

EU DIGITAL SINGLE MARKET CROSS BORDER DATA SHARING EU DATA SPACES (incl. MOBILITY DATA SPACE)

DIGITAL TRANSPORT AND LOGISTICS FORUM (DTLF)

PLUG & PLAY

FEDERATION

TECHNOLOGY INDEPENDENT SERVICES

SAFE, SECURE, TRUST

FEDeRATED CORE OPERATING FRAMEWORK

DATA QUALITY

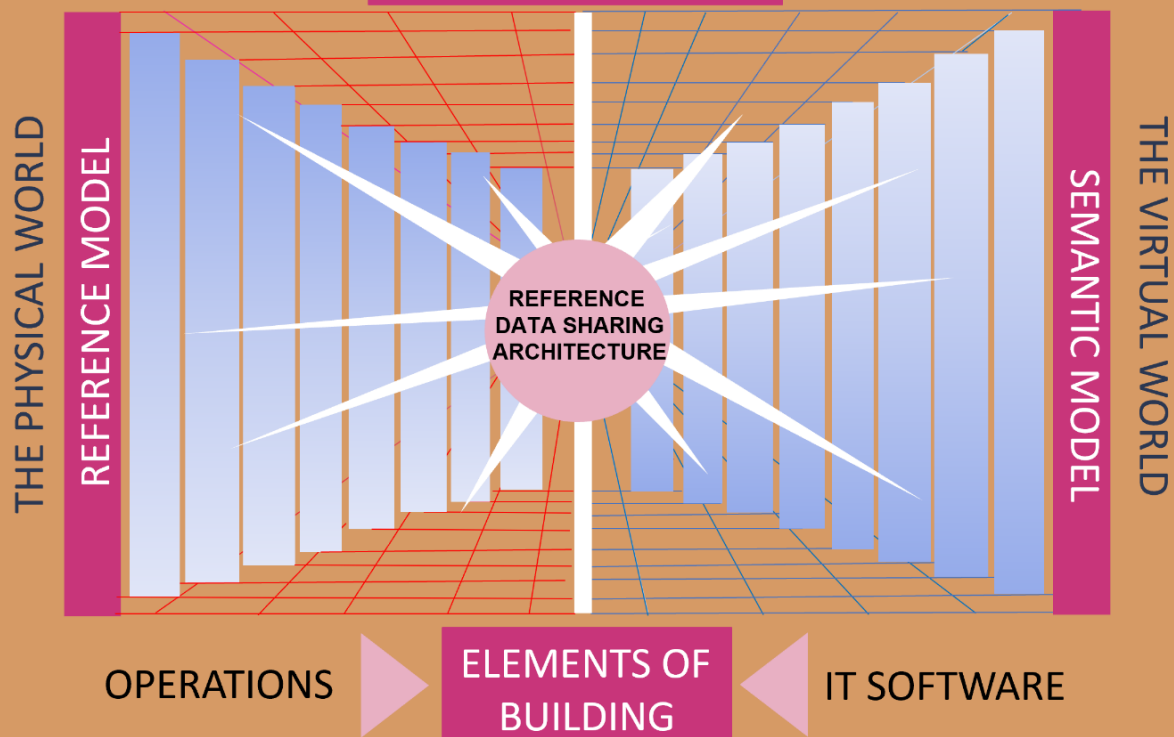
OPEN & NEUTRAL

TRUST

INTEROPERABILITY

DATA SOVEREIGNTY

LEADING PRINCIPLES



VALIDATED MASTERPLAN - NODE PROTOTYPE

This publication has been produced with the assistance of the European Union. The content of this publication is the sole responsibility of the FEDeRATED project consortium and can in no way be taken to reflect the views of the European Union.

Introduction

How to enable logistic operators and connected public authorities to structurally benefit from real-time logistics data in an agile and future-proof way? The answer appears to be easy. Enable all stakeholders to share data with another in a trusted and secure environment. Plus, apply the data at source and data sovereignty principles and establish Plug and Play (see illustration on the left page). Such a decentralized data availability would make federated data sharing in logistics become reality; - allowing data holders and data users to execute business transaction, to provide services, comply with legal obligations and to enforce legislation in the most effective and collaborative way. The anticipated outcome: seamless multimodal transport and logistics operations: smarter, greener, and more competitive.

Let's get it done. From 2019 onwards, the CEF FEDeRATED project started practical business cases in correlation with the development of an operational framework for federated – read: decentralized – data sharing. An IT Architecture Board and Semantic Modelling Group set the pace, developing a Reference Architecture, a FEDeRATED semantic model, and some tools. After 5 years, it's time for a next step. To team-up with many many stakeholders in the logistics chain to finetune the federated architecture design and make it work for us all.

Easier said than done. The logistics market is huge and complex. Many different and competing interests, business models and dependencies exist. Teaming-up for federated data sharing means getting ready for change. Thereto, it is essential to be transparent on what steps to take. To make texts aimed to connect and convince IT professionals accessible to all stakeholders. This Elements of Building booklet serves as a helping hand in just doing that.

This Elements of Building booklet challenges you to apply 21st century digital technology for enhancing your operational brainpower, incorporating the power of pull in your business processes. To assist you in developing and building a federated data sharing network for logistics with business partners, whilst enhancing EU operational brainpower applying the European Interoperability Framework.

With this booklet, we aim to provide you with accessible information trying to persuade you to get involved in federated data sharing. We have focussed to be as concise and informative as possible. The 17 information factsheets (chapters) are captured in four parts: Design, Capabilities, Migration, and Tools.

In case you feel tempted to take the first steps towards a customized federated data sharing approach, please note: the elements of building covered in this booklet are elaborated in depth in various documents and articles accessible through the FEDeRATED website:

www.federatedplatforms.eu and especially in its page on products:
<https://federatedplatforms.eu/index.php/products>

We hope you may find this booklet useful.

November 2023
www.federatedplatforms.eu



Design	Context	1
	Data Sharing Options	2
	The Federated Approach	3
	Data Sharing Grid	4
Capabilities	Operational Framework	5
	Semantics	6
	Service Registry	7
	IAA	8
	Index	9
Migration	Roadmap – Migration Strategy	10
	Stakeholder Engagement	11
	Adoption Strategy	12
	Legal Framework	13
Tools	Getting Started Implementation	14
	Semantic Adapter	15
	Node Installation	16
	Node Interfacing	17

Design - Context

Digital technology plays an important role in changing our ways of living, enabling global connectivity and collaboration. Computing power per Euro spent is increasing. We keep finding more use for it. Machine learning/AI is about to utterly transform—and enormously expand—the global economy. The World Wide Web is evolving from a 2D working space - designed to connect pages of information viewed through a browser - into a 3D or even 4D (including time) dominated Spatial Web, which enables people, places, events, and things to connect and collaborate in active ways; - innovation at the edge based on data at source.

The Spatial Web will fundamentally alter the economy as we know it; affecting every industry from healthcare to education, to manufacturing and transportation. Fundamentals must be managed well: liquidity, productivity; cost of the future. Digital literacy is a cost of the future. We are moving away from an Electronic Data Interchange (EDI) and Application Programming Interface (API) based data exchange between a limited number of partners towards semantic interoperability, empowering freedom to move and preventing vendor lock in and centralized and monopolistic platforms to dominate the market.

In the world of the Spatial Web, any logistic operation or event - business transaction, service provided and compliance event - will be executed and examined based on a digital twin of the logistic chain operator. The reactions of a digital twin will be perfectly informed, even by what cannot be seen. Digital twins can reveal the real time functioning of any logistics operation, also monitoring every environmental factor affecting safety or productivity via sensors of unparalleled accuracy and ubiquity, to seeing snags in a global supply chain.

Increasingly, supply chain management practices – the sequence of events and processes that take a product from cradle to grave¹ – dominate logistic chain operations, requiring public authorities and enterprises to set up a new framework for data sharing to:

- Interact with, authenticate, authorize, and analyse events in real time anywhere across time and space.
- Allow continuous productivity improvement urging all to collaborate and find a structure for innovation.

Europe

The EU and national policy and business practices in freight transport and logistics aim to deliver seamless multimodal freight transport operations. Digital technology can assist achieving green and more competitive transport operations.

The European Union pursues a sustainable and smart transport agenda, also in connection with developing an EU Digital Single Market. For that purpose, the European

¹ The Supply Chain Council identifies five features: Plan, source, make, deliver, and return.

Commission initiated various policy initiatives, such as the EU Digital Transport and Logistics Forum (DTLF²), an expert group, and the EU Data Strategy³.

The EU DTLF aims to create an open, neutral, and trusted data sharing infrastructure that can be applied by all logistics stakeholders with the so-called 'federated network of platforms' approach. The DTLF approach is supported in the EU Data strategy, that has proposed the concept of data spaces. Data spaces are decentralized infrastructures, where diverse actors can share and use data in a secure, reliable, and trustworthy manner, following governance, organizational, regulatory, and technical mechanisms. They will interact various data ecosystems in a demand-driven process⁴. For logistics the DTLF and Data Space approach should preferably converge in an EU Mobility Data Space, which should also cover passenger transport.

The CEF FEDeRATED Action is an EU Member States driven initiative to contribute to the establishment of a viable federated network of platforms for data sharing, as prescribed by DTLF, in the freight transport and logistics domain at EU level (and beyond).

² The DTLF is established in 2015 and acts as an expert group consisting of EU Member States and Industrial Stakeholder organisations representatives. raised and chaired by EC DG Move. The DTLF identified 4 building blocks for federated network of platforms: plug and play, independent technology services, federation and safe, secure and trust.

³ The EU Data strategy related and led to a number of EU legal initiatives, such as the General Data Protection Regulation ePrivacy Regulation, Digital Operational Resilience Act, Data Governance Act, Digital Market Act, Digital Service Act, NIS2 Directive, Data Act, and Artificial Intelligence Act. New proposals are under preparation.

⁴ Source: EC Joint Research Centre (JRC). The data spaces principles are:

- Data sovereignty: Keep ownership and autonomy over data
- Security: Prioritize data security via encryption and ensure confidentiality
- Control: Revoke access at any time and retain control
- Interoperability: Consistent formatting and nomenclature enabling seamless integration
- Adaptability: By design, versatile and accommodating to various tech, use cases and industries

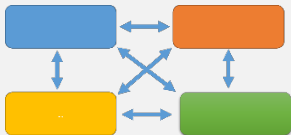

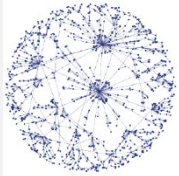
Design - Data Sharing Options

2

Three data sharing architectures can be identified, namely:

- **Messaging** – EDI data is duplicated from one organisation to another. Messages mostly reflect business documents like orders and invoices. Specification of message semantics is in a syntax, e.g. XML, JSON, or EDI
- **API/web services** – one organisation provides an IT (web) service – can be done through an API - enabling another organisation to provide or access data. specification of message semantics is in a syntax, e.g. XML, JSON, or EDI
- **Semantic Web** – all organisations have access to the same semantics (ontology) and can decide which data to share with others.

There are three data sharing Designs to choose from:

1 BILATERAL Peer2Peer	2 PLATFORM	3 FEDERATED Multiple, open and neutral
		
One organisation shares data with another organisation through a direct link	A central entity provides the platform to which individual parties connect, enabling these parties to share data with each other, greatly reducing the links for parties to share with each other	Any party (node in a grid) is capable to non-prescribed M2M querying of any other party (node) and to share readable data through an access point with any other party, while keeping the data at source and applying security mechanisms

Data sharing Designs and the European Interoperability Framework (EIF)

1 BILATERAL Peer2Peer	2 PLATFORM	3 FEDERATED Multiple, open and neutral
TECHNICAL INTEROPERABILITY		
Message architecture	Open API	Semantics
SEMANTIC INTEROPERABILITY		
Message model	Message and Data model	Semantic model - ontology
ORGANISATIONAL INTEROPERABILITY		
Individual business case	Multi stakeholder business case	Multi stakeholder and sustainability business case
LEGAL INTEROPERABILITY		
Bilateral agreement	Platform setting	Transnational agreement – possibly legal setting

The Design Characteristics (including pros and cons)

1 BILATERAL Peer2Peer	2 PLATFORM	3 FEDERATED Multiple, open and neutral
IDENTIFICATION & AUTHENTICATION		
<ul style="list-style-type: none"> • Use webtokens or Open OAUTH standard 	<ul style="list-style-type: none"> • Use webtoken or Open OAUTH standard • Platform can provide security in a data space. • Verifiable Credential (VSs) issued by Registration Authorities can be applicable 	<ul style="list-style-type: none"> • Must apply independent mechanism. • Requires application of Verifiable Credential (VCs) issued by Registration Authorities
LINKING WITH EXISTING PLATFORMS		
<ul style="list-style-type: none"> • Not easy – an agreement on what and how is needed 	<ul style="list-style-type: none"> • Linking to central platform required • Easy - Message exchange through platform between already connected parties 	<ul style="list-style-type: none"> • Quick linking possible due to M2M prepared linking to new parties
UNAMBIGUOUS CONCEPTUAL FRAMEWORK		
No: must be discussed specifically	Yes: enforced by message format standard	Yes
DATA DIRECTLY FROM SOURCE		
Yes	Yes	Yes
ADVANTAGES		
<ul style="list-style-type: none"> • Easy to implement for limited number of links • Often adopted and implemented in supply and logistics. • Trust is no issue • Liability clear 	<ul style="list-style-type: none"> • Easy to connect many parties • Large variety of interfaces (often API) between organisations • Wide range of standard services • Secure data and data communications • Liability clear 	<ul style="list-style-type: none"> • Data at source • Scalability • Open to all based on set of agreements. • Interaction patterns for data sharing of real-world objects and their status (Digital Twins, events), makes it possible to fully digitize processes. • Low risk vendor lock-in due to open standards
DISADVANTAGES		
<ul style="list-style-type: none"> • Complicated and time consuming to scale-up • Management issue (many links) 	<ul style="list-style-type: none"> • Limited space for innovation • Hard to deviated from existing services • Follow data sharing rules • Often conservative business model • API requires IT investments (SME problem) 	<ul style="list-style-type: none"> • Technology under development • Few semantic industry standards available • Liability issues need set of agreements • Generic governance model

Design - The federated approach

3

Federated data sharing is a means to overcome current bottlenecks preventing the structural realisation of digital solutions for seamless multimodal freight transport operations which can be characterised through:

- No common language.
- No level playing field.
- Insufficient interoperability.

As such, a federated network of platforms has to rely on the comprehensive consideration of certain definable design requirements as well as legal and organisational boundaries, constituting the following key principles:

1. Ensure data sovereignty and data quality.
2. Create trust among platforms and participants.
3. Provide a framework to enable interoperability.
4. Be open and neutral to any participating party.
5. Rapid deployment of new IT services to support business processes.

Federated data sharing is about data accessibility (pull data) by authorized users to:

- Make data-based logistics feasible for all stakeholders.
- Develop - just like one internet, made up of many different networks and services - one (common) data sharing grid where all data users and holders can qualify – based on a set of requirements - as a node. The market that this will unlock will be much bigger than any of them could create alone.
- Provide any stakeholder the freedom to safely browse the (data sharing) grid: to explore new business opportunities, conduct data-based business transactions and compliance procedure, under the condition of safeguarding data autonomy.

The biggest challenge is interconnectivity. Open standards need to be devised and adopted. Portability of data, or for that matter property rights, over virtual items will drive standardisation and interoperability over time.

The major reasons to choose a federated data sharing design are:

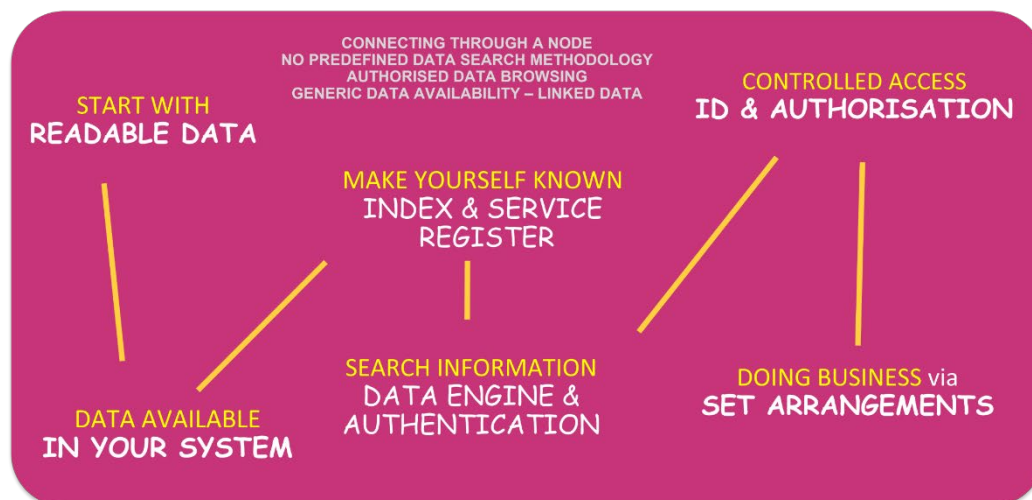
1. Focus on decentralized, open, neutral, trust and data at source, enabling innovation at the edge.
2. Querying for each organization formulated against any infrastructure that is configured with all relevant semantic models (upper – and aligned ontologies) and contains matching functionality⁵.
3. Integration of various interaction patterns and their data⁶ allowing each organisation applying different technologies (messaging, APIs, triples) to select its own way of integrating to a federated infrastructure.
4. Matching various data sharing between environments using different semantics.

⁵ From a regulatory perspective, these queries must be published allowing enterprises to provide required data sets and be compliant with applicable regulations The data set required for eFTI (electronic Freight Transport Information) Regulation can be formulated as such a query.

⁶ The upper ontology contains concepts for specifying interaction patterns for sharing data about real-world concepts. These are called the Technology Independent Services (TIS).

5. Flexible and extendable alignment of organizations, data spaces, platforms, industry associations, and regulators can specify their data requirements based on the common (Web based) concepts and make them available to all others.
6. Data quality - A node supports data quality validation - correctness and completeness of event data and query (results) and either in its internal IT systems or by its index.
7. Governance of only the common concepts⁷ and specialization, alignment, and matching procedures must be agreed upon and require governance. These can be standardized, thus requiring a light-weight governance structure⁸.

The federated data sharing design is based on the notion of interoperable Nodes, enabling platforms and organisations to enjoy full interconnectivity. The Node is based on complying with the federated capabilities. A prototype example of one approach to such a Node has been developed.⁹



Query the graph, connect and share – to boldly go where you haven't been before.

⁷ These common concepts for data sharing in supply and logistics are the so-called upper ontology. This upper ontology can be standardized.

⁸ These common concepts for data sharing in supply and logistics are the so-called upper ontology. This upper ontology can be standardized.

⁹ See chapter 11B (Installation and configuration instructions are available on the FEDeRATED Github page: <https://github.com/Federated-BDI/Docker-BDI-Node>)

Design - Data Sharing Grid

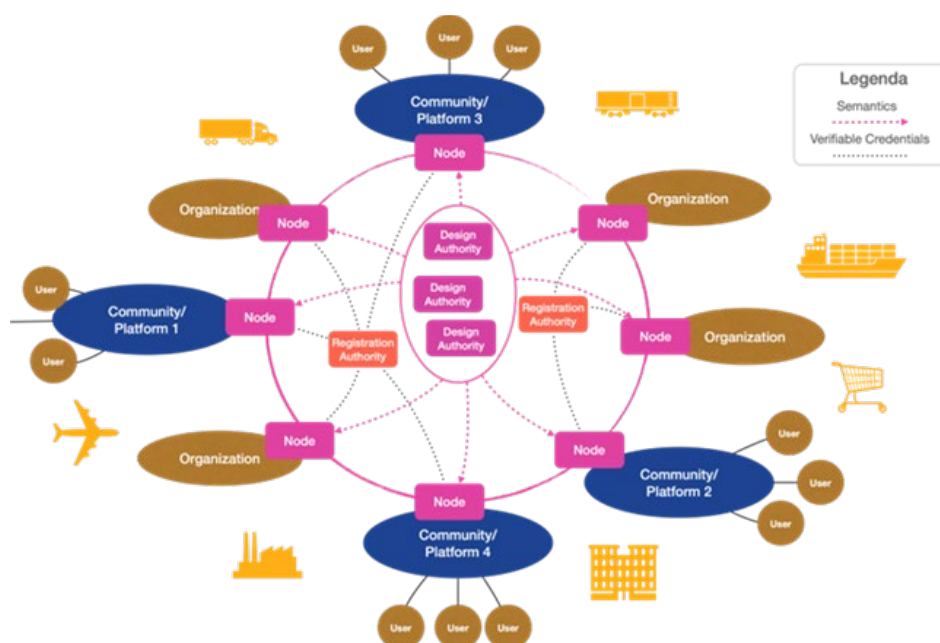
Getting operational is based on a design. The purpose of the design is to enable all operators and public authorities to understand each other, enable non-predefined data queries and to connect a huge amount of data that are based on non-compliant standards.

4

The design considers a data sharing grid¹⁰. The most important design features are:

- The grid consists of a wide variety of data generating and receiving nodes.
- The nodes generate their own data from various sources (mostly supply chain operators) and exchange the data on request through the grid.
- Within the grid the nodes are orchestrated in a decentralized manner using specific language models (harmonized data interoperability), data discovery tools, and access control and security mechanisms.
- A set of agreements enable nodes to execute transactions and legal compliance procedures, also governing the functionality of the nodes (change management procedures) and the applicable methods to calculate the ESG impact¹¹ of the transactions.
- Each organization or platform locally interface to a node tailored to its requirements and configures its node for the required business services or -goals and their interaction patterns.

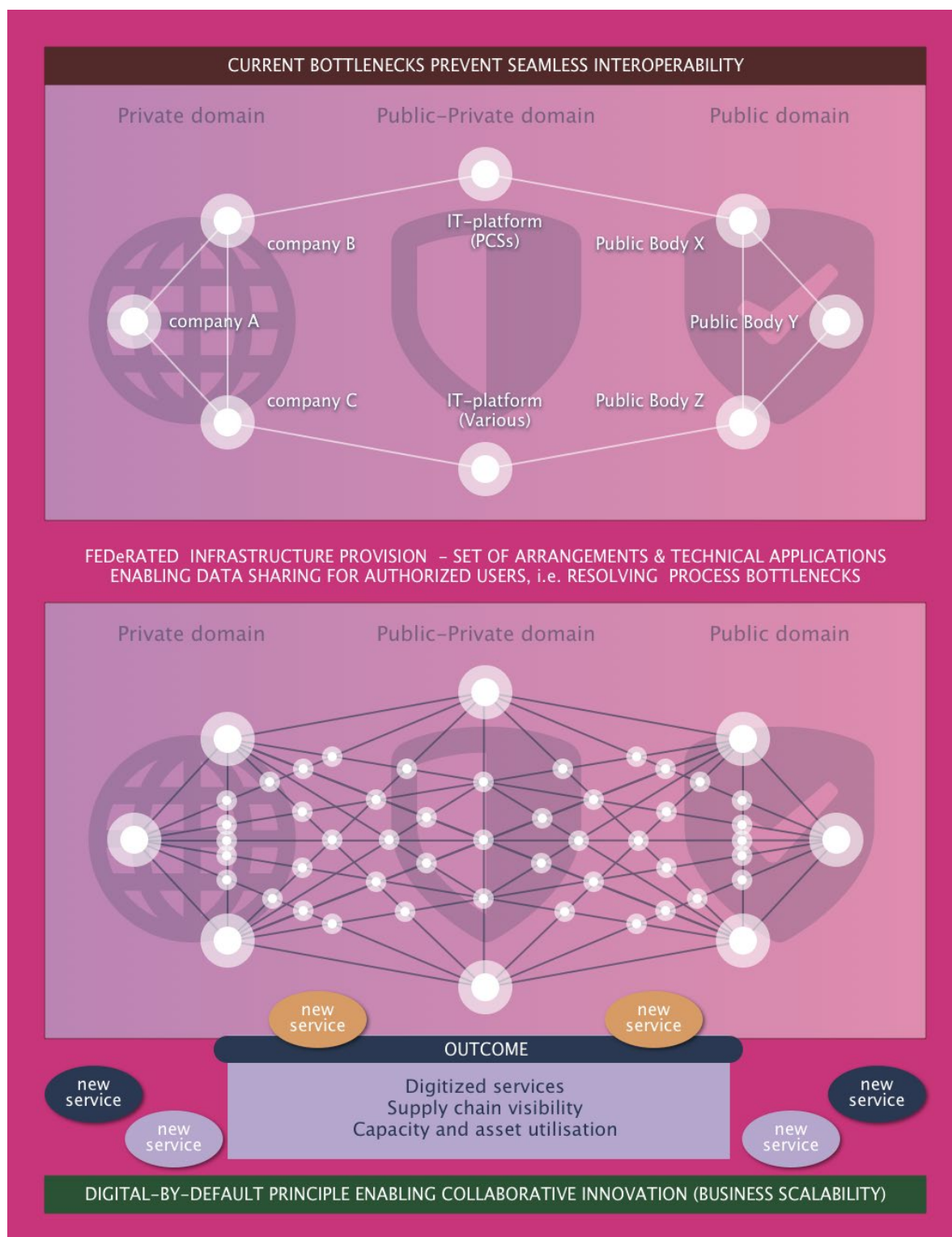
Leading Principles (37)¹² set the basis for the operators to function in the grid, complying with the need for semantic, organisational, technical, and legal interoperability. An operational framework translates these 37 principles into requirements and capabilities.



¹⁰ Data is defined as information in a specific representation, usually as a sequence or symbols. When speaking of computer data, semantics deals with entities, both physical and conceptual and with the relationship between those entities. Ontologies show the properties and the relations between a set of concepts and categories, i.e., automatically deriving data from large data sets.

¹¹ ESG stands for ecology, societal and governance, being identified in the Green Deal implemented and various pieces of EU legislation like Directive (EU) 2022/2464 as regards corporate sustainability reporting.

¹² Available at FEDeRATED website – Chapter: <https://federatedplatforms.eu/index.php/products>



The FEDeRATED Vision in a nutshell

The Operational Framework

The operational framework for federated data sharing sets the requirements to realize federated (decentralized) real time data availability (pull data) for authorized users for seamless multimodal freight transport and logistics.

A. The **organisational requirements** are:

1. Stakeholder engagement - identification, interaction, and involvement.
2. Valid business cases based on data exchanges between data users and data holders.
3. An EU and national governance structure, including a set of agreements ¹³ on:
 - a. the collaboration between the stakeholders.
 - b. installing and maintaining hardware and software.
 - c. a manual on how to hold and use the data, also for providing services and fulfilling compliance procedures.

B. The **functional requirements** of the infrastructure provision refer to the need for:

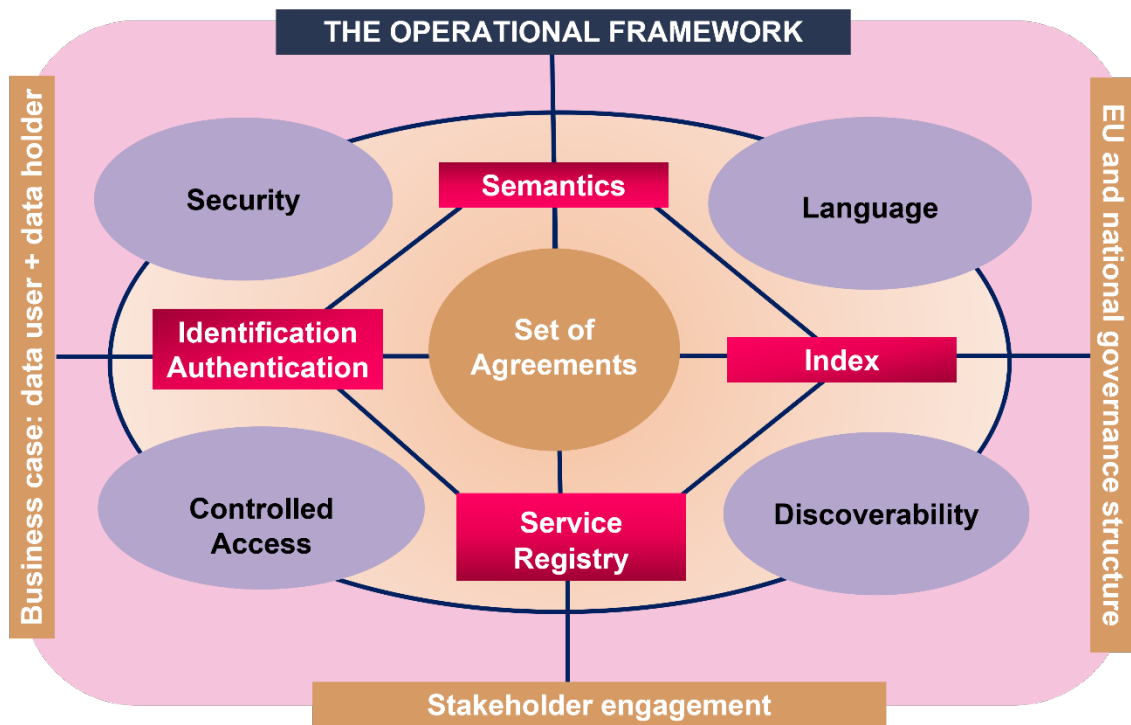
1. “Common” language – the semantics and interaction order (process choreography) for data processing by heterogeneous systems or platforms.
2. Discoverability of data – it is about being able to search and find (query) service providers and data that an organization needs for its tasks. The latter is filled in with 'Linked Data': an organization receives a link to data as an indication of the data they may access.
3. Security for all participants - to provide trust for all participants.
4. Controlled Access to all participants – enabling any company to give another company or competent authorities access to data that either the company is willing to make available to others or need to provide in accordance with legislation. This can be done through open data or via links that have been shared. In practice, this access will be limited, thus controlled access.

C. The **technical specifications – capabilities** - for any data holder or user to participate are:

1. Apply the semantic web technology and a common semantic model (Semantic adapter). Semantics - discussed in the context of semantic web, instead of modelling data – can add contextual meaning around data so it can be better understood, searched, and shared within supply chains, full of varied and complex logistic operations and compliance procedures.
2. Utilize an Identification and Authentication (IAA) infrastructure – the unique identification and authentication of an organization and its authority granted by a recognized registration authority.
3. Apply a Service Registry – enabling organisations to formulate their capabilities, **specify** the maximum of queries, events, and digital twins they can support, identify the infrastructure they use, and the business service(s) they require or support.

¹³ A set of agreements can be structured in various ways like Legal acts, Standards, proprietary Terms of Use, bilateral/multilateral Agreements or legal contracts. Can also be a combination. Based on technology developments, these requirements and specifications will be constantly updated.

4. Deploy an Index – providing any participating organisation a transparent overview of the event-based data being available to share for conducting business and administrative compliance procedures.



Capabilities - Semantics

Exchanging information requires providing meaning and mutual understanding, whereby the virtual – data world – has to fully represent the physical world. This is difficult to do. Often semantics is applied based on standards with different structures. As communities have different (implementation guides of) standards this hinders seamless multimodal data interoperability. To overcome the problem that many transport modes use different standards preventing data from being exchanged an alignment (ontology) framework is necessary.



The Reference model constitutes the basis of the alignment framework (concept of the ontology). It identifies all relevant real-world objects of a logistic and compliance procedure or operation. These objects are translated to a digital reality consisting of 'Digital Twin' and 'Event' data:

- Digital Twin is a taxonomy of real-world objects (container, truck, barge, etc.) and infrastructure.
- Event is the association between two or more Digital Twins in time and space.

The structure of semantics relates to connectivity or rather linkage – i.e., linked data.

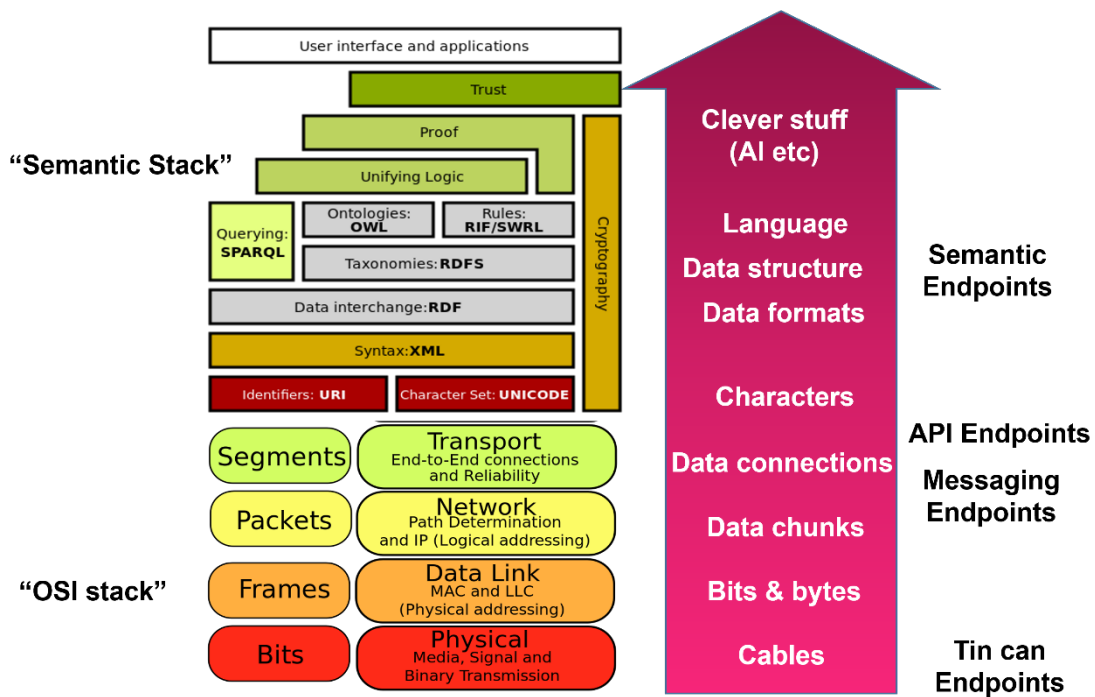
Semantic web technology is promoted. The features:

- The Resource Description Framework (RDF), a standard format to describe and link data. It adds contextual meaning around data, thus discovering relationships in the data.
- Triples. Linked data produces a graph-like representation of data with nodes and edges that are easy to share, easy to combine.
- SPARQL, a standard query language designed for semantic data.
- Triple Store, a semantic database, enabling any organisation to freely use available facts to enhance their own propriety data, and to make interferences by using this combination of data to substantially increase their organizational knowledge.

Based on the above, a FEDeRATED Semantic Model has been developed and is available to be used. It contains an open-linked data repository of general knowledge¹⁴, and a Semantic tree house, which supports the transition from hierarchical message models towards ontologies, based on Linked Data principles and the Semantic Web. See also FEDeRATED Semantic Model - Development Portal.¹⁵

¹⁴ FEDeRATED Semantic Model organizes and presents a massive amount of contained data, available in various knowledge bases of potential data users and holder into triples that can be queried by SPARQL.

¹⁵ Available at <https://federatedplatforms.eu/index.php/products> check [Developer Portal \(federatedplatforms.eu\)](#)



How computers communicate

Capabilities – Service Registry

The Service Registry enables any organization to specify its data requirements and to define the business services it wants to provide. The Service Registry can be applied to specify:

- The business activities of an organization and their interaction patterns (Design).
- An organisational profile presenting its business services, data quality, electing the various lower layer protocols it supports (including endpoints), and access control to data, which links are shared by events, that is stored (Configuration).

The Service Registry must be able to select a technical protocol for data sharing like openAPIs and XSDs.

An interface is required to interconnect. Discoverability is implemented at two levels, based on known SPARQL endpoints of Service Registries:

- **Technical level** - re-use of a design for configuration. SPARQL queries are formulated on the data sharing ontology and query results are specified as constraints to the multimodal ontology by SHACL. Query results may include any choices made in design with respect to a lower layer protocol.
- **Business level** - SPARQL queries represent business goals that match business services. The result of a query also provides access to an organization profile for further support of digital business and compliance.

Any two stakeholders can share data for those parts of their organisational profile that are common - goals and business services can be matched. A user selects the constraints applicable for its organisation i.e., selecting the relevant logistic Digital Twins applicable to its organisation. The endpoints of Service Registries must be trusted.

The choice between API or semantic (SPARQL) endpoints:

API	API requests include fixed parameters that help the server find the right data	When you use API's you have to know exactly what the endpoint is for what formulate you query correctly
	WHEN TO USE: API's and message endpoints are SUFFICIENT for connectivity within a platform or local community	
SPARQL	SPARQL (semantic) requests include a flexible query definition that will be executed by the server to find the right data	When you use SPARQL you can discover what data the server has and formulate different queries to meet your needs
	WHEN TO USE:: SPARQL endpoints are NECESSARY for platforms and organisations cross many borders, languages, and cultures around the globe, facilitating general-purpose network connectivity	

The options to deploy the Service Registry functionality:

- **Minimum** - Design and configuration combined generate and publish an openAPI with an endpoint for a single interaction, like a transport order, a business document, or a visibility event, including a connectivity protocol like CEF eDelivery over TLS (Transport Link Security).
- **Maximum** - Design and configuration are separation where at design time the complete data sharing ontology is applied as input for configuration. Data sharing is implemented with semantic technology, only a (semantic) endpoint is specified



The Yellow Pages – a paper version of a Service Registry

Capabilities – IA

Identity and Authentication or IA is about trust in access to data. The data is business data (e.g., order data), a design, or an organization profile. IA relates to authorization of users, i.e. employees of a participant, and architectural components that provide (access to) data. Safe and secure data transfer is addressed separately by connectivity protocols for the Index.

IA is built upon two pillars¹⁶:

- Organizational trust – each organization that requires to be a node must implement measures that assure trust, for instance cyber security measures and an Identity and Access Management (IAM) registry. Rules for creating this type of trust will be formulated by a legal framework. This covers trust in processes, employees, etc.
- Inter-organisational trust – each organisation must share an identity with another organization that can be verified by that other organization when sharing events, queries, and/or query results.

Authorization is internal to each organization and constitutes the basis for access control. Organisations do not know authorized users of other organizations, but rather trust that authorisation is properly implemented by others (organizational trust).

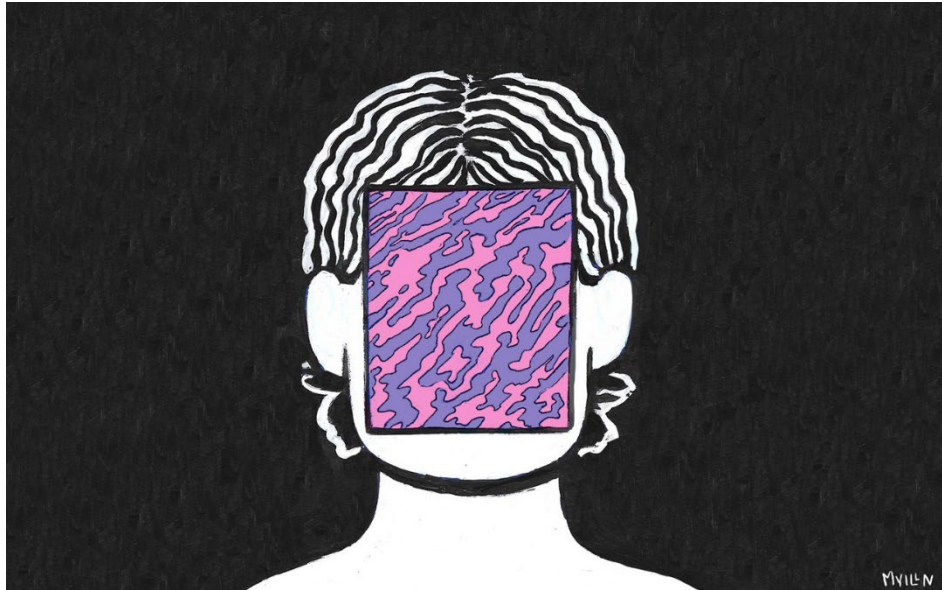
8

Identity and authentication must be based on a completely distributed solution. Each node must have at least one endpoint with inter-organizational trust (Identity and Authentication); - it may have multiple ones (e.g. one for its business services and another one for data sharing). Trust is based on provided and governed by:

- a regulator providing/ establishing a legal data sharing framework as issuing policy of identities, (e.g. EC),
- a trusted registration authority as issuer of verifiable credentials, and
- a certification body for organizational trust (separation of concerns).

The implementation of such a distributed solution based on Verifiable Credentials (VCs) containing an organization profile(s) for plug and play requires alignment with other initiatives. The existing standards, and solutions (like OAuth2.1) can still be applied to create inter-organizational trust (applicable to data of a Service Registry and an Index). This intermediate level requires one or multiple Identity Brokers acting as intermediate Registration Authorities. Preferably, a regulator is a public body.

¹⁶ There is also trust at business level i.e., the trust in properly executing business activities for customers according to agreements made with them. This trust is outside scope of IAA.



Capabilities - Index

The index shares and stores events (with links to data) between a data holder and -user and supports a data user to formulate queries based on links received via events and share these with a data holder.

Each organization has its own private index that stores all events (with links to data) sent as data holder to data users and received as data user from data holders. This index can be implemented by existing IT system(s) of an organization or as a separate (front-end) system like the Node. Events with links to data are shared in a commercial or legal relationship between any two stakeholders. Some events, like an order event, can have links to different types of data like parties involved including their role (shipper, carrier, forwarder), whereas others represent visibility of the execution of a business activity (e.g., an ETA event that links to an order event).

An index supports:

- A data user to retrieve additional data via the links it has received and a data holder to provide a query result by validating the link was shared (link-based access control) and accessing data. The latter is data either stored by a data holder itself or by another organisation (data provenance). In the latter case, a query is federated to the data source (query federation).
- Event distribution - sharing an event with the proper data holder(s) - based on input of a data holder initiating a commercial relation, an existing commercial relation (previous events are stored by an Index), or legal compliance.

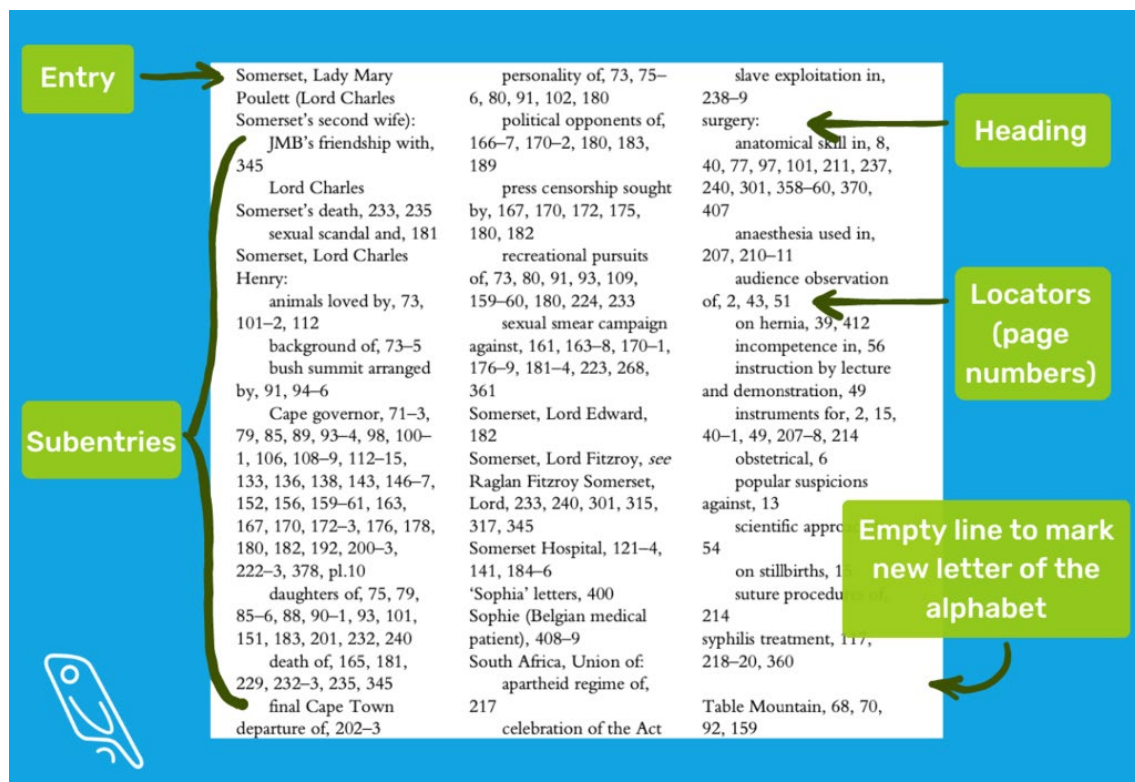
According to the protocols specified by a configurator in its profile, each index must integrate with back-end IT systems of an organization and must implement trusted, safe, and secure data sharing with other Indexes. Integration with back-end IT systems will be supported by the semantic adapter.

The options to deploy the functionalities of an index:

- **Minimum** - to share visibility events with no link to additional data. This is only about progress validating the quality of event data¹⁷.
- **Maximum**¹⁸ - to support:
 - Data quality validation (correctness and completeness of event data and query (results),
 - Event logic (validating the sequence of events),
 - Event distribution (sharing an event with the proper data holder(s)),
 - Enable access for replying to data users' queries (link-based access control),
 - Query federation (data provenance).

¹⁷ This relates to the minimal functionality of the Service Registry.

¹⁸ This maximum functionality requires a complete design (and configuration) of a business activity choreography with the Service Registry. There are all types of options between minimal and maximal functionality, like support of the Index visibility only (one of the business activity choreography phases).

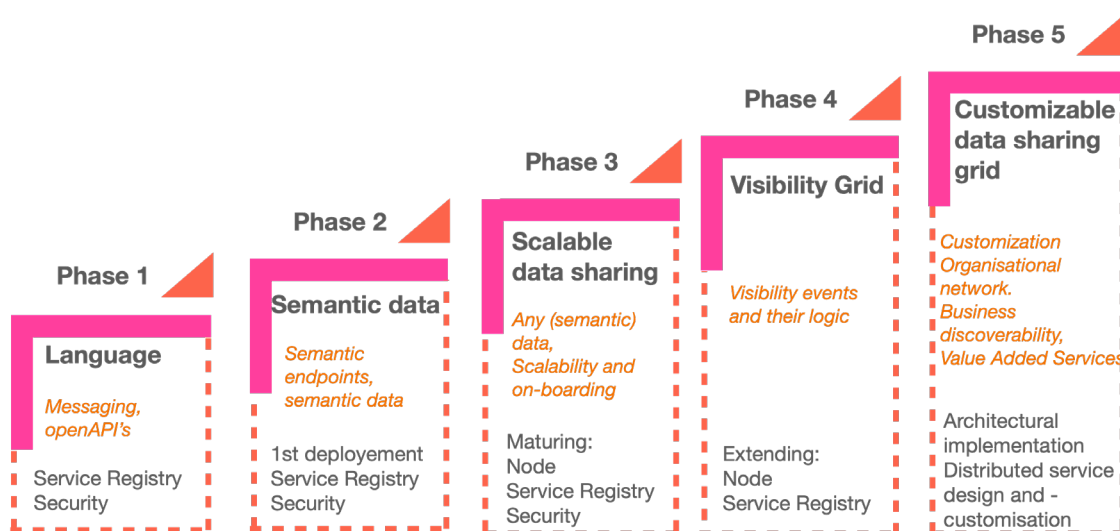


An example of an Index in a paper format, based on the need for the users to maximum query (Index created by Mark Swift for Rachel Holmes' *The Secret Life of Dr James Barry*).

Roadmap - Migration strategy

From EDI message to API webservices to semantic web data sharing practices is a long road ahead. The steps for any organization or platform to evolve as a Node depend on a long-term horizon – return on Investment time scale - requiring a governance structure establishing and supervising a data sharing grid. This governance structure is not specified in this document, which focusses on competence building.

In a migration strategy – roadmap – the starting point for many organizations is their investment in open standards and IT solutions. Many stakeholders are in the process of migrating to APIs for (IT) services to their customers and service providers. Introduction of a federated design thus requires new skills and expertise¹⁹. The steps to be taken towards becoming a participant in an open and neutral data sharing grid for (multimodal) freight transport and logistics are illustrated and explained hereunder.



Migration path - The five steps towards an open, neutral data and secure data sharing grid for freight transport and logistics

1

1. Language – ontologies are applied for specifying messages and/or APIs. This already requires two capabilities, namely the ability to interpret an ontology and specify a message or API with an ontology. A software tool may hide this type of complexity enabling a user to specify an interaction (in a sequence diagram) and generate a message structure or API specification (or API code). Of course, a Design

Authority may provide messages or APIs based on an ontology.

2

2. Semantic data – messaging architecture push data from a data holder to a data recipient, which reflects current business transactions. API architectures already provide a data pull. The Semantic Data Sharing Architecture takes the same approach by sharing events with links to data.

¹⁹ One should reckon with the fact that many SMEs have no knowledge and financial means for investment in this technology available and require ready-to-use applications.

This is about sharing semantic data with the links (data at source). In the case of business transaction data, the implication is that data remains at the source. In case of transaction chains, upstream data must be accessible to downstream stakeholders. It requires federated querying combined with authorization mechanisms and the creation of a semantic endpoint.

3

3. Scalable data sharing – After having created a semantic endpoint and sharing data via events, this is about scaling these applications. It is about creating security: identity and authentication with registration are crucial. It is about the identity of organizations and internal authorization of employees. These provide rapid on-

4

boarding by applying a generic configurable node and creating a grid.

4. Visibility grid – This is about adjusting business processes and IT to standardized interaction patterns for multimodal, multi-cargo visibility (also known as ‘smart contracts’ when applying blockchain technology). A node provides visibility services that can locally be integrated and configured by a participant.

5

5. Customisable data sharing grid – This is about migrating existing interfaces with other parties towards adoption of the capabilities and services²⁰. The technology implemented by an organization must support existing interfaces, to be phased Services can be customized by individual organizations.

²⁰ Services are synonymous to Technology Independent Services

Migration - Stakeholder Engagement

Data sharing is a collaborative effort: interaction and involvement are paramount²¹. Generally, sharing data is between different entities within one organisation – internal collaboration – or between various organisations (companies and public authorities). The goals can be manyfold: (new) business services, legal compliance, process innovation, effective law enforcement, Return on Investment, ESG goals (CO2/NOx emission reduction, less congestion)²², faster lead times, less administrative burdens, more safety and improved emergency response.

Data sharing is about trust (often data sensitivity) dealing with different dimensions:

- Technical - 'is my data only accessible to authorized organizations?'
- Liability – 'what is done with my data?'
- Business - 'do I get paid or is the service/product delivered to me?'

The most important steps to take executing a data sharing project are:

1. The identification of the partners²³, based on a business case (can also be legal compliance, facilitation, etc.).
2. Project definition (scope), i.e. objective, collaboration, governance and continuation, on-boarding, finance, and feedback loop.
3. To define and agree on the services to be provided.
4. To identify the data flows that need to be exchanged between data holders and users of the use case.
5. To apply the appropriate design – identify how to technically share data (i.e. API's, semantics, nodes, digital twins, etc).
6. To implement the appropriate tools (i.e. dashboards, nodes) to allow data to be seamlessly exchanged and shared.
7. To test
8. To deliver
9. To quantify the impacts, such as contributions to policy objectives, benefits, and savings.
10. Communication and evaluation.

Committing various stakeholders to share data requires a lot of care, clear understanding what's in it for who with what purposes, proportionate action, a sound governance structure and a common sense of purpose. Often, stakeholders lack digital savviness, thus jeopardizing data-based logistics project development.

Successful data sharing requires a future proof data sharing design enabling all stakeholder's freedom to operate, supply chain visibility and situational awareness.

²¹ Leading to a common understanding of the problems between stakeholders and shared responsibility

²² ESG stand for ecology, societal and governance, being identified in the Green Deal implemented and various pieces of EU legislation like Directive (EU) 2022/2464 as regards corporate sustainability reporting

²³ It starts with stakeholder commitment. The major operators involved are Shippers, Transporters, Forwarders/agents, Terminal operators, Retailers, and Public authorities. In addition, there can be many third parties involved, such IT Services providers (platforms), software companies, standardization bodies, ports, bankers, insurers, etc.

Between 2019-2023, 5 Member States, transport industry associations and other associations with a particular interest in transport and digitization, e.g., trade associations, were consulted on a basis of a collaborative stakeholder ship in the FEDeRATED project. Respondents represent a cross-section of industries both directly and indirectly concerned by transport visibility.

The key conclusions of the consultation process can be summarised as follows:

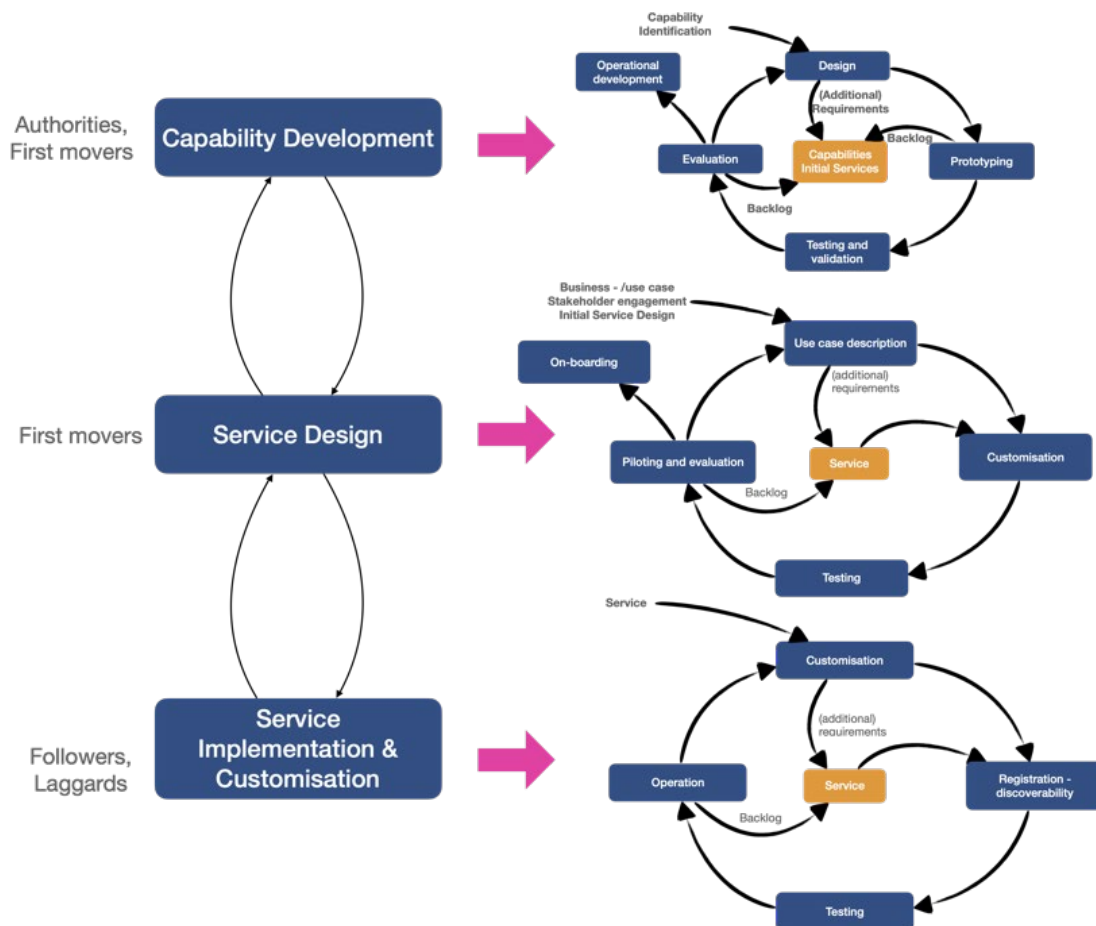
1. Supply chain visibility has become a serious issue which needs addressing.
2. Possible EU visibility measures should focus on real time data exchange mechanism rather than on technical specifications and data sets only.
3. Digital readiness assessment is important and should be further studied.
4. No guarantee for structurally enhanced visibility is likely to be achieved in the medium term.
5. The measures must consider the realities of the market.
6. Any measure should be EU wide to avoid distortion between markets and, as far as feasible, apply to all transport modes.
7. A harmonizing and decentralized EU approach for effective and similar control between various organisations and traders is appreciated, aimed at realizing a systematic approach.

Migration – Adoption Strategy

A data sharing grid is like the creation of the Internet or the introduction of the first phones; it consists of services²⁴ provided by the capabilities. Initially, capabilities need to be developed and validated with first versions of potential services. There will not be many users of new services; the intention is to create a snowball effect.

Innovation adoption theory distinguishes three types of users: first movers, followers, and laggards. Like with all innovations, first movers are approached. They are (most often large) organizations willing to experiment with new services with sufficient financial means and IT technology.

Authorities are the ones that require a data sharing grid. They interface with a great many enterprises of all sizes for regulations that re-use business data. Access to business data by authorities is a driver for enterprises to also digitize business-to-business (B2B) relations. It is basically about multimodal visibility both in B2B and B2A (business-to-administration)!



12

Implementation of a visibility service can be followed by other services like optimization of capacity utilization and booking services. The data sharing grid supports these

²⁴ Services are synonymous to Technology Independent Services. Customisation makes these services, amongst others, technology dependent.

developments and thus capabilities development is initially driven by authorities, supported by first movers. After having developed prototypes and validated the design, software – and services providers can adapt and further develop these capabilities. All requires governance by an independent body.

Innovation adoption is about investigating the potential of a service in a data grid with the intention of implementation. It requires Living Labs, where the potential impact of new services on for instance IT and business processes are exploited. Innovation adoption also tells us to avoid centralized functionality and focus on functionality that does not affect existing business models of participants. Innovation adoption takes time!

Use cases are crucial in shaping new services like a multimodal visibility service. Each use case requires a selection and an initial design of those services that can be applied by many use cases in more than one Living Lab. Each use case finetunes the functionality of a new service of the data sharing grid.

FEDeRATED has selected and made an initial design of multimodal visibility as the initial service of a data grid. It is required by most organizations, but not yet implemented as scalable by all that have such a service. This multimodal visibility service covers all modalities and cargo types and interfaces with other similar services designed by others. A first mover can select its application for a type of cargo and/or modality based on a variety of criteria as a means of exploiting and finetuning this service. Waste reduction can be a selection criterium for a use case, like a focus on logistics of food. Product quality is another example of a selection criterium, thus focusing on logistics of perishables. The transport of high value products and commodity transport are other criteria, where commodity logistics will most likely already have low costs but might be missing certain aspects. Think of container transport where end-to-end visibility is not yet supported.

First movers must be willing to act as opinion leaders and make results available as best practices to the followers. There is a main difference between first movers and followers: first movers can finetune a new service that must be made available as commodity to followers. These followers are the many Small and Medium-sized Enterprises (SMEs) we have in Europe. Most of them don't have sufficient means and knowledge to invest in the development of new services. They must be able to install new services, configure them, and make use of them. The services of a data sharing grid have become a commodity. Nodes supporting (smart) apps for services can be downloaded and customized by an SME to become a participant in the data sharing grid.

Capabilities and services can be upgraded, and new capabilities and services can be developed. Governance must be in place for this cyclic development.

Migration - Legal Framework

Real-time data exchange has become a vital worldwide issue enabling transport, cargo and container visibility. It concerns the European Union whose role as trading partner relies on effective, secure and green transport by all modes and at all levels.

The use of real-time data to improve European transport and cargo visibility has already seen considerable recent improvements. In recent years there have been various efforts to improve supply chain visibility levels in a limited number of well-identified key areas. Within the Customs domain, many years of experiences have geared up the volume of seamless digital reporting. Considerable progress has been achieved recently enhancing maritime reporting to ports and public authorities. In aviation the IATA One Record standard is developed enabling paper to be replaced by data. In road, rail and inland waterways transport data is slowly replacing paper as major information source.

In 2019, the EMSWe Regulation (European Maritime Single Window environment) was approved enabling a more cohesive and harmonized transport and cargo tracking approach for and between EU Member States, thereby also pursuing collaboration between transport and Customs authorities. In 2020, the eFTI Regulation (electronic Freight Transport Information) was approved by European Parliament and EU Member States. The eFTI Regulation will enable companies the opportunity to electronically – paperless – comply with a great many B2A legal compliance procedures for various transport activities.

Many EU legal acts relating to ESG (environmental, society and governance) issues require monitoring progress and compliance, also in connection to specific transport modes. The proposal for transport emission monitoring will be set in place, requiring data to be exchanged between stakeholders and product owners. Most logistics centers and logistics operators have invested in data exchange technologies and applying service level agreements with platform providers. Access control rules, through the focus on data sovereignty, are being implemented and many operators have introduced identity, authentication, and authorization eFTI procedures for employees and third parties. There is growing awareness that digitization could substantially enhance operational brainpower as well as provide new business opportunities.

Overall, these initiatives do not add up to an overarching EU framework enabling companies and public authorities to fully benefit from the opportunities real time data provide to supply chain operators and public authorities towards establishing supply chain visibility. In addition, new legal initiatives and sustainability goals make it increasingly important for operators connected to the supply chain to create full transparency.

All these developments are welcomed. However, they are limited in scope, and do not result in a systematic interoperability approach. The FEDeRATED Action has provided further insight into how adoption of the DTLF federated network of platforms approach the potential must greatly contribute to overall Supply Chain Visibility through collaboration with, and involvement of, stakeholders not limited to their current sphere of operation, rather reaching out to enable data sharing both within a multimodal transport perspective as well as embracing the entire chain from manufacture to delivery.

A common EU reference methodology, ensuring the data generated within the supply chain operations can be exchanged according to the European Interoperability Framework, is lacking. This often leads to a significant discrepancy in data availability and access between the various stakeholders connected to the supply chain. As a result, the current state of play in digitized business transactions, and related compliance procedures, lacks the opportunity for providing visibility. To this extent it is conceivable that it be more appropriate to develop a Community visibility framework for the supply chain instead of opting for a patchwork approach. The framework concept should contain baseline standards relating to the required capabilities for all operators of the supply chain, as well as specific technical rules where warranted. In all cases, The framework should allow for regular, simple updates.

To develop a proactive EU data policy, such an EU regulatory framework which also coincides with a EU Data Space policy development and multiple digital technology based initiatives, both from EU Member States and business, could be a move in the right direction. Data at source as well as data sovereignty trust, and reliability are important connotation towards developing an operational concept, i.e., the development towards an EU Mobility Data Space in connection to the EC strategy on Smart and sustainable transport is ongoing.

Despite these policy developments, also in connection to various studies and practical experience and lessons learnt established in many EU projects, it must be recognized that most supply chain operators and public authorities are reluctant to fully engage in digitization. A poor level of digital readiness in many segments of transport has been identified. To combat this lack of competence, a digital strengthening within all forms of transport systems, including the enhancement of the legal framework and the improvement of preventive mechanisms, is called for. Any such community measure should:

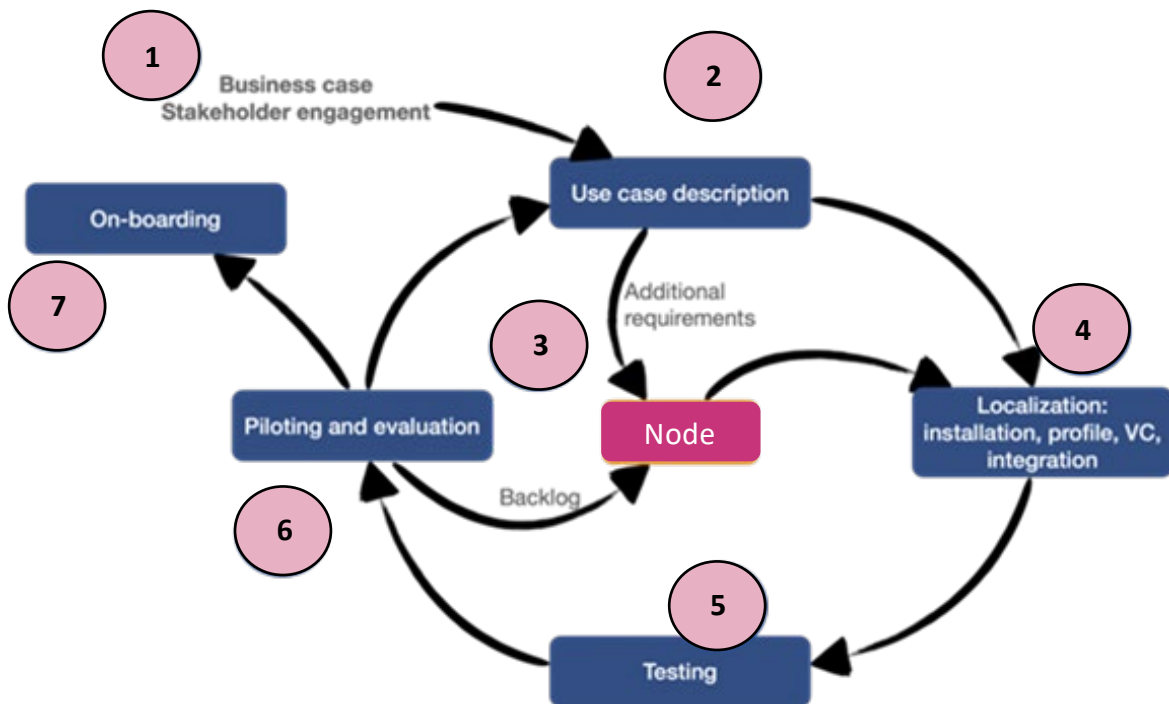
- Consist of a framework for supply chain visibility. It should not limit itself to addressing specific areas but include these areas in an appropriate way to allow regular updates.
- Must strike a balance between highly prescriptive total visibility and the need to ensure a free flow of trade whilst allowing for a gradual tightening and harmonization of baseline data sharing standards or interoperability requirements.
- Cover all freight transport operations and a low threshold to participate.
- Ensure every operator in the supply chain is responsible for providing or using real-time data for its own actions. It cannot renounce this responsibility and be made responsible for other operators' activities. The aggregate of individual measures to manage visibility provides for the visibility standard of the complete supply chain.
- Recognize that for the execution of their public tasks, the state is dependent on supply chain visibility. Supply chain management is industry's responsibility. A cooperative state/industry approach is necessary.
- Adopt the federative approach and thereby set specific standards for operators involved in the supply chain for complying with the minimum set of capabilities of a Node to share data in a trusted and federative manner. The four capabilities required within the various business operations are:
 - Semantics – enabling operators to communicate in a common language based on linked data.
 - Service Registry – enabling operators discoverability.
 - Identification, and Authentication – providing operators security.
 - Index – enabling organizations to share links to their business process data in a machine-readable format (M2M).

Getting started – Implementation

The federated approach aims to assist organisations to evolve as a node in a data sharing grid. The building of a grid²⁵ and the investment of any organisations or platforms aspiring to become a node go together.

The Node can be a data holder and a data user (or both), providing services, engaging in business transactions and fulfilling compliance procedures. The Node integrates the various capabilities, enabling smooth multimodal or logistics operations based on decentralized data to take shape. The context is the supply chain. The goal being full interconnectivity, or rather supply chain visibility (SCV).

The onboarding process for organisations or platforms to function as a Node starts with piloting based on a use case. Some major steps are illustrated hereunder, followed by some explanatory texts:



1 Step 1. Business case – stakeholder engagement - To function as a Node an organisation needs to have incorporated part of the semantic model for its business activities. The business case – sense of purpose – to do so should be clear, as well as connecting to various stakeholders.

2 Step 2. Describing a use case - Sequence diagrams and data models

are developed, although message models are also used. Use case descriptions formulate requirements that must be mapped with interaction patterns and data semantics enabling use case participants to become a Node. Any interaction structures based on the FEDeRATED Semantic Model are specified by the FEDeRATED semantic tree house.

²⁵ Based on the organisational requirements

1. Search for existing interaction structures that match the use case requirements. There may already be other use case with similar requirements that have formulated interaction structures. These can be re-used.
2. Specifying new interaction structures with the FEDeRATED Semantic Model:
 - The root of the tree must be selected. An interaction representing a trip of a truck starts for instance from 'truck'.
 - All related data must be selected. In the example of 'trip', the locations and their addresses that will be called upon by that truck at a date/time should be selected. For a trip, an 'event' is selected associating a truck to a location at a time. Additionally, data of cargo to be pickup and/or dropped of a location might be required.
 - Specification: selecting particular code values like applicable location codes (only those in country X for a truck operating in country X) and the packaging types.
3. The mapping of the data generates a code based on a tree structure using:
 - openAPI code – an openAPI with for instance a JSON syntax – to perform the necessary data validation.
 - Semantic validation – code (SHACL – SHape Constraint

Language) – for semantic data.

- A Semantic adapter for transforming JSON to semantic data. Each interaction structure is represented by its SHACL document and can thus be shared amongst stakeholders. To find these structures, metadata is included like its function in an interaction pattern.

3

Step 3. Validate the Node functionality.

The previous step shows how to implement interaction structures. Potentially, new functionality is required not yet supported by the Node, e.g. event logic.

4

Step 4. Localisation, installation, profile, Verified Credentials (VC)²⁶. Each organisation or platform integrates the various capabilities into its technical setting to start sharing data as a Node. Individual stakeholders may also integrate the required capabilities in their IT environment with local APIs.

5

Step 5. Technical testing of the various components and their integration with IT back-end systems.

6

Step 6. Piloting and testing the operational applicability of the solution. This is at business level, potentially with a limited number of partners.

7

Step 7. Operational application of the solution and on-boarding of other partners.

²⁶ To act as a Node a data sharing grid should be in place. This requires a border crossing and national governance, probably based on an EU or global legal setting. Such framework would

enable any participant to register itself according to its business activity(-ies) and receive credentials for participation after validation (or certification) by an external body.

Tools – Semantic Adapter

A semantic adapter supports the transformation of any data format and structure of an IT system with the FEDeRATED Semantic Model. A semantic adapter is part of the local interface of an organization or platform to a Node. Data is shared between IT systems of different organisations via the grid.

In addition, a semantic adapter can transform:

- Internal and semantic data to (open or de facto) standards.
- Internal data and semantic data structured according to the FEDeRATED Semantic Model in case an organization implements an index with semantic technology and shares semantic data (RDF, SPARQL)²⁷.

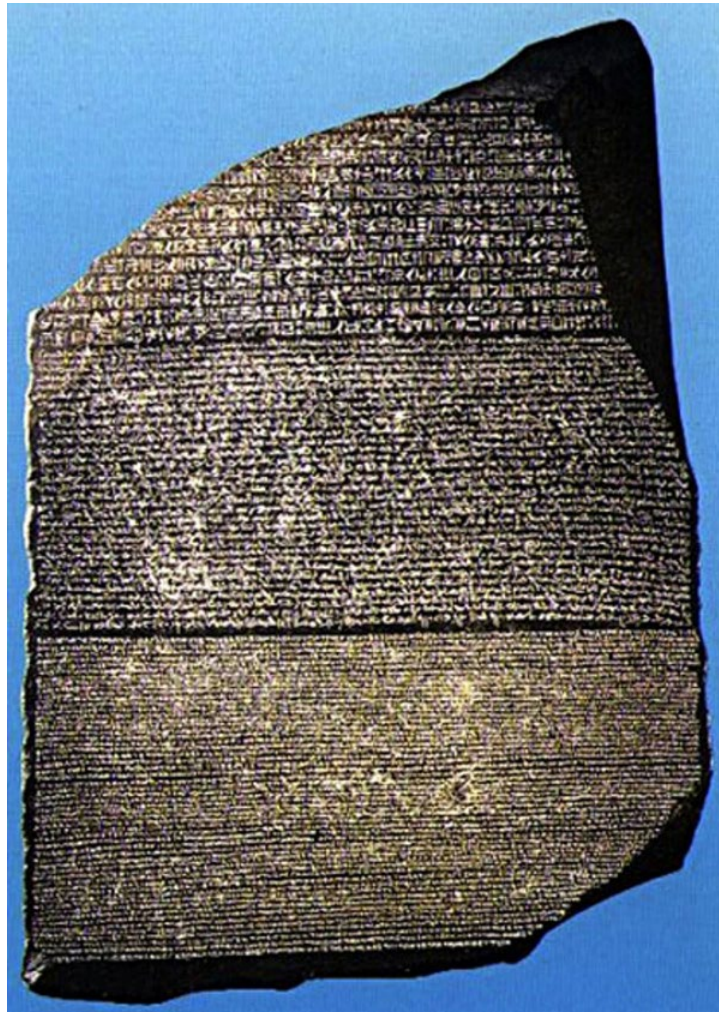
The semantic adapter is part of the local interface, that also consists of processing, like an openAPI and/or batch interface with a local IT system. Any openAPI based local interface must reflect functionality specified by a profile of an organization. In case a batch interface is applied, additional node functionality is required to support events.

Whenever a designer or configurator chooses to implement an interaction of a pattern with a (JSON) openAPI, traditional integration technology can be applied. In that case, the nodes of the grid are not applied, but each organization must implement the openAPIs of others that it would like to interface with.

The options to deploy the functionalities of a semantic adapter:

- **Minimum** - the support of a JSON file structure that reflects the structure of (a part of) the semantic model. It is up to an organisation to interface with the intermediate JSON file structure. This minimal functionality may not yet support a link for querying, since that requires additional mapping functionality.
- **Maximum** - an ontology alignment and mapping tool for configuring an RDF plugin for a relational database, combined with a process engine in case data is stored in different systems. All types of intermediate versions are foreseen, like the mapping of semantic queries to (open)APIs (which is complex) combined with a prescription of openAPIs for an internal IT system (e.g., an openAPI for all relevant parts of the Digital Twin taxonomy implemented by an organization).

²⁷ If you always exchange the same type of data with the same partners you can use an API. If you need to exchange data with new or unknown partners use SPARQL



The Rosetta Stone – Egypt 196 BC - is a stone with three versions of a decree .
The top text is in Ancient Egyptian script using hieroglyphic. The middle texts is in
Demotic script, The bottom text is in Ancient Greek.
The decree has only minor differences between the three versions, making the Rosetta
Stone key to deciphering the Egyptian scripts

Tools - Node Installation

There are a few steps that need to be performed to successfully install a Node. A Node is composed of several components; all these components must be installed by each of the participants in the network. After successful installation of the components a Node has to perform a registration process in order to be able to participate in the network. During this registration process a Node acquires a certificate required for identification and access to the network.

Nodes communicate with each other over TCP (Transmission Communication Protocol)²⁸. A Node needs to have at least a (public) IP address on which it can be reached by other nodes in the network. A Node stores the events it sends and receives in its local (GraphDB) triple store with the semantic model. The Node API can be configured to support events; it will not perform any validation or data transformation when not configured.

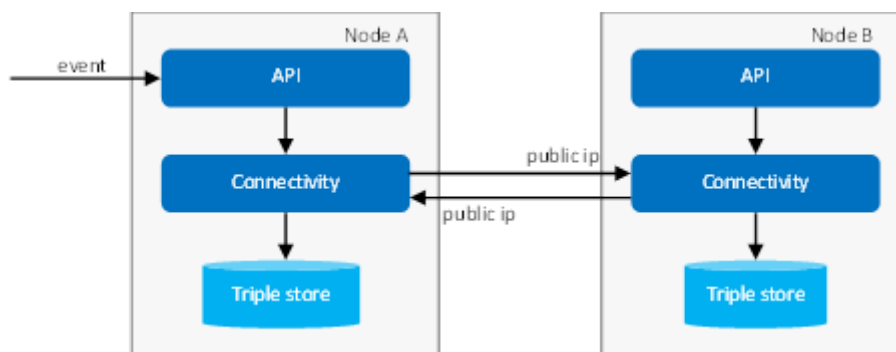


Figure 1: Example of two nodes, the components and communication between them.

For ease of installation, all components are made available as containerized images. The following images must be installed and configured for each Node:

- API
- Connectivity component
- Triplestore

FEDeRATED specific images are hosted on Docker hub: <https://hub.docker.com/u/federatedbdj>. Installation and configuration instructions are available on the FEDeRATED Github page: <https://github.com/Federated-BDI/Docker-BDI-Node>.

There are Helm scripts available for installation of a Node on a Kubernetes cluster: <https://github.com/Federated-BDI/Kubernetes-BDI-Node>. Note that these scripts might have to be modified to match specific infrastructure requirements.

²⁸ TCP – The Transmission Communication Protocol is an important part of the Internet protocol suite. It is a transport layer that facilitates the transmission of package from source to destination.



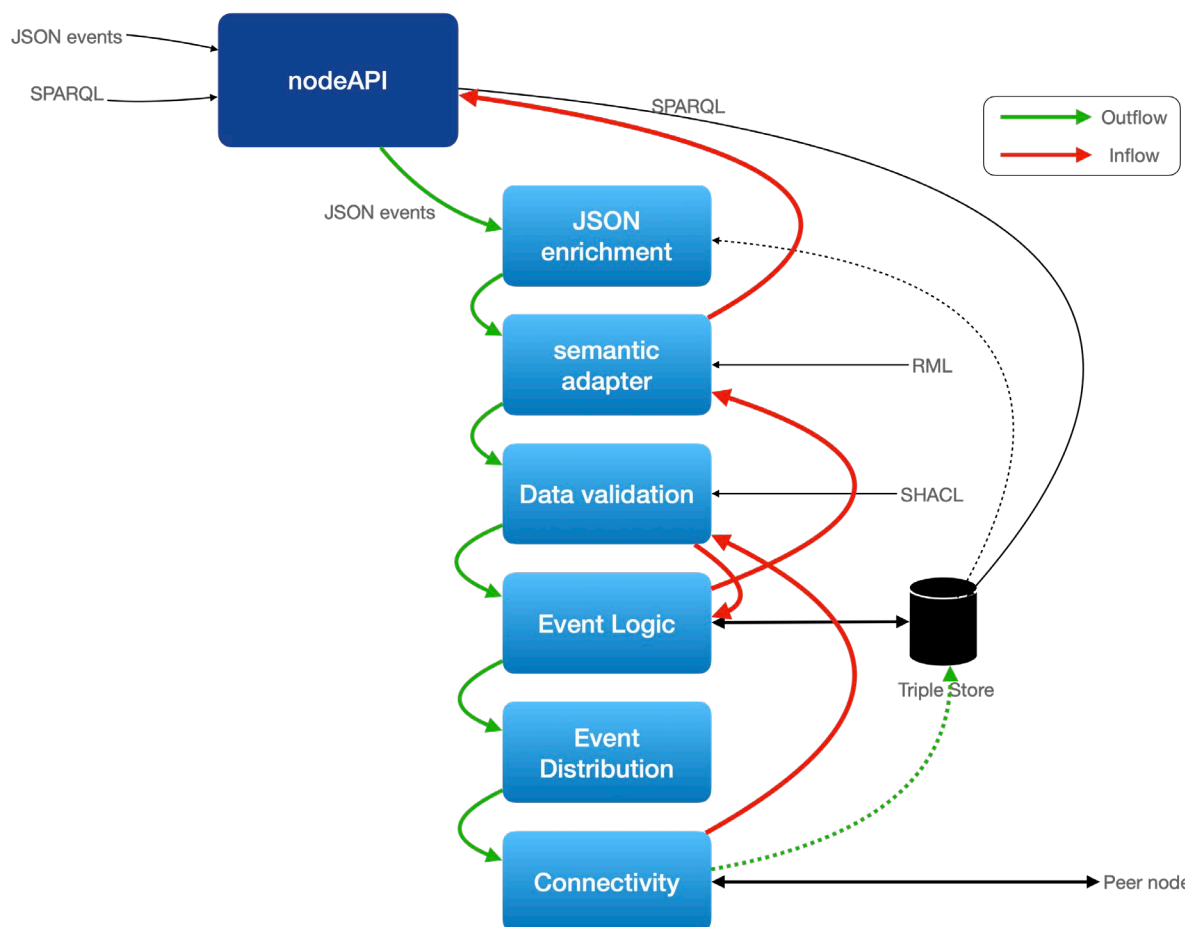
*Node Installation will requires dedicated follow-up actions,
incl. maintenance and collaboration.
Deployment in splendid isolation is not an option.
Interconnectivity is essential.*

Tools - Node Interfacing

In a grid, a node supports sharing events with links to additional data and queries to retrieve that data. Each node can share data with all other nodes in the network, using a connectivity protocol which is independent of the data that is shared.

Each node provides an eventAPI to internal IT systems. This eventAPI can be configured for:

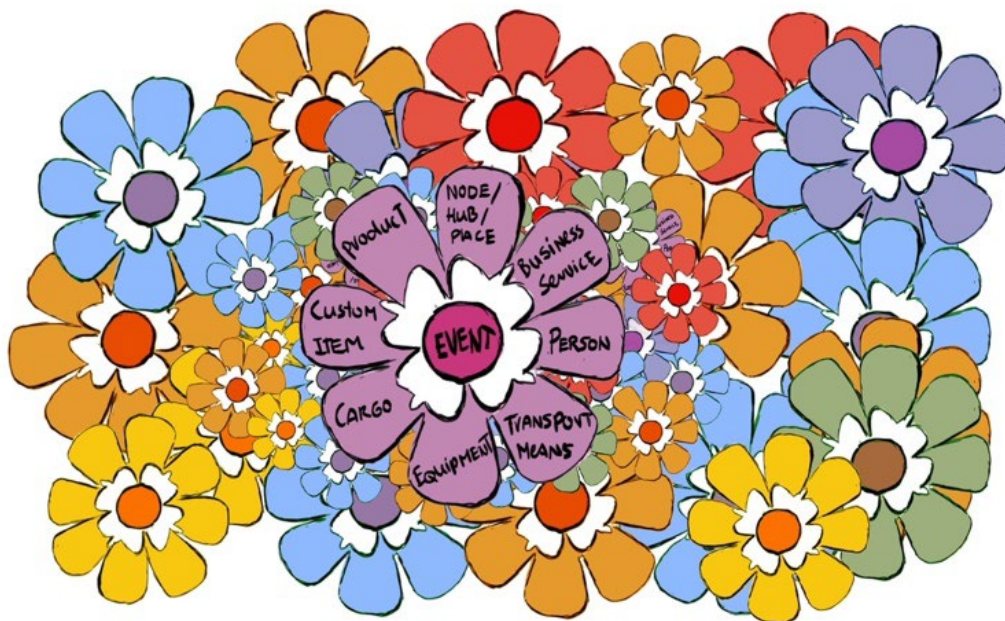
- Data transformation by the semantic adapter.
- Data validation based on SHACL specifications representing event data structures.
- Event logic to validate the processing according to an interaction pattern.
- Event distribution to route events to the proper nodes.
- Querying by storing shared events in a triple store (index functionality).



JSON enrichment is to assure that the proper identifications are inserted, the UUIDs of the events. The semantic adapter is configured by RML (Rule Markup Language), and supports a JSON type of interface reflecting the semantic model. Both RML and SHACL are generated by a tool.

The current version of the node supports the Corda connectivity functionality. It is a Corda proprietary protocol for peer-to-peer data sharing using Corda Identifiers and message queueing (AMQP) over TLS (Transport Link Security). Corda Identifiers are issued at registration. Additional, non-repudiation is supported by the Corda Notary Network, an underlying blockchain network running on existing Corda nodes.

Event logic is not yet implemented. Queries are shared using Corda and need to be mapped into openAPIs of existing IT systems.



"Letting a 100 flower (and more) bloom" is feasible when structured around 21st Century proof semantics.

www.federatedplatforms.eu

