Trustworthy data infrastructures

Gabriella Laatikainen, Ph.D., docent @JYU IT

Gabriella Laatikainen



- Senior Scientist at VTT Technical Research Centre of Finland, and docent at the University of Jyväskylä, Finland
- Current research focus: data-based innovations, including data spaces, data governance and AI-ready data products and ecosystems.
- Special focus on industry-research collaboration, how to apply innovative research ideas in an industrial context
- Publications in several journals, such as Decision Support Systems, Journal of Systems and Software, Information and Organization, Information and Software Technology, and Computers and Security.

About the VTT Data Innovation Lab

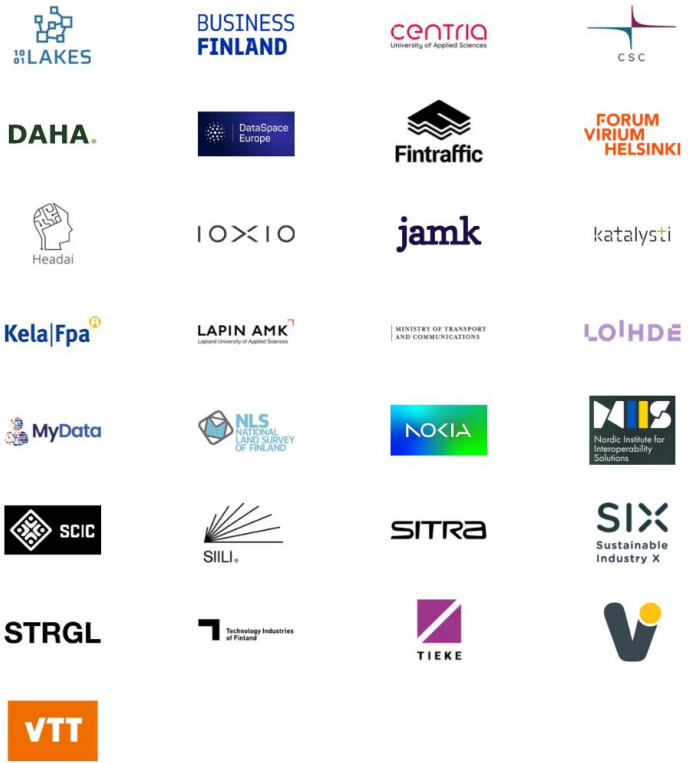


- Technical, governance, business and legal capabilities
  - Sandbox environment for data spaces as a co-creation with partners
  - Use case clinic, ecosystem orchestration, data productisation and governance
- Reference work from more than 30 pilots
- Data-space specific knowledge complemented with domain-specific capabilities
  - A large dedicated team focusing on data spaces
  - VTT has domain-specific teams focusing on most industries
  - Expertise in digital identity-related topics, incl. DPPs
- VTT is IDSA Hub Finland, a founding member of BDVA, a Gaia-X member, and part of various data spaces-related projects, such as the Data Spaces Support Centre

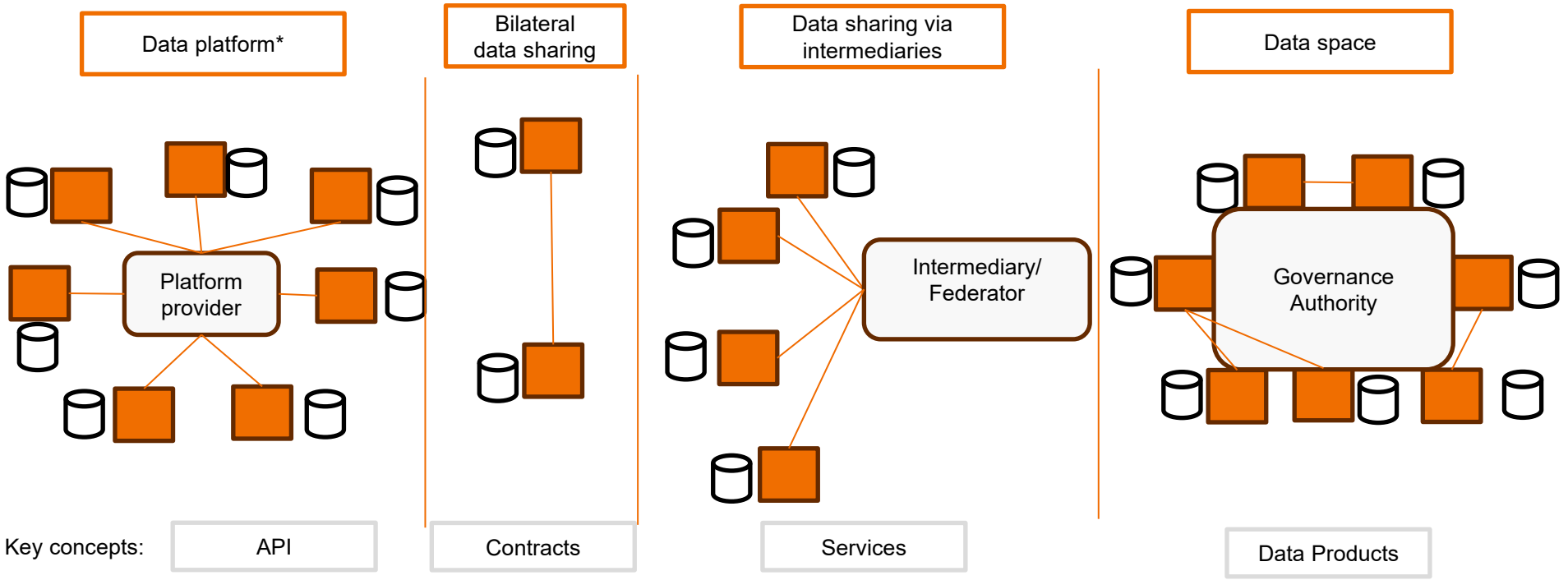
www.vttresearch.com/en/ourservices/data-spaces

About the Data Spaces Alliance Finland

- A community of organisations interested in building, utilising and benefiting from data spaces in Finland
- It strives to accelerate the growth and maturity of the Finnish data space initiatives to boost the Finnish data economy



Data sharing approaches



*Data lakes, warehouses and hubs share similar characteristics as centralized platforms in this comparison

Instead of talking of these approaches separately, the world is heading toward **data spaces-enabled data ecosystems** composed of elements of each approach

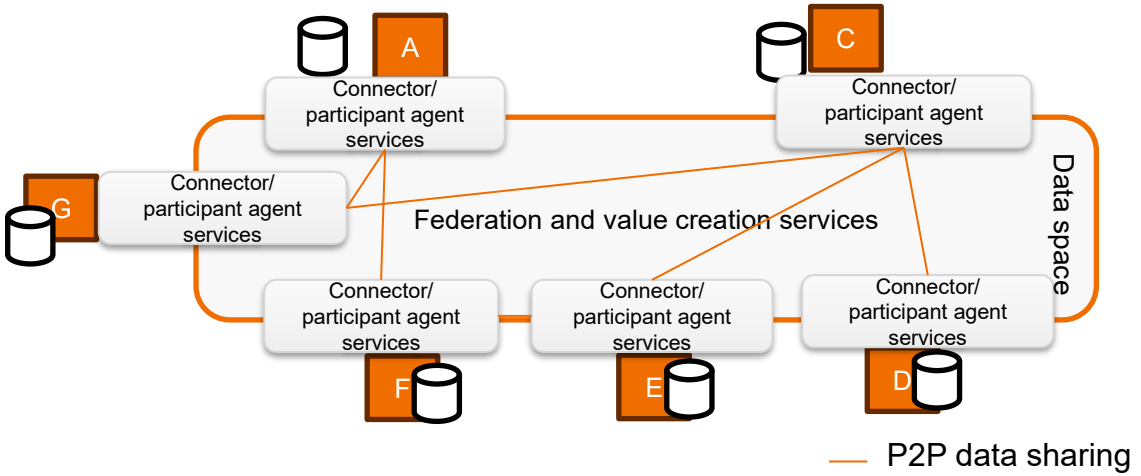
What is a data space?

governance framework and supporting **services**

to build **trustworthiness** and enable **data sharing** through an agreed set of

- policies,
- semantic models
- protocols and
- processes

[Definition from ISO/IEC DIS 20151]



Key characteristics of a data space

- ❑ distributed system
- ❑ ruled by a governance framework/rulebook
- ❑ secure and trustworthy data transactions
- ❑ data sovereignty
- ❑ enables value creation through use cases

Data space: a trusted environment for organisations to share data according to agreed rules

Data spaces enable trusted data transactions between data providers and data consumers.

DATA DISCOVERABILITY AND PUBLICATION

- Discover data
- Share the description of the data, and their access and usage terms
- Contract negotiation

DATA INTEROPERABILITY

- Semantic interoperability through common data models

DATA SOVEREIGNTY

- Participant-defined access and usage policies
- Data sharing contracts

REGULATORY COMPLIANCE

- Contractual framework
- Machine-readable policies, rules and agreements

TRANSPARENCY AND TRUSTWORTHINESS

- Transparent, clearly defined rules
- Traceability
- A set of verifiable evidence forming trust

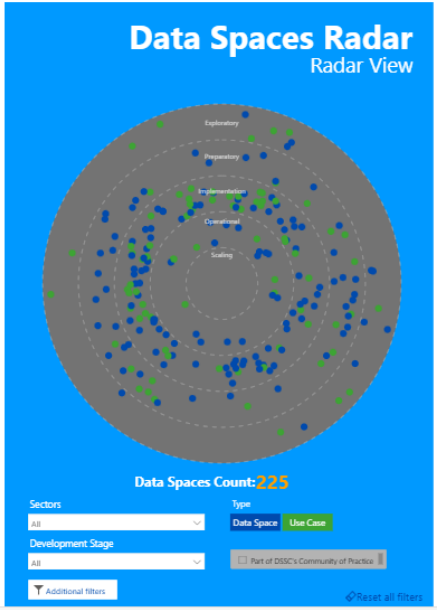
Data spaces are of different kind …

- Publicly oriented data spaces vs. commercial data spaces
- Data spaces may be built on different technology stacks
- Different sectors have different requirements and priorities
- Data spaces can be built by one organisation or a set of organisation

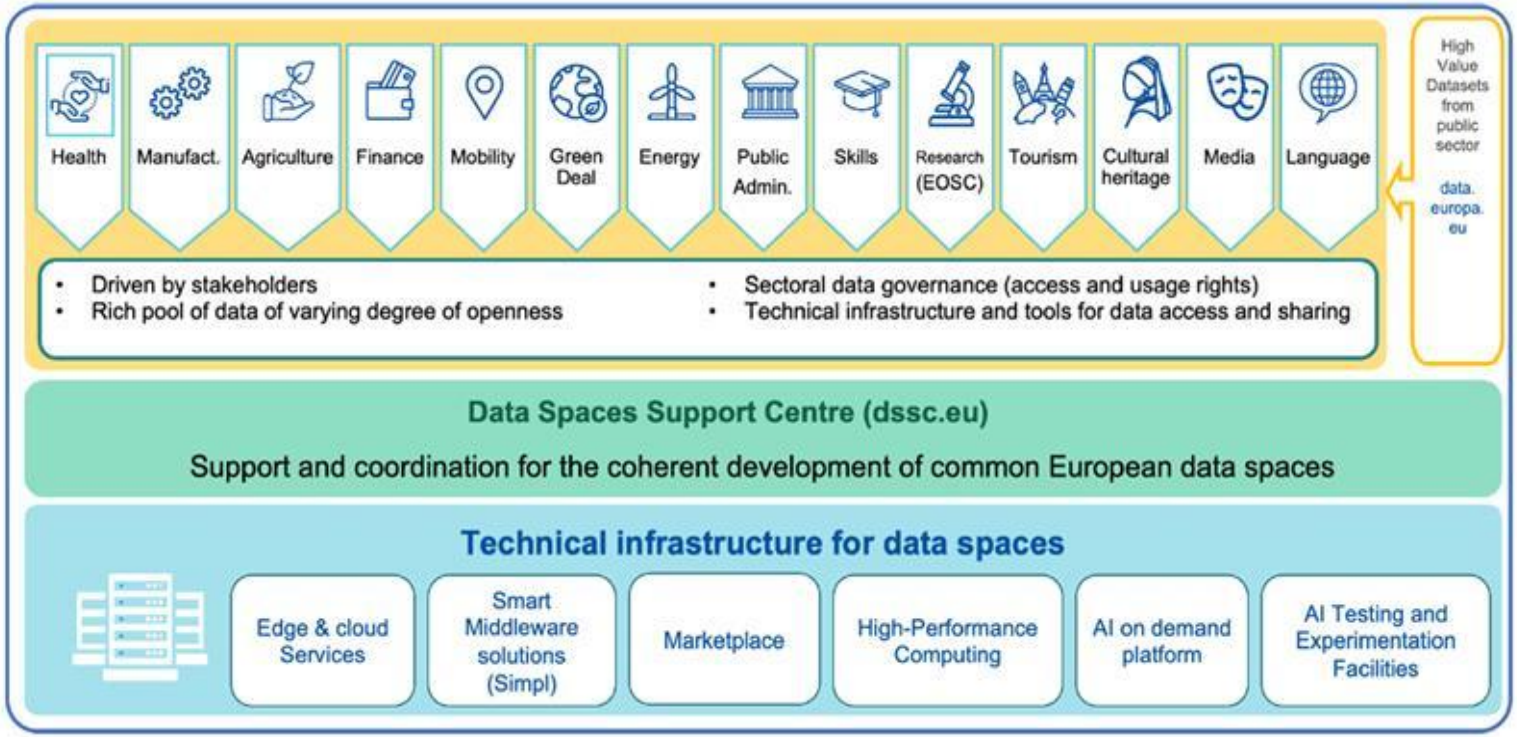
Data Spaces Sustainability and Business Models - Data Spaces Support Centre

Currently, the Data Spaces Radar lists 245 data spaces worldwide

Data Spaces Radar



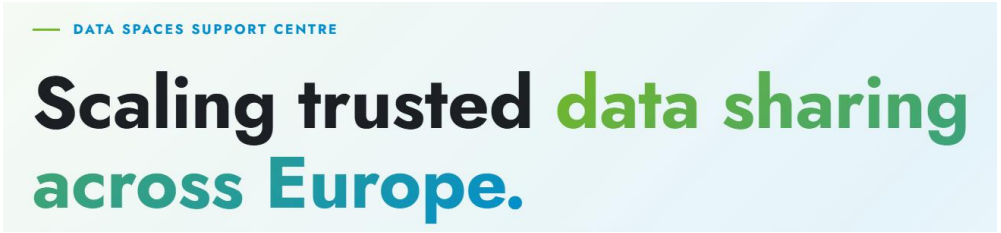
European Data Spaces Are Evolving in Multiple Strategic Fields



Common European Data Spaces

Data Spaces Support Centre (DSSC):
Designing and delivering the European Single Market for Data

The Data Spaces Support Centre supports the deployment of the common European data spaces, that collectively create a **data sovereign**, **interoperable** and **trustworthy data sharing environment**, to enable data reuse within and across sectors, fully respecting EU values, and supporting the European economy and society.



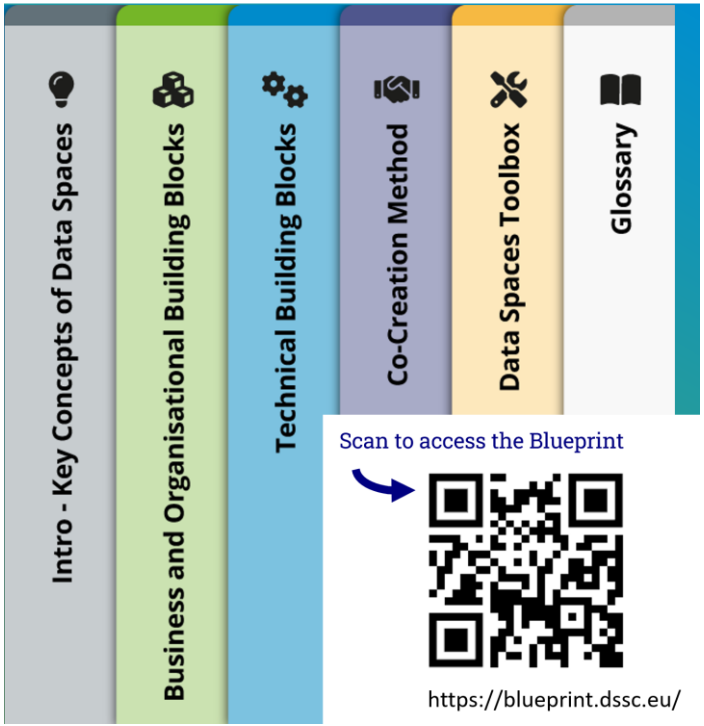
https://dssc.eu



DATA SPACES SUPPORT CENTRE

Data Spaces Support Centre Assets: a European knowledge base for data spaces

- Glossary
- Building blocks
- Co-creation model
- Toolbox
- Maturity model



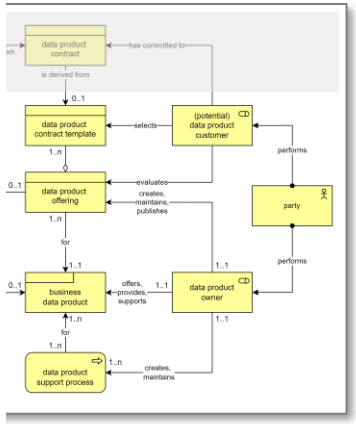
DSSC Blueprint

The Glossary

- Data spaces is a novel and emerging discipline
- Among practitioners, concepts and terminology can be ambiguous or subject to different interpretation
- Converging on language is necessary for clarity and collaboration
- The Glossary provides a consistent and coherent terminology for the data space community

Introduction - DSSC Glossary

- 1 Key Concept Definitions
- 2 Data Space Use Cases and Business Model
- 3 Evolution of Data Space Initiatives
- 4 Data Space Services
- 5 Data Products and Transactions
- 6 Data Policies and Contracts
- 7 Interoperability
- 8 Identity and Trust
- 9 Building Blocks and Implementations
- 10 DSSC Specific Terms
- 11 Foundation of the European Data Economy Concepts
- Alphabetical List of All Defined Terms in Blueprint v1.5



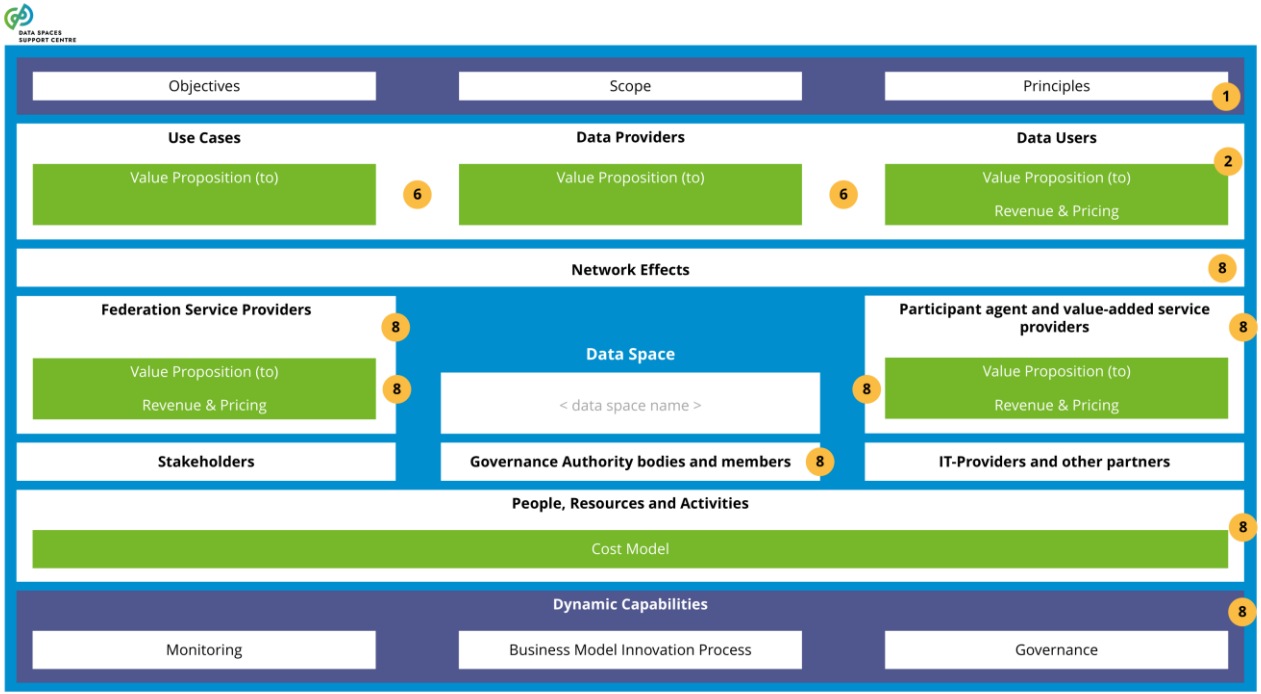
Learn more

Business and organisational building blocks

Business	Governance	Legal
Business model	Organisational form and governance authority	Regulatory compliance
Use case development	Participation management	Contractual framework
Data Space Offering		
Intermediaries & Operators		

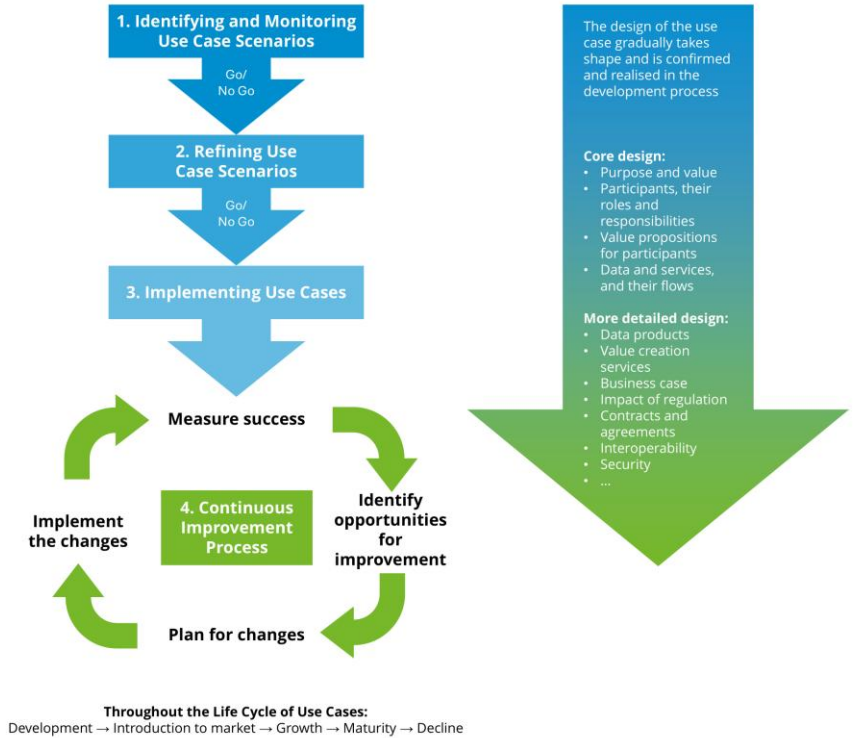
Business Model (Business)

- Multisided
- Collaborative
- Governed



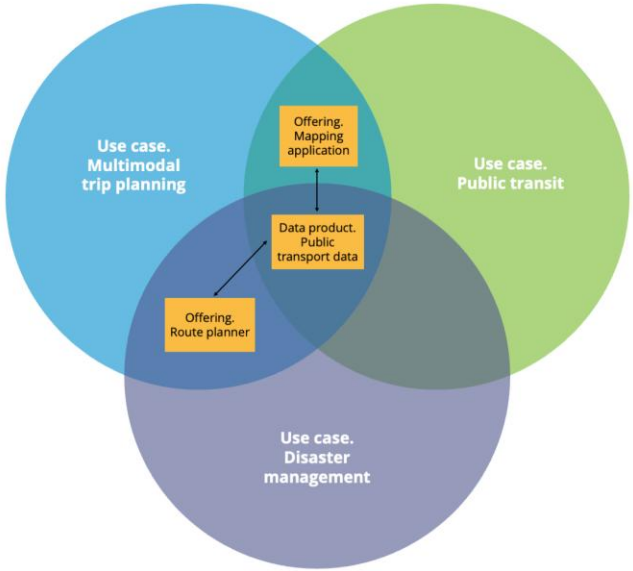
Use Case Development (Business)

- A data space use case is a specific setting in which two or more participants use a data space to create value (business, societal or environmental) from data sharing



Data Space Offering (Business)

- Data product: Data sharing units, packaging data and metadata, and any associated license terms.
  - The data product may include, for example, the data products' allowed purposes of use, quality and other requirements the data product fulfils, access and control rights, pricing and billing information, etc.
- Bundling data and service components into offerings allows for more flexibility of value creation in data spaces, such as e.g., creating value through AI models and data streaming services.

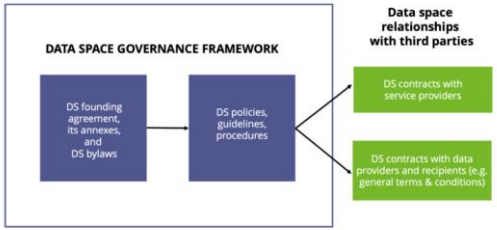


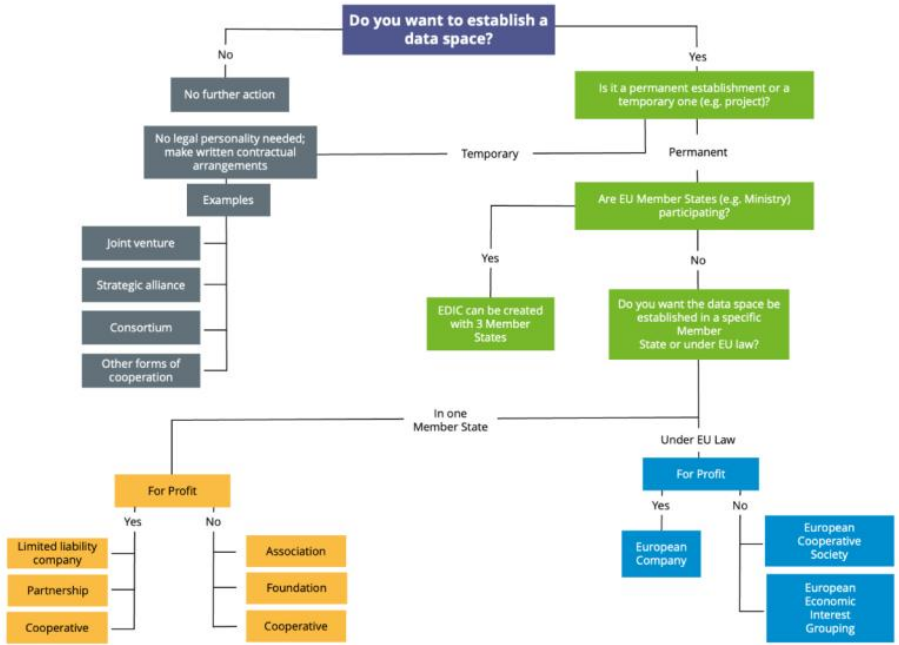
Intermediaries and Operators (Business)

- Service providers provide enabling services that are common to all participants of the data space. In common usage interchangeable with ‘intermediary’.
  - *We use typically the term ‘operator’ when a single actor is assigned by the governance authority to manage the enabling services and when it is responsible for the operation of the data space, ensuring functionality and reliability as specified in the governance framework.*
- Enabling services: federated services and participant agent services, hence services that are needed to enable trusted data transaction in data spaces.

Organisational Form and Governance Authority (Governance)

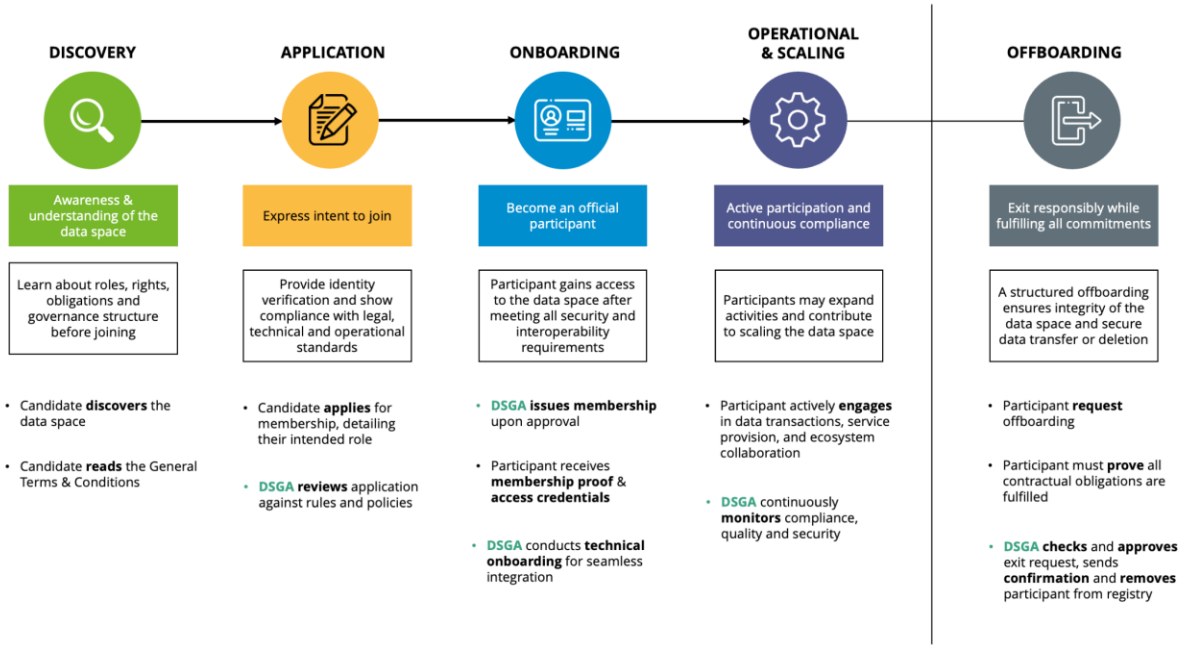
- Determining an organisational form (legal entity) and establishing a Data Space
- Creation of Governance Authority – body of the data space, consisting of participants that is committed to the governance framework for the data space, and is responsible for developing, maintaining, operating and enforcing the governance framework
- Creation of the Data Space Governance Framework – a set of internal rules and policies applicable to all data space participants





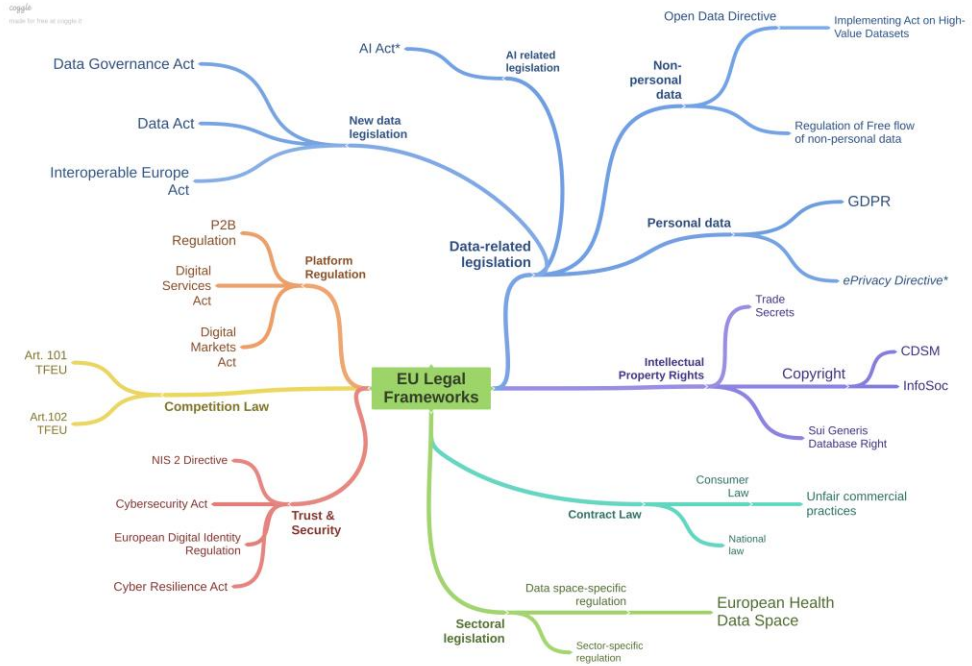
Participation Management (Governance)

- Governance processes to manage how participants engage in data spaces:
  - Specification of roles
  - Onboarding
  - Offboarding

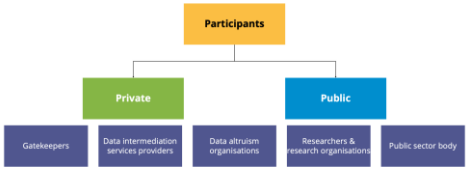


Regulatory Compliance (Legal)

- Guiding the data space governance authority in applying legal rules to a data space's design and operation
  - **Triggers** - legislation ”triggered” depending on the type of data, the type of participants and the type of domain or use case
  - **Data space requirements** –legislation that directly regulates data spaces (e.g., Data Act and European Health Data Space Regulation)
  - **Additional legal considerations** – e.g., cybersecurity law (NIS-2 Directive) and competition law

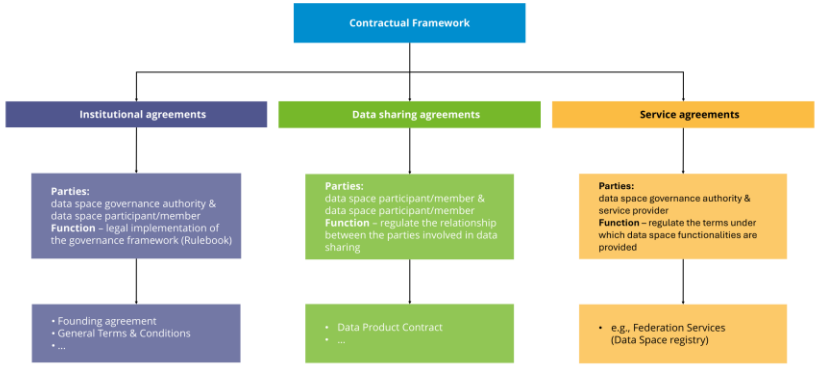






Contractual Framework (Legal)

- Legally enforceable agreements binding data space participants and transaction participants
  - **Institutional agreements** – Define the foundational governance and common rules binding all data space participants.
  - **Data-Sharing Agreements –** Regulate rights and obligations between parties exchanging or accessing data.
  - **Services Agreements -** Set terms for providing and using data-related or enabling services within the data space.



Standardisation efforts are ongoing

More information about the key data space standards: DSSC Blueprint

The latest report on standardisation landscape around data spaces: IDSA-Position-Paper-Standardisation-Landscape-Europe-and-international-version-2.0.pdf

Table 1: List of new European standards and European standardisation deliverables to be drafted and deadlines for their adoption

	Reference information	Deadline for the adoption by the ESOs
1.	Harmonised standards on Trusted Data Transactions Part 1: Terminology, concepts and mechanisms	1 June 2026
2.	Harmonised standards on Trusted Data Transactions Part 2: Trustworthiness requirements	1 November 2026
3.	Harmonised standards on Trusted Data Transactions Part 3: Interoperability requirements	1 May 2027
4.	Technical specification(s) on a data catalogue implementation framework	1 March 2026
5.	Technical specification(s) on an implementation framework for semantic assets	1 September 2026
6.	European standard on a quality framework for internal data governance	1 March 2027
7.	Technical specification(s) on a maturity model for Common European Data Spaces	1 September 2026

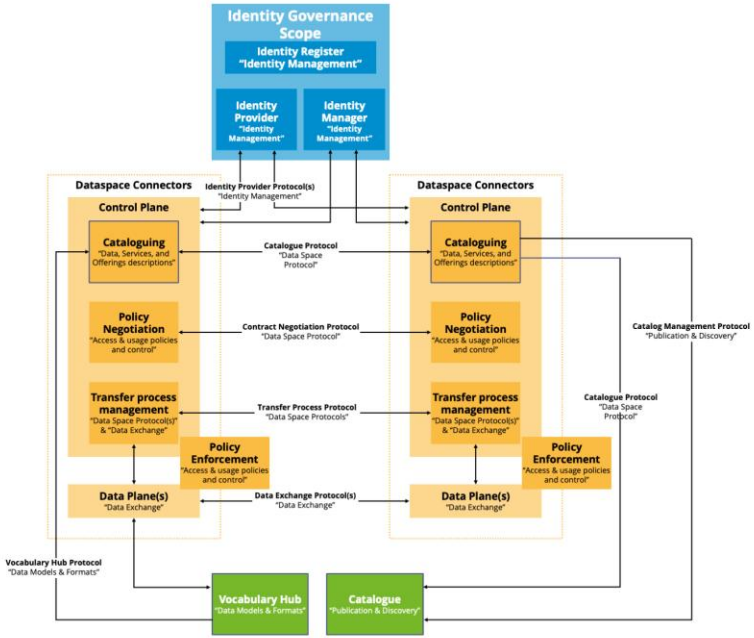
Standardisation request regarding the European Trusted Data Framework

Building on Top of Foundational Technical Standards

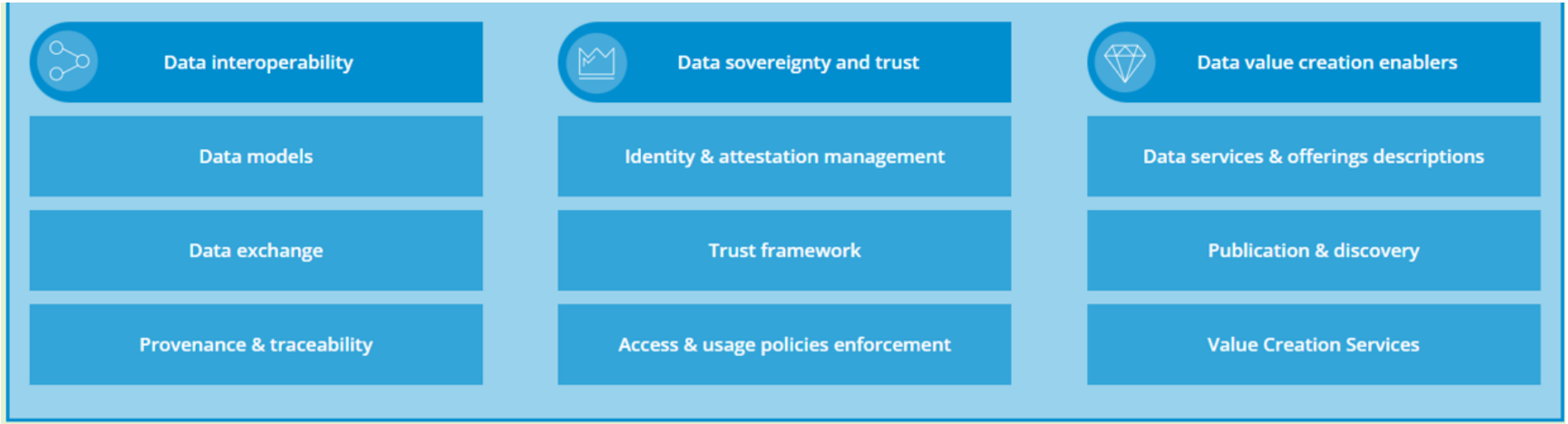
- **W3C Verifiable Credentials**: enable the creation, sharing, and verification of digital credentials in a secure and privacy-respecting manner. Tamper-evident and cryptographically verifiable. Issued for identification and to indicate that a participant is compliant with something
- **W3C DCAT (Data Catalogue) v3**: facilitate the interoperability of data catalogues published on the web. Framework for describing datasets in a way that makes them easily discoverable and accessible. Facilitate data integration and reuse across different platforms and domains
- **W3C ODRL** (Open Digital Rights Language): policy expression language o provide a flexible and interoperable framework for representing statements about the usage of content and services. Express access and usage policies to a dataset.
- **Data Space Protocol (DSP):** explains (i) how datasets are deployed as DCAT Catalogs and usage control is expressed as ODRL Policies, (ii) how agreements that govern data usage are syntactically expressed and electronically negotiated, (iii) how datasets are accessed using Transfer Process Protocols.

Control plane vs. Data plane (Technical)

- The control plane is responsible for deciding how data is managed, routed and processed.
- The data plane is responsible for the actual sharing of data.



Technical Building Blocks



Data interoperability

Building Block	Description	Specifications
Data Models	• Data accurately and consistently interpreted when exchanged • Data models: structured representation of data elements and relationships used to facilitate semantic interoperability	• **Vocabulary** (Simple Knowledge Organization System (SKOS), JSON, CSVW, …) • **Ontology** (OWL (Web Ontology Language), RDF (Resource Description Framework), or UML (Unified Modelling Language). • **Application profile** • **Data schema**
Data Exchange	• Actual transmission of the data between participants of a data space • APIs specific to a particular use case/data space • Linked to a specific data model	• **Data Exchange protocol** REST API • **Transfer Process Protocol** (Data Space Protocol)
Provenance & Traceability	• Auditing and compliance • Trace of transactions taking place in data space • Linked to the Data Product concept	• **Provenance ontology**: W3C PROV-O • **Provenance, Authoring and Versioning ontology:** PAV

Data sovereignty and trust

Building Block	Description	Specifications
Identity and Attestation Management	• Identify (i) data space participants, (ii) services and data products offered in the data space, (iii) represents any other information • Verifiable and decentralised digital identities • Verifiable credentials for such digital attestations • Exchange of attestations • Onboarding and offboarding of participants	• W3C Verifiable Credentials (data model) • W3C Resource Description Framework (RDF) • W3C Decentralised Identifiers DID • eIDAS and EDIW
Trust Framework	• Definition of Conformity Assessment Scheme • Establish a list of Trust Anchors / Trust Service Providers • Validation and verification of claims and attestations	• ISO/IEC 17000:2020 CASCO (Conformity Assessment Scheme) specifications • W3C Shapes Constraint Language SHACL • Gaia-X trust framework
Access & Usage Policies Enforcement	• Policy Definition and Governance • Policy Implementation and Enforcement -> Negotiation phase and enforcement phase	• Policy representation using ODRL (Open Digital Rights Language • PxP points (PEP, PIP, PAP, PDP

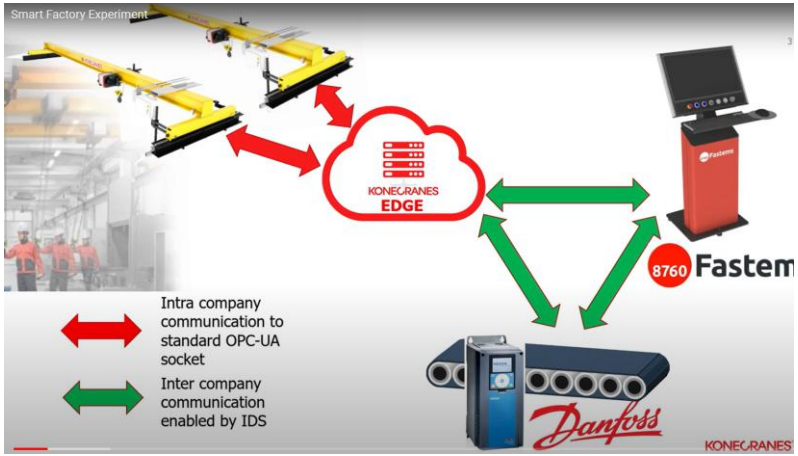
Data value creation enablers

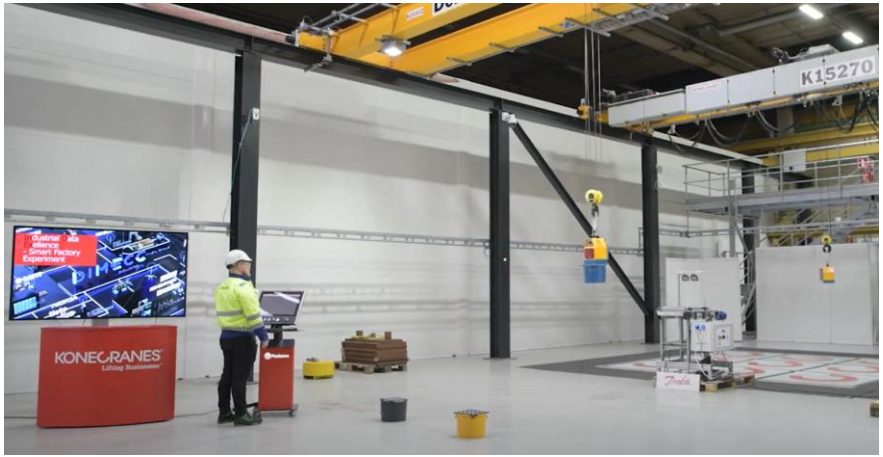
Building Block	Description	Specifications
Data, Services, and Offerings Descriptions	• Datasets, services and offerings described comprehensively and precisely • Creating, validating, updating and managing descriptions	• W3C Data Catalogue (DCAT) v3.0 • DCAT Application Profiles (AP) v3.0 • More specific application profiles • W3C ODRL, DQV (Data Quality Vocabulary)
Publication and Discovery	• Provision and discover offerings of data and services in a data space • Providers to expose descriptions, customers to search among descriptions, assess characteristics, and terms and conditions, request access. • Publication, update, removal, discovery	• Catalogue capabilities in the control plane • Catalogue protocol (Data Space Protocol) • Metadata broker
Value creation services	• Definition of services aimed to create value by different means out of the data shared in the data space • Setting up the conditions to ensure the appropriate provision, delivery and utilization of those services	• Generic technical elements of a value creation service • ISO/IEC 20000 Service management system • Service management framework specifications

Examples

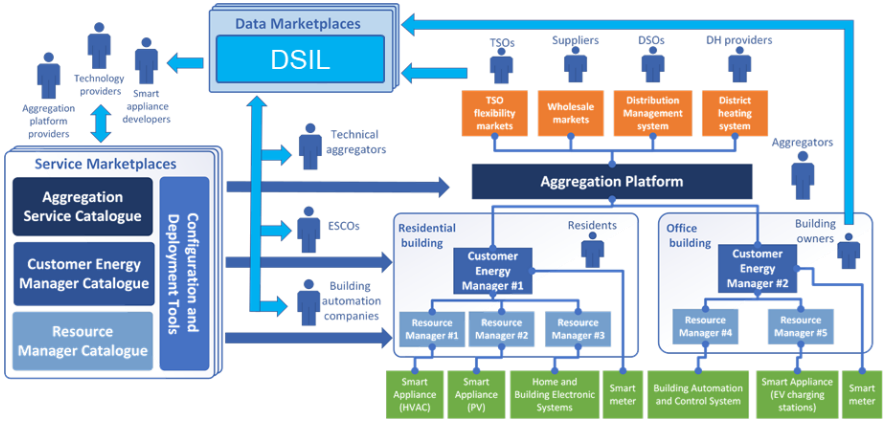
Example: A data space for factory environment

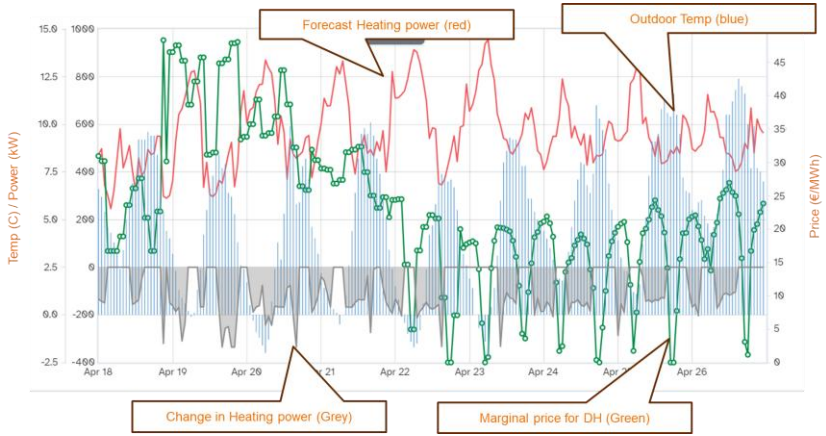
- A data space infrastructure within a factory environment, integrating machines from multiple companies using heterogeneous interfaces and protocols
- OPC UA compatible IDS connector was used
  - The connector was tested and demonstrated in the Konecranes factory premises in Hyvinkää





Example: a data space supporting peak control using price signal





Cross-sector ecosystem for demand-side flexibility management:

- Implicit demand response was implemented.
- The calculated marginal cost for district heat is shared via data space.
- 7 buildings were connected => significant energy savings both for the customer and DSO

Average flex utilized (24h) 42 kW. First estimation of value of production savings 18 k€/MW/year

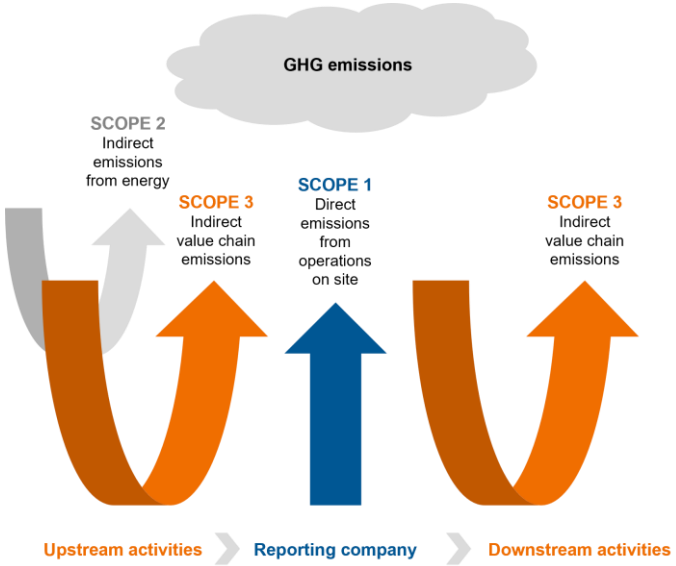
Another example: sustainability data sharing

OPTIMI research project

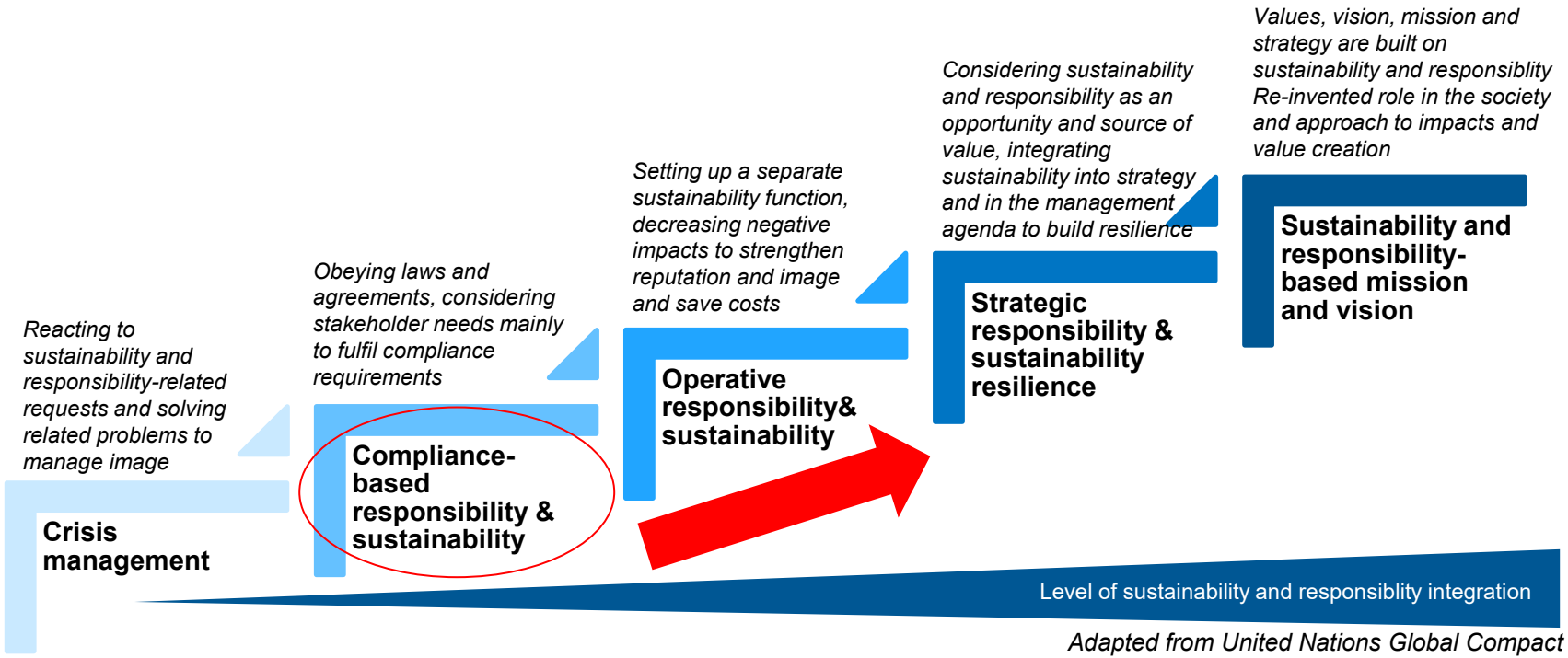
Sustainability data sharing: SCOPE3 emissions

The EU Corporate Sustainability Reporting Directive requires companies to monitor and report their Scope 3 emissions.

- In a VTT-led research project OPTIMI, we aimed to help companies turn this reporting obligation into commercial success



Steps of sustainability performance and data use



Adapted from United Nations Global Compact

Interconnected fields of actions

Own operations

Products (goods & services)

Value chains

Communities & societies

Sustainability data sharing in forestry and logistics: challenges

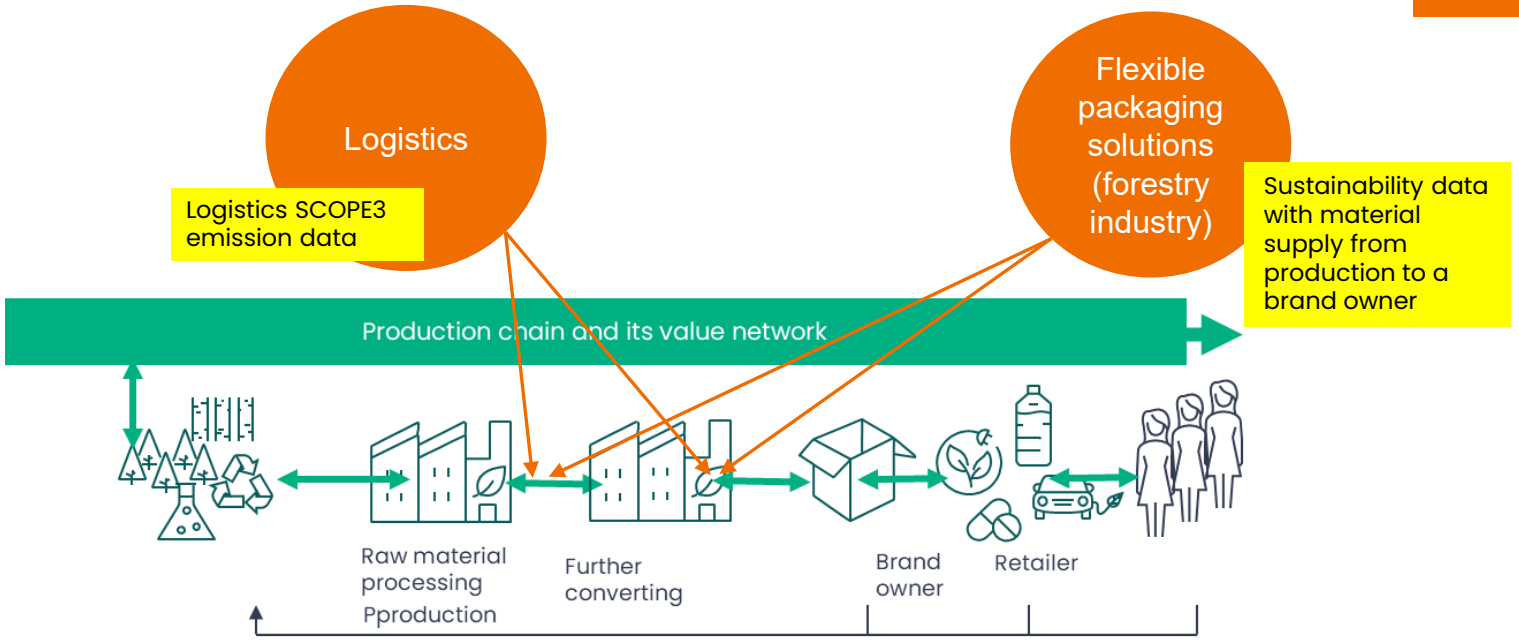
- Lack of reliable, accurate and specific data
- Limited data transparency and traceability
- Lack of standardized methodologies for calculation
- Lack of automated and scalable tools for data extraction
- Lack of resources
- Complexity of supply chains
- Limited influence over most Scope 3 categories
- Evolving and changing reporting requirements

A concept for digital SCOPE3 passport and its value creation

- **Goal**: A concept for transparent and reliable SCOPE3 emission data collection. It enables optimization of production chain upstream and downstream operations to minimize emissions, and it provides new value addition to a passport owner's customers. Calculations are based as far as possible on accurate primary data with precise methods.
- **Context framing** for the concept:
  - Value chains of flexible packaging solutions (forestry industry) and logistics
  - Sustainability parameters related to SCOPE 3: emissions due to energy and fuel consumption, fuel and energy quality – in production, use and logistics
  - Sustainability KPI frame: SCOPE 3 is analysed through environmental, social and economic aspects

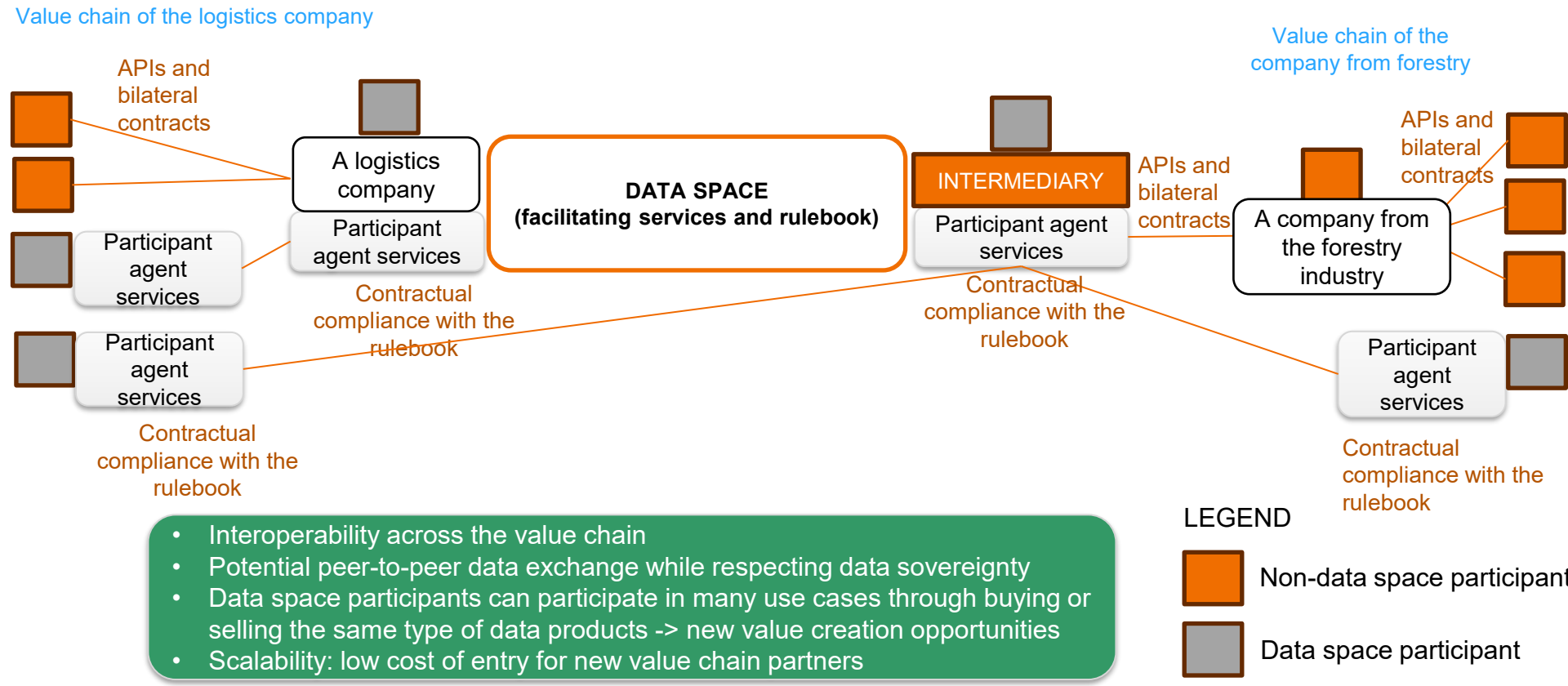


Lenses for development



SCOPE 3 sustainability data collection in a value chain

Concept for Scope 3 solution using a data space



- Interoperability across the value chain
- Potential peer-to-peer data exchange while respecting data sovereignty
- Data space participants can participate in many use cases through buying or selling the same type of data products -> new value creation opportunities
- Scalability: low cost of entry for new value chain partners

Beneficial use cases for sustainability data sharing in the logistics domain


Trip-related Scope 3 data (traceability)


Vehicle load optimization


Real-time quality improvement


Investments (stock exchange)

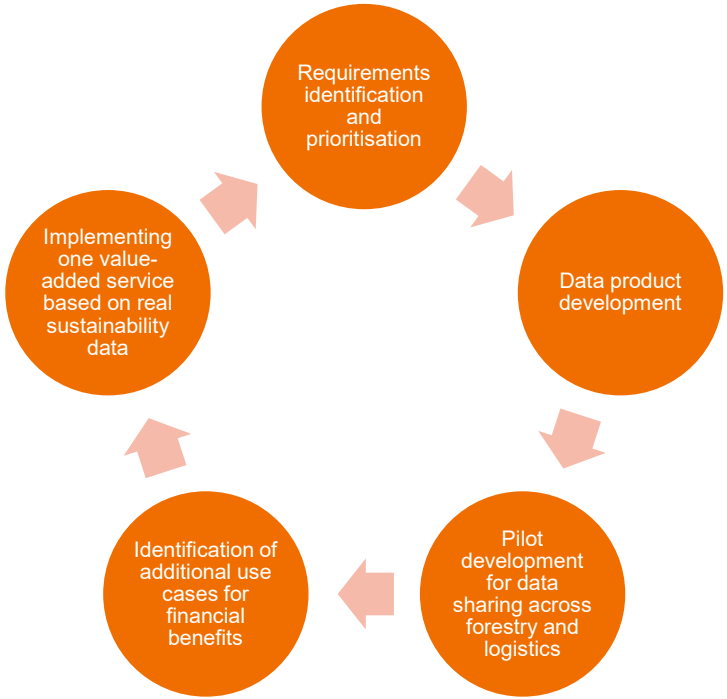

Scope 3 reporting

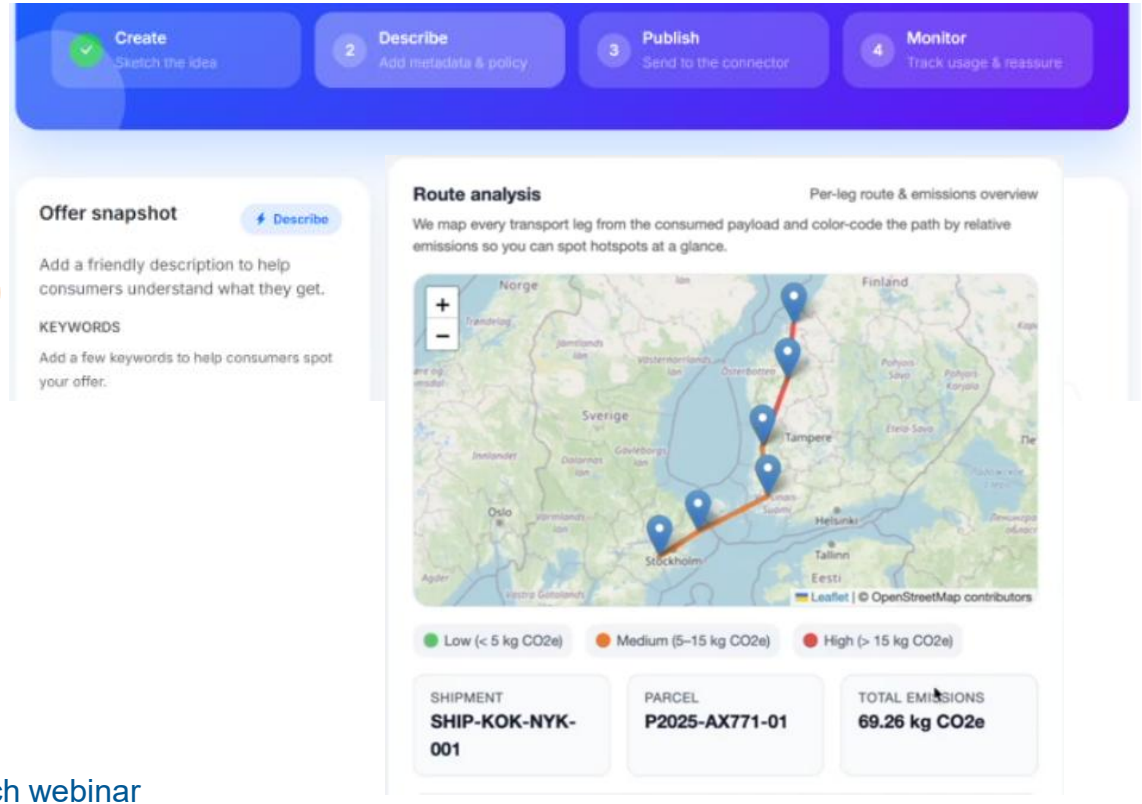

Digital Product Passport

Key benefits

- Operational cost reduction
- Internal decision making
- Brand building
- ESRD reporting
- New partners and new revenue streams

A pilot for sustainability data sharing using the companies's real data and data space architecture





[Demo visible on Youtube: VTT Data spaces research webinar]

Thank you ☺
Questions and comments?

gabriella.laatikainen@vtt.fi