe is too Fair, Safe, Ethical!"*

■ Turning Weaknesses into Strengths

■ Strength in Diversity

- Larger share of SMEs, Startups
- All own valuable data
- European stakeholders must join forces!

■ Strength in Regulation and Compliance

- Example: GDPR Compliance = Safer, Less Risk in Sharing Data
- Entities more comfortable sharing data



EC Data Strategy (2020)

■ The View on Data

- Data crucial for AI to improve our lives
- Data can transform all sectors of the economy
- Data is a source of innovation for new products and services

■ A “European Single Market for Data”

- Enable Free-flow of Data across regions and sectors
- Achieve greater access to data for AI innovation
- Respect European Rules, Values
- Establish clear Data Governance for fair access and use of data

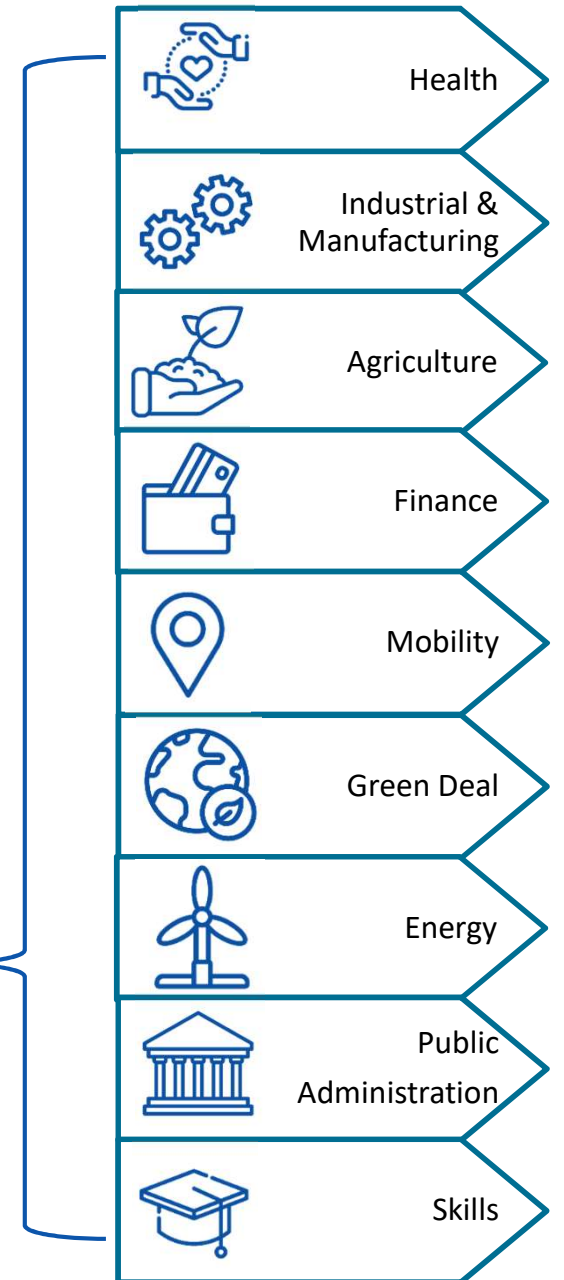


Source: Yvo Volman, Head of Unit G1 - Data Policy and Innovation, DG CNECT, EC

EC Data Strategy (2020)

■ Strategy Deployment Plans

- 4 Pillars: Balancing 'free-flowing' data with European values & rules
- Rollout of common European data spaces in crucial economic sectors
- In line with the Horizon Europe & Digital Europe Programmes 2020-27



Source: Yvo Volman, Head of Unit G1 - Data Policy and Innovation, DG CNECT, EC

WHAT IS THE VISION AND WHICH ACTIONS TO TAKE?

WHAT?





- The BDVA Vision 
- Widespread Data Sharing: Challenges to overcome
- How-to enable conditions for a Data Sharing Space?
- Actions by Stakeholders
- BDVA Recommendations

Challenges to Widespread Data Sharing

■ Identified Challenges

- Consultation with BDVA members, relevant associations, organizations and initiatives
- Confirmed: Biggest concerns are around non-technical issues



 Technical	 Legal Compliance
<ul style="list-style-type: none">• Sharing by Design• Ownership• Decentralisation• Veracity• Security• Privacy Protection ←	<ul style="list-style-type: none">• Data Protection• Free-flowing Data• Privacy preservation• Regulatory Compliance ←
 Business & Organisational	 National & Regional
<ul style="list-style-type: none">• EU Values• Global Competition• Dynamic Ecosystems• Dynamic Skills• Digital Transformation• Trust ←• Valuation Standards	<ul style="list-style-type: none">• Workforce Skills ←• Resistance to Change• Investment Evaluation• EU-wide policies• Policy Compliance

Source: Towards a European Data Sharing Space - Position Paper (April 2019) <http://www.bdva.eu/downloads>

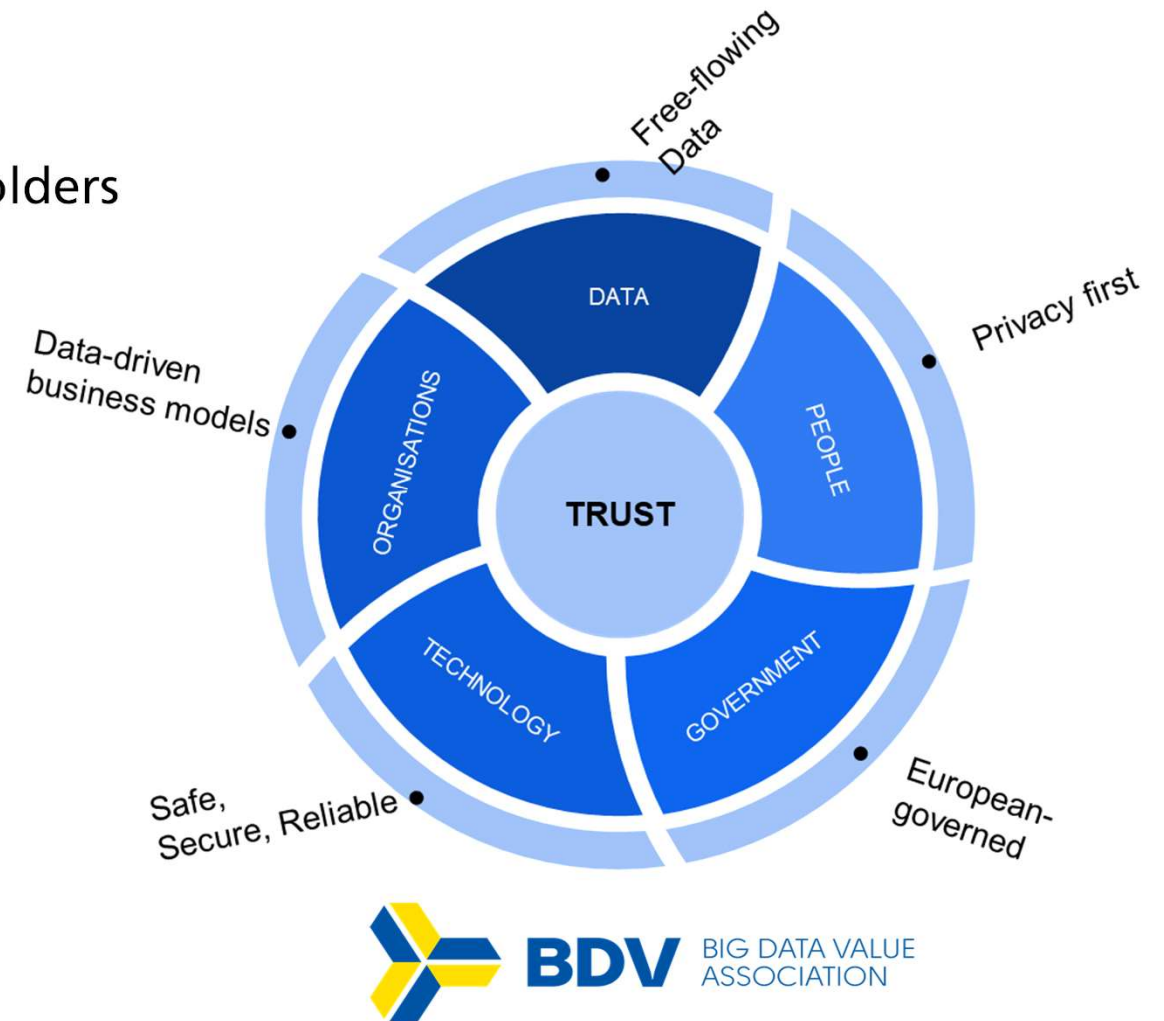
Enabling Conditions for a Euro-governed Safe & Trusted Data Sharing Space

■ The BDVA Position

- Supports the EC Data Strategy
- Need for Urgent Action by wide array of Stakeholders

■ Principles

- **Trust** is central to Vision
- Concerted effort between
 - Organisations (Business, Research & Academia)
 - Government (EC, Member States)
- Include opportunities for all!
 - Industry (B2B)
 - Science (B2S, S2B)
 - Government & Public Bodies (G2B, B2G)
 - Private citizens (C2B)



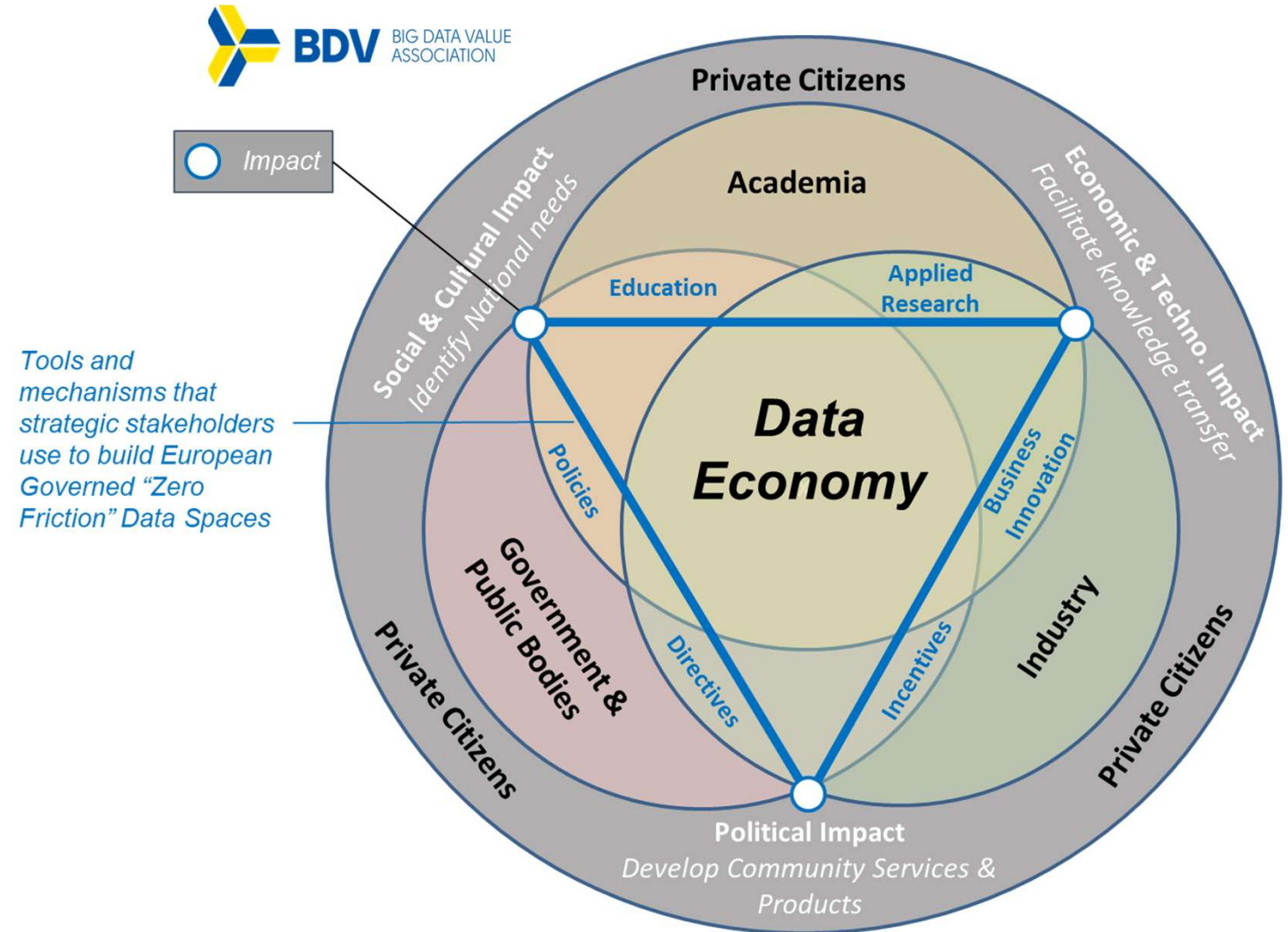
BDVA's Call for Urgent Action by Stakeholders

■ Data Economy Actors

- Government & Public Bodies
- Research & Academia
- Business/Industry
- (Private Citizens)

■ Actions Categories

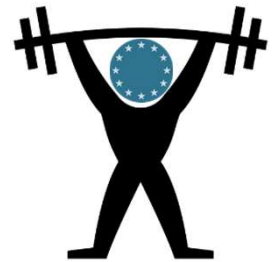
- 'Environment Enablers'
- 'Infrastructure Development'
- 'Technical Innovation'
- 'Safeguard Provision'
- 'Workforce Upskilling'



BDVA Actions: Environment Enablers & Infrastructure Development

■ EU policy makers & Member States as “Enablers”

- Setup **policy environment** for the realisation of a European-governed Data Sharing Space
- Ensure inclusive **engagement** amongst all regions, stakeholders
- **Alignment of local/national strategies** to realise framework for exploitation by industry
- **Consultation** with Standardisation bodies
- **Showcase evidence-based benefits** for diverse stakeholders
- Produce **Guidelines & Certification Mechanisms** for trustworthy participation



■ Business & Organisations as “Developers”

- **Converge existing efforts** and steer development of Trusted Data Sharing Framework
- Identify and **study use-cases** and realise **horizontal data sharing pilots**
- Joint development of **High-quality Trust Solutions** following Guidelines, Mechanisms etc.
- Explore Sustainable Business and Organizational Models for **Data Valorisation**



BDVA Actions: Innovation, Safeguards, Workforce

■ Research & Technology Organisations as “Innovation” Leaders

- Raise trust by **evolving technical solutions** for critical matters
 - Data Governance,
 - Sovereignty
 - Data rights Management
 - Veracity, Accuracy, Quality
 - Usability
- Data life-cycle management strategies that natively incorporate data sharing events and transactions
- Promote interoperability, security, quality, privacy and ethics by-design
- Cover both ‘Data’ and ‘Algorithms’



■ EU Policy Makers, Member States and RTOs to “Safeguard” Innovation

- Support EU businesses to **safely embrace** new technologies, practices and policies
- Provide facilities offering **safe testing and de-risking** of new business and innovation models
- Promote existing innovation mechanisms (Digital Innovation Hubs, BDVA i-Spaces)
- Establish **Euro-wide Regulatory Sandboxes**



■ EC & Member States to re-equip Workforce for new data economy

- Assemble EU-wide digital skills strategy
- Revisit educational policies holistically
- Reskill workforce and prepare professionals for dynamic careers, new jobs and abilities



BDVA Data Sharing Consultation – Have your say!

■ **Data Sharing Practices Survey:** Share your experiences!



“Has your organisation participated in any kind of data exchange or sharing activities with strategic partners? Regardless of whether you represent a business, research institution, governmental or non-governmental organisation your input is very valuable!”

Help shape the future by answering the 12-minute Survey!



<http://www.bdva.eu/DataSharingPracticesSurvey>

- Results will help BDVA support the convergence of efforts into realising ***a safe and cross-sectoral 'made in Europe' international data sharing space***
- Insights to be included in 2nd position paper: “Towards a European-governed Data Sharing Space”

HOW COULD IT BE TECHNICALLY ACHIEVED?

HOW?

- A Secure and Trusted Data Sharing Space
- Concerns that need a technical solution
- The IDSA Reference Architecture



Solutions for Safe and Trusted Data Sharing Spaces

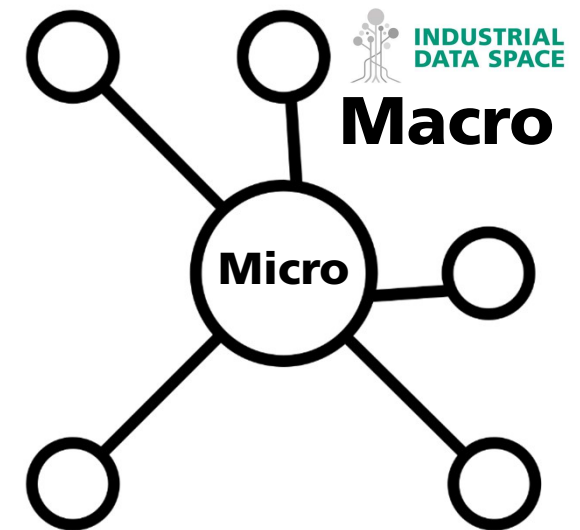
■ **'Macro' concerns – Data Sharing Ecosystem issues**

- Sovereignty concerns – managing the realities of 'transferable' data ownership
- Trust – data governance for clear cut responsibilities and obligations
- Interoperability – minimizing need for human intervention

■ **'Micro' concerns – Internal Data sharing node issues**

- Quality & Veracity – ensuring high-quality, consistency
- Secure Access – approaches to minimize risks
- Legal concerns – privacy protection and legal compliance

■ **IDSA Reference architecture addresses 'Macro' concerns**



The IDS “Network of Trusted Data”

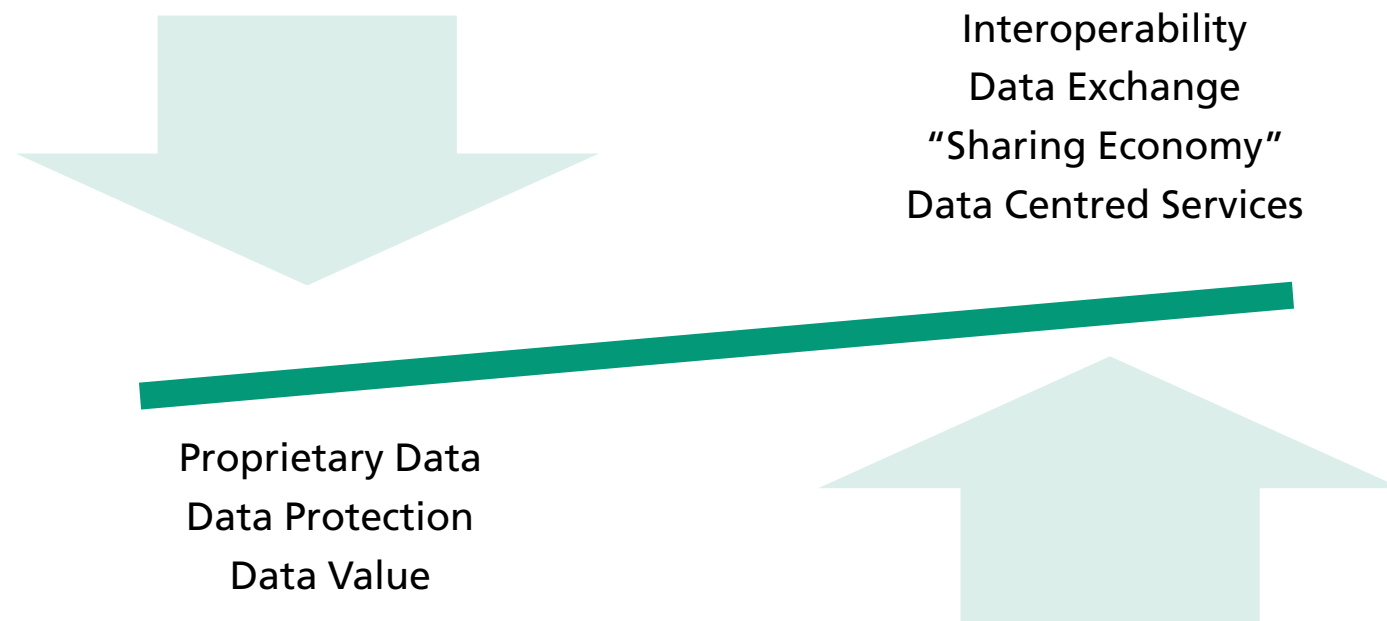
“Secure data exchange between entities where the producer(s) of data maintains sovereignty over the use of that data by authorised consumers”

- Defines conditions and governance for a reference architecture and interfaces
- Addresses **Sovereignty, Trust & Interoperability** concerns



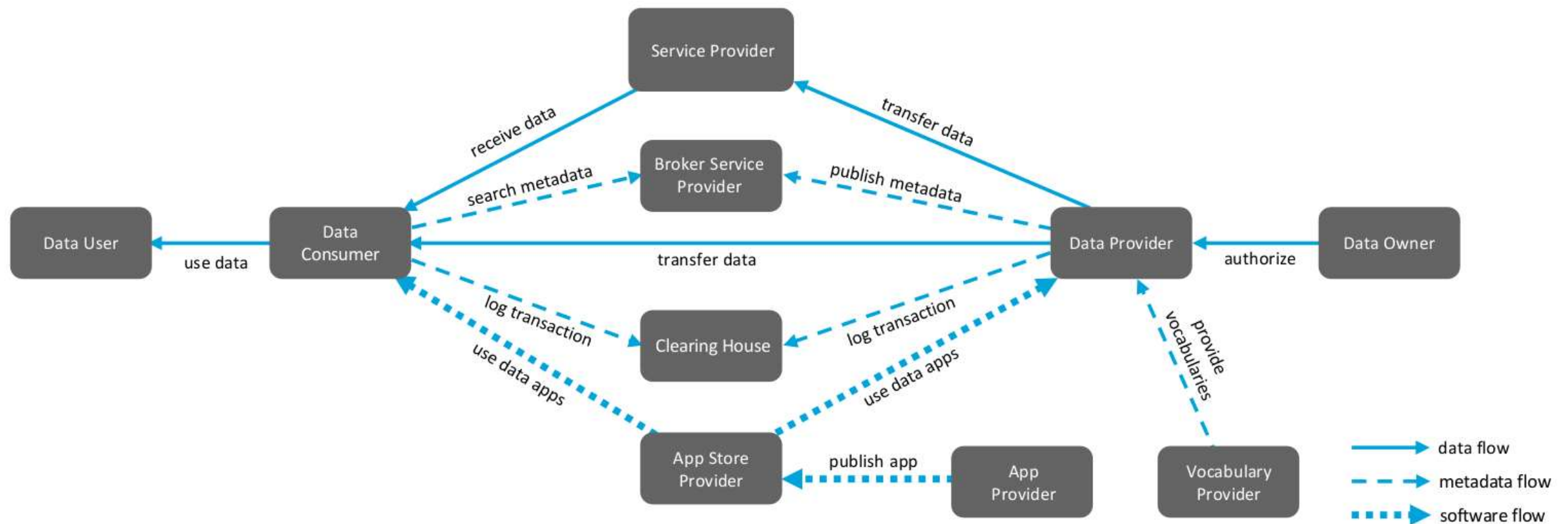
Balancing Data Management Rights & Value

***“Data Sovereignty** is the ability of a natural or legal person to exclusively self-determine their use of data assets”*



Roles and Interactions in an IDS Ecosystem

- An IDS **Connector** enables any organization to be a compliant **Producer** or **Consumer**
- Simplest **Broker** form offers a registry of available assets
- Simplest **Apps** offer means to extract the agreed upon data subject to contracts



IDSA Reference Architecture Model

Version 3, 2019



APPLYING FOR IDS_READY

Compliance to IDS reference architecture for components & organizations



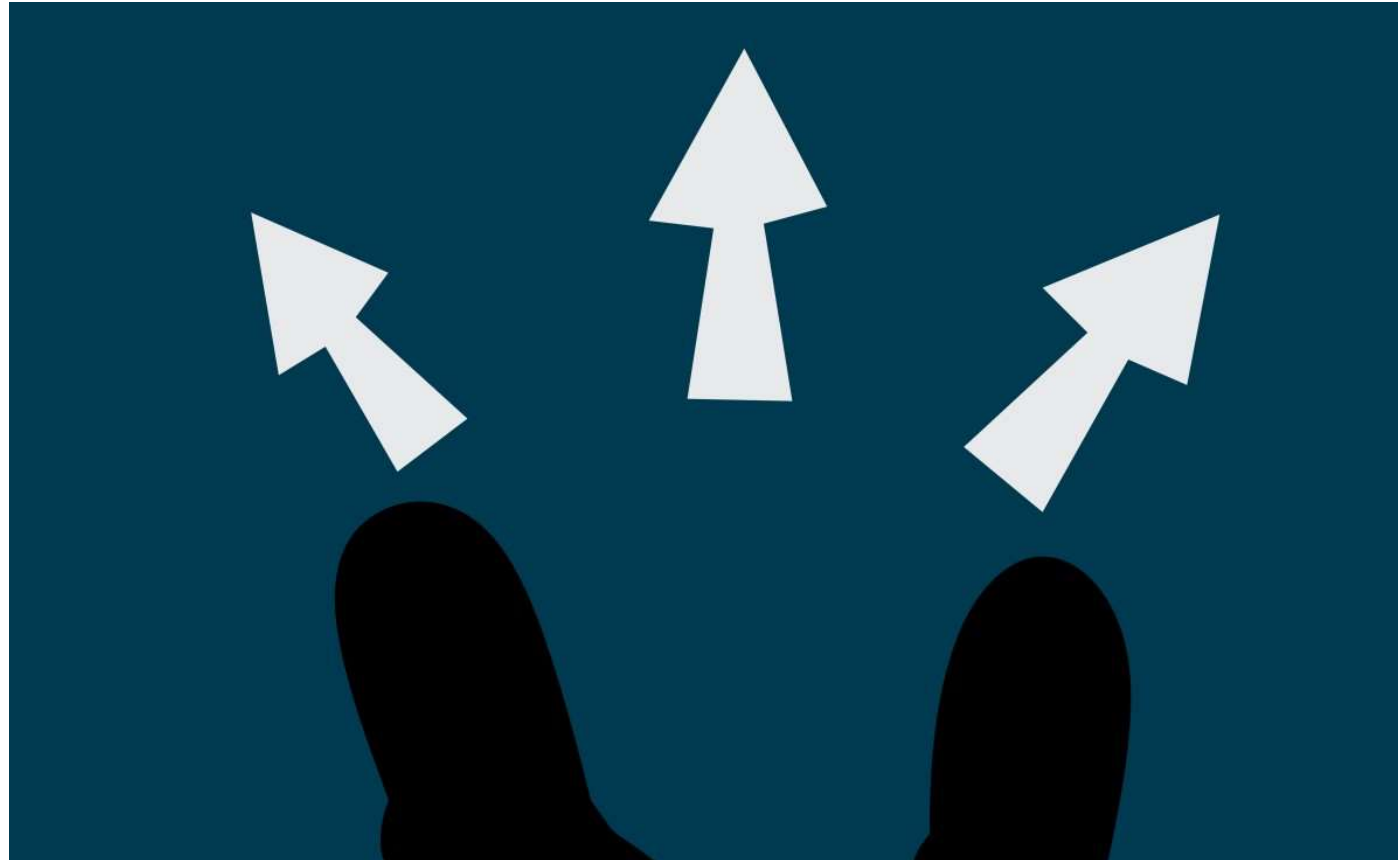
**INDUSTRIAL
DATA SPACE**

One of the core concepts/models behind the growing European **GAIAx** initiative

<https://www.internationaldataspaces.org/the-principles/#architekturmodell>

REALISING THE VISION

Research Focus



Foresight: Requirements for A functional Data Sharing Space

■ **Technical Issues that need further investment**

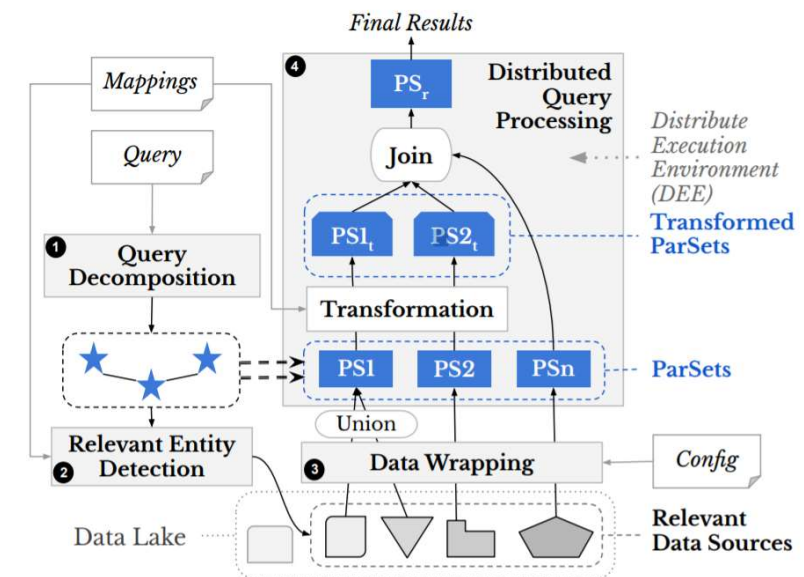
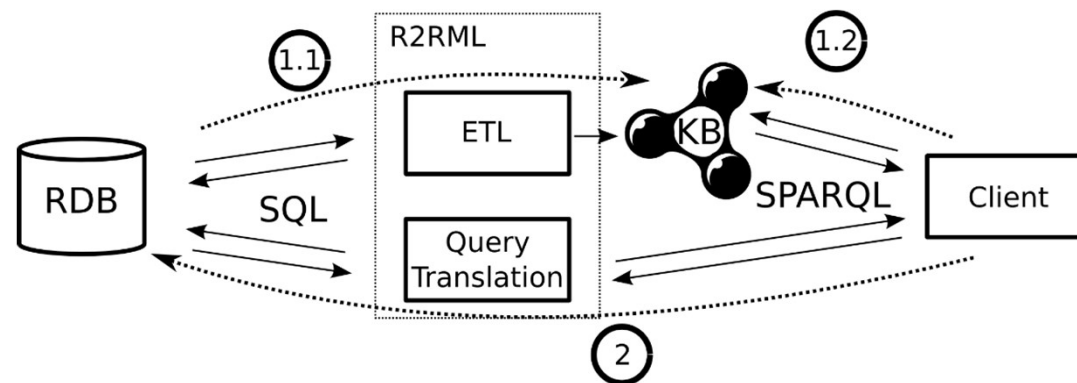
- Interoperability
- Quality, Veracity, Provenance
- Privacy-preservation
- Data Rights Management (rather than 'Ownership' rights)
- **Decentralised Data Processing Architectures** ←



Foresight: Top Research Question

■ How to realise and operate a “Semantic Data Lake”?

- *Decentralised* data: retaining data at source, in original format (optional transformation)
- *Centralised* data mappings: vocabulary-based (e.g., Broker in Marketplace setting)
- *Virtual Semantic Data*: on-demand query translation (no transformation, only mapping)
- Federated Querying over Heterogenous data in a Decentralised Architecture → ***Scalability is key!***



Source: Linking Data and Knowledge in Enterprises, Research and Society. Lange et al., 2014. Source: Uniform Access to Multiform Data Lakes using Semantic Technologies. Mami et al., 2019.

THANK YOU!



Dr. Simon Scerri simon.scerri@iais.fraunhofer.de

Senior Researcher at Fraunhofer IAIS

Member of the Board of Directors & Data Sharing Spaces Group Lead in BDVA

