Data sharing in supply and logistics as commodity

The Digital Transport and Logistics Forum Second Mandate (DTLF II)
Subgroup 2: Corridor Information Systems

Interim report

– Final –

Table of contents

Table of Figures

List of abbreviations

ABBR	Description
API	Application Programming Interface
B2B	Business-to-Business
B2G	Business-to-Government
CEF	Connecting European Facilities
DTLF	Digital Transport and Logistics Forum
EC	European Commission
EU	European Union
EU CDM	EU Customs Data Model
G2B	Government-to-Business
GDPR	General Data Protection Regulation
PCS	Port Community System
SME	Small and Medium sized Enterprise
TCO	Total Cost of Ownership
WCO	World Customs Organization

1 Introduction

1.1 Federated network of platforms: context and proposed solution

1.1.1 Context

The creation of an accessible and neutral solution to facilitate data sharing in the supply chain and logistics sector faces several challenges that must be addressed.

Digitalization enables appropriate tools to support this process and is an important driver and enabler for efficiency, sustainability, simplification, cost reduction and better utilization of resources and existing infrastructures. Digitalization also creates new opportunities for business and has the potential to change the way cargo and traffic flows will be organized and managed in the future.

To be effective, and to serve properly the needs of business and authorities, the future solution for seamless data sharing should meet specific conditions and prerequisites:

- Transparency, neutrality, and inclusivity: the federated network must be able to support seamless data sharing by all entities (both private and public sector) operating within the supply chain and logistics sector across the EU, including small and medium-sized enterprises (SMEs).
- Support to the supply chain and logistics sector: focus on enabling the seamless data flow pertaining to the movement of goods and logistics processes, and smart mobility services in transport.
- Support to specific data flows between businesses and authorities: while businesses share data as part of collaboration they also must make data available to authorities for compliance with regulation.
- Innovation: future innovations in the supply chain and logistics should be more easily supported by performant data sharing applications.

These conditions and prerequisites may require businesses and authorities alike to modify their business processes and IT systems.

1.1.2 Challenges

Existing data sharing applications between businesses (B2B) and between businesses and governments (B2G and G2B) encompass sharing of orders, business documents like Air Waybills (AWBs), Bills of Lading (B/Ls), and CMR consignment notes (CMR stands for 'Convention of the Contract for the International Carriage of Goods by Road'), planning data, and support of for instance import/export procedures and safety regulations like dangerous goods declarations.

In practice a wide range of different interfaces are deployed, each often with their own data requirements and representation of semantics. Many of these interfaces are implemented in a bilateral or community setting like (air-)ports, focusing on a transport modality (e.g. air or inland waterways), particular type of cargo (e.g. containers), and/or particular products (e.g. commodities like palm oil and grain).

Many large enterprises (e.g. retailers, suppliers and manufacturers) have developed their own systems and methods for implementing existing open standards, resulting in many variants. Similarly authorities have developed their own guidelines relating to international standards: for instance a

national EU Customs administration will have its own method for accessing the EU Customs Data Model (EU CDM), which is a variant of the World Customs Organization (WCO) data model.

Whilst it is the case that businesses effectively handle these types of data sets with their individual systems and applications, it is also true that their Total Cost of Ownership (TCO) is higher on account of the complexity caused by so many interface variants. This high TCO and complexity reduces business flexibility and is a barrier to business innovation, supply chain resilience and agility.

The current approach also lacks inclusiveness. In general, SMEs have insufficient knowledge and financial means to invest in the current methods of data sharing. They need to have ready- and easy-to-use IT applications. Platform providers will offer these solutions; however SMEs may have to connect to multiple platforms depending on the requirements of their customer base, which leads to higher operating costs.

The lack of digitization within the SMEs category causes challenges for large enterprises. The 80/20 rule can be applied: large enterprises conduct 80% of their business working with a small number of other enterprises and 20% with many SMEs, leading to high operational costs.

The lack of harmonized interfaces also leads to a large variety of platform solutions, each with their interfaces and a focus on a particular market. Having their interfaces for a particular market solution is a barrier to those platform providers wishing to extend their services to new markets.

In short, existing data sharing solutions requires lower cost and needs to be more inclusive. They hinder the creation of an open and neutral data space.

1.1.3 Proposed solution

A solution must be technology-independent in such a way that it can be implemented by various organizations providing and/or governing commercial services in mobility and freight, using an enabler of choice. Such an enabler of choice can be an (innovative) IT service provider or a peer-to-peer solution, using a technology of choice. Such a technology-independent solution is a set of agreements that allows organizations to share data with any other (relevant) party to support their business processes and enable innovation. These agreements consist of:

- Business process collaboration – the set of IT services required to support collaboration of business processes for providing and governing commercial services. These IT services are called the Technology Independent Services; they can be -among other solutions- implemented by REST (REST – Representational State Transfer) Application Programming Interfaces (APIs).
- Common language – both the semantics and their representation for data sharing should be clearly specified to automatically share and process data by different IT systems. This is called the semantic model.
- Identity and Authentication – each organization should have an unique identity that is issued by a certified identity provider and can be authenticated. Multiple identification domains may have to be specified, each based on its certification mechanism supported by an identity broker. eIDAS (electronic IDentification, Authentication and trust Services) is an example where the EU Member States have implemented an agreed certification mechanism for B2G data sharing, both for users and IT systems. Open standards should be applied, in combination

with the implementation of the Technology Independent Services (e.g. OAUTH2.0 and REST API identity tokens).

- Data sovereignty – each enterprise should be able to control its data sharing, compliant with any restrictions (e.g. GDPR) and data requirements of authorities based on regulations (fit for purpose). This is part of access control.
- Discoverability – for inclusiveness and optimization it should be possible to discover commercial information, business services, available logistics capacity, and the past (e.g. a trace or container track), present, and future (e.g. a planned flight, itinerary, or voyage with available capacity) state of supply and logistics chains in networks. State changes are shared via events that support business collaboration. Additionally, information services like weather conditions must be findable.

Platform – and solution interoperability should be provided, i.e. various platforms and solutions should be interoperable to provide an open and neutral environment. This aspect is addressed by constructing a reference architecture focusing on development of the various interfaces between components supporting all functionality identified as part of the agreements.

Implementation of these agreements by IT Service and Solution Providers compliant with the architecture enables individual organizations to register themselves as member of the data space and integrate these agreements with their existing IT solutions. This is called plug and play.

1.2 Organization - the Digital Transport and Logistics Forum (DTLF)

1.2.1 Context and structure of the organization

To contribute to the EU Mobility and Freight data space, DG MOVE established the Digital Transport and Logistics Forum (DTLF) - a consultative platform for the coordination and cooperation between stakeholders in a cross-modal and cross-sectoral perspective. The DTLF produces recommendations to DG MOVE regarding the development of the EU Mobility and Freight data space. The DTLF is currently in its second mandate, DTLF II.

The Forum is subdivided into two subgroups, dealing respectively with paperless transport in the context of the electronic Freight Transport Information (eFTI) Regulation and the establishment of digital corridor information and management systems. Each subgroup is further split into dedicated teams examining specific issues and topics.

DTLF II meets in the form of a Plenary, but all the groundwork is performed by working groups (sub-groups and teams). The activities of the Subgroups are led by Rapporteurs, who stimulate the discussions, coordinate work of the respective Team Leaders, monitor progress and deliverables, and report to the Commission. The Commission maintains the overall oversight and supervision of the Forum and provides a technical secretariat for its operations.

This document was produced by Sub-group 2 based on work undertaken by the various teams within this sub-group.

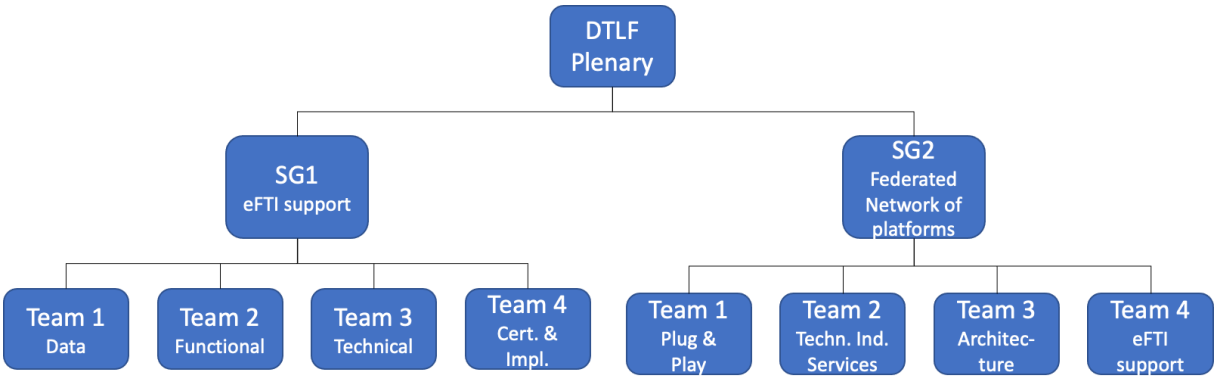


Figure 1 Structure of the Digital Transport and Logistics Forum (DTLF)

1.2.2 Teams for creating the federated network of platforms

In its first mandate, DTLF Sub-Group 2 recommended the creation an open, neutral, and inclusive (i.e. level playing field) data sharing infrastructure constructed by existing solutions like (community/commercial) platforms and peer-to-peer solutions. This federated network of platforms, or what can be called an EU Freight data space, should serve as a commodity. It consists of four main aspects mentioned before, namely:

- Team 1 - plug and play: the ability of individual organizations to use the commodity and share data with all relevant other users.
- Team 2 - Technology Independent Services: the IT services and common language required for business collaboration and compliance with regulations between any two entities.
- Team 3 – architecture: functionality and components with their interfaces to support the technology independent services and plug and play.
- Team 4 - trusted, safe, and secure: any relevant governance structure and procedures to ensure trust that the commodity can be used in a safe and secure manner.

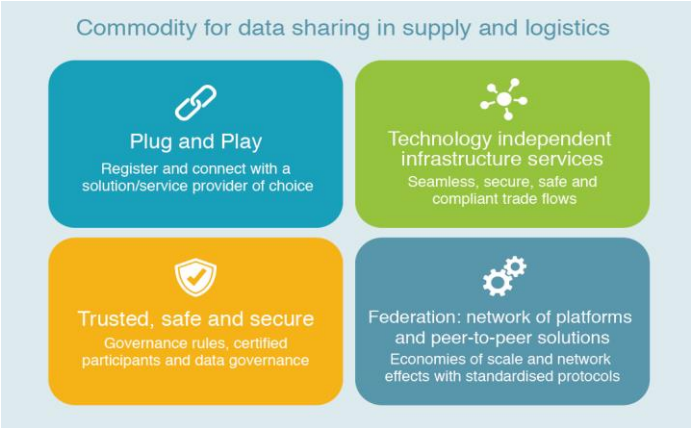


Figure 2 Building elements of the federated network of platforms

The work by the various teams in DTLF II Sub-GroupG2 is supported by the CEF (Connecting European Facilities) funded FEDeRATED (http://www.federatedplatforms.eu) – and FENIX (https://www.fenix-project.eu) Actions. Both Actions also run various Living Labs and pilots to further validate and test the proposed solution in practice.

1.3 This document

The solution proposed by DTLF in this document is called a 'federated network of platforms', and information regarding the solution is available in several forms:

- Leaflet: an overview of the solution and its components.
- Executive summary: a more detailed overview of the solution.
- Annexes: these provide technical details of the various components of the solution.

This document is an Intermediate Report providing technical details regarding the proposed solution as an Annex to the Executive Summary.

The structure of this document is as follows:

- Section 2 – rationale and key requirements underpinning the work.
- Section 3 – data sharing in the supply chain and logistics sector.
- Section 4 – business and authority data sharing.
- Section 5 – modelling data sharing, towards a semantic model(s).
- Section 6 – how to share data, the architecture as currently developed.
- Section 7 – general perspectives such as security. Other perspectives are yet under development.

2 Rationale and key requirements

The rationale for and key requirements underpinning the creation of a federated network of platforms are embedded within the challenges facing the supply chain and logistics sector with regard to data sharing. These are set out below.

2.1 Harmonization of interoperability - EIF

The European Interoperability Framework (EIF) sets out the basic conditions for achieving interoperability, for relevant initiatives at all levels including European, national, regional, and local, embracing public administrations, citizens, and businesses. The authors of the document have opted to follow this framework when creating the specification within this document.

The framework comprises four types of interoperability as shown in Figure 3 – these are then summarized in the subsequent paragraphs.

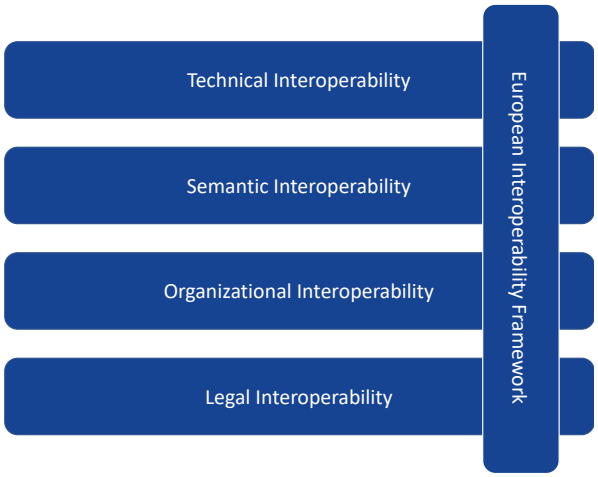


Figure 3 Overview of the European Interoperability Framework

The technical interoperability covers the applications and infrastructures linking systems and services. Aspects of technical interoperability include interface specifications, interconnection services, data integration services, data presentation and exchange, and secure communication protocols.

Semantic interoperability ensures that the precise format and meaning of exchanged data and information is preserved and understood throughout exchanges between parties, in other words 'what is sent is what is understood'. In the EIF, semantic interoperability covers both semantic and syntactic aspects.

The organizational interoperability refers to the way in which public administrations align their business processes, responsibilities, and expectations to achieve commonly agreed and mutually beneficial goals. This framework can of course also be applied in the private sector. In practice, organizational interoperability means documenting and integrating or aligning business processes and relevant information exchanged. Organizational interoperability also aims to meet the requirements of the user community by making services available, easily identifiable, accessible and user focused.

Each public administration implements European regulations in its own national legal framework. Legal interoperability is about ensuring that organizations operating under different legal frameworks, policies and strategies can work together.

The focus of DTLF II Sub-Group2 is on organizational and semantic interoperability, both in the context of business-to-business (B2B), business-to-administration/government (B2A/B2G), and administration/government-to-business (A2B/G2B).

Within this context there are several constraints arising from a range of regulatory instruments: GDPR Regulation and the EU Cyber Security Act, the eIDAS Regulations, and the Data Governance Act. There are also various regulations in the context of freight data sharing: public regulations such as eFTI and private regulations, treaties, or conventions such as the Hague-Visby and UNCECE CMR. These need consideration in the context of data sharing within the supply chain and logistics sector. The challenge for DTLF II is to identify any additional regulation that might be required to support the utilization of a federated network of platforms.

2.2 Federation – open and neutral

There are two specific challenges that are being addressed by DTLF II. The first is the ever-increasing number of data sharing solutions utilised within the supply chain and logistics sector, each with their own governance structure and business model. These solutions provide a variety of IT services with their APIs (Application Programming Interfaces) for a particular customer segment with a view to meeting the needs of those customers. These data sharing solutions come with their own technologies. Whenever an organization requires to share data for a particular purpose with another organization and both use different solutions or platforms, they must connect to both solutions. The variations within IT services with respect to data semantics, process functionality, and supporting APIs also create a potential risk of vendor lock-in: it can be expensive to change from one platform to another.

The second challenge is the ability to integrate business processes within the supply chain and logistics sector. Whilst authorities specify regulation at EU level pertaining to data sharing protocols, at Member State level the national guidelines on implementation can vary, making it difficult for enterprises operating in the EU. This aspect can only be dealt with through legal intervention, possibly supported by a common IT architecture. Another aspect in this context is the bilateral or multilateral (community) agreements made between enterprises for applying an open standard in data sharing. DTLF I Sub-Group 2 has concluded that such agreements would also lead to conflicting guidance on implementation and therefore not contribute to interoperability nor openness.

So, what is needed is to enable optimal use of existing solutions and create an open and neutral data sharing infrastructure? The proposed solution is based on the creation of harmonized Technology Independent Services that can be implemented by platforms and solutions and facilities that enable organizations to integrate with a solution of choice and be able to share data with any other organization, without common agreements. This will support rapid on-boarding of any organization to the federation of networks.

For organizations to be findable in such a network, additional services need to be developed. These are all identified by the architecture. Since there is a variety of platforms and solutions, organizations sharing data require a degree of trust in the overall solution. Furthermore, data sharing needs to be safe and secure.

To address the issues above, the teams within Sub-Group 2 envisage that there could be four types of implementations:

A. Peer-to-peer data sharing: organizations use their own internal solutions to share data with each other. They must implement identified interfaces and components of the architecture themselves. There is no shared functionality.
B. Single platform – each organization interfaces with a single platform, where the platform implements (a subset of) the Technology Independent Services.
C. Multiple platforms – each organization connects to a platform of choice and can share data (via another platform) with another organization.
D. A combination of peer-to-peer and a platform – one organization uses a platform and another its own data sharing solution. They must interface with one or more platforms and other peer-to-peer solutions.

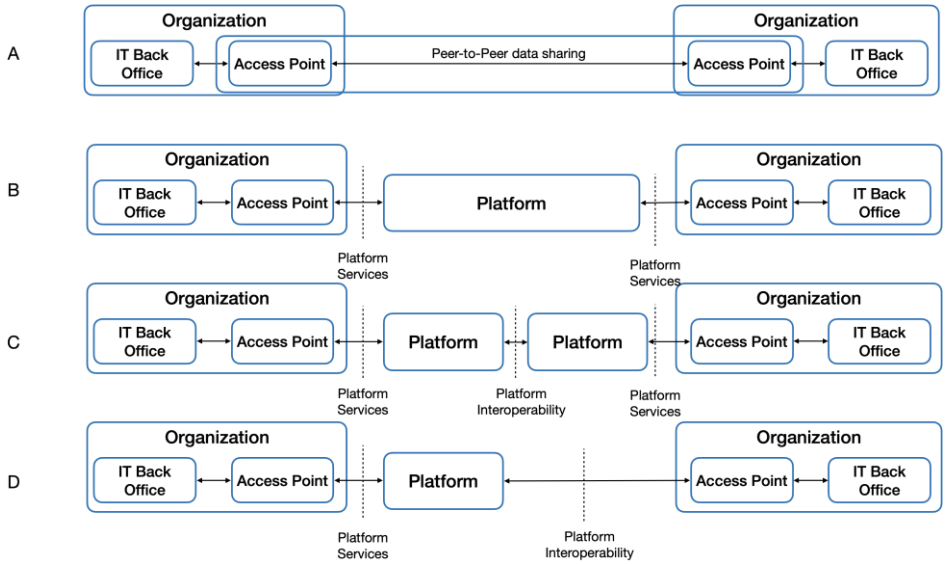


Figure 4 Implementation variants

2.3 Level playing field

Besides being open and neutral, the federated network of platforms needs to be inclusive towards SMEs. A level playing field enabling SMEs to use a solution of choice and be able to share data with all others needs to be developed.

A level playing field is based on harmonized Technology Independent Services and the ability to register once and be able to share data with all other relevant organizations based on a common set of agreements, trust, and safe and secure data sharing via a network of platforms and solutions.

2.4 Innovation – creating a market

Finally, harmonization of Technology Independent Services and federation of platforms creates a market for innovative solutions. The level playing field enlarges the market by including SMEs, which might lower prices for data sharing and lead to a commodity for data sharing supporting the principles with respect to data sharing (e.g. data sovereignty).

Having these Technology Independent Services also enables development of new services, for example value added services for propagating event data generated by a truck or vessel to customers and administrations. These services can of course also be developed by existing platforms and solutions, thus distinguishing themselves from other providers. The challenge for existing solutions is to create value with the data that is shared with consent of their customers.

Innovation should also address the fact that a solution must evolve. Where the focus is currently on commercial (business) transactions between organizations, these might also be supported by autonomous vehicles, for instance to support truck platooning and corridor management. Any solution should be able to be extendable with new functionality. This puts a requirement on platforms to rapidly deploy these new services or any changes in existing services, to support innovation.

2.5 Solution and architectural approach

The solution referred to in the introduction comprises some key elements of functionality, as shown in Figure 5.

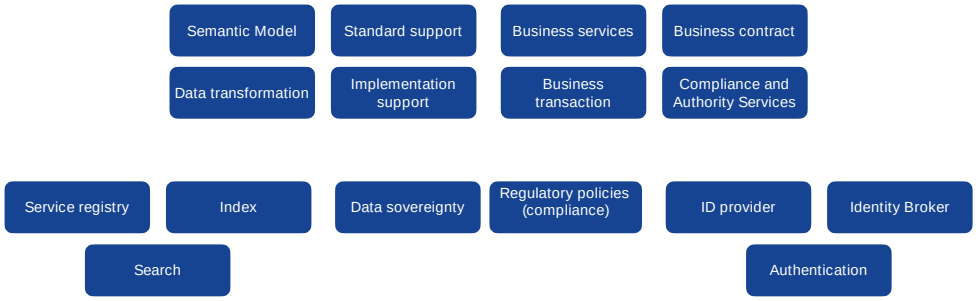


Figure 5 Key elements of functionality within the solution

The functionality comprises:

- Process aspects which support business services resulting in business contracts with business transactions. These must be compliant with regulations and need to support the necessary data for compliance.
- Common language is disaggregated within a semantic model(s) to support data sharing in the supply chain and logistics sector. These must be mapped to existing standards and organizations have to configure data transformation according to plug and play. Each enterprise must provide relevant data that complies with applicable regulations and supports business services. To assist organizations, tools for implementation support must be provided.
- Identity and Authentication is broken down in Identity Providers and Identify Brokers supporting Authentication.

- Accessibility covers data sovereignty for business-to-business data sharing and data access by authorities in relation to regulatory compliance.
- Discoverability mainly consists of an index for sharing event data and a service registry for publication of business services. Search functionality will have to be specified on each of these components, where the functionality is implemented in a distributed way.
- With respect to data sharing solutions, these must comply to (a relevant subset of) the aforementioned components and will have minimal implementation requirements.

Another way of visualizing this functionality is by means of an IT architecture that integrates the various components. The architecture has to describe the components and their interfaces from the perspective of data sharing, i.e. in the context of the Mobility and Freight data space. Thus, the focus will be on the various interfaces between solutions, supporting different types of use.

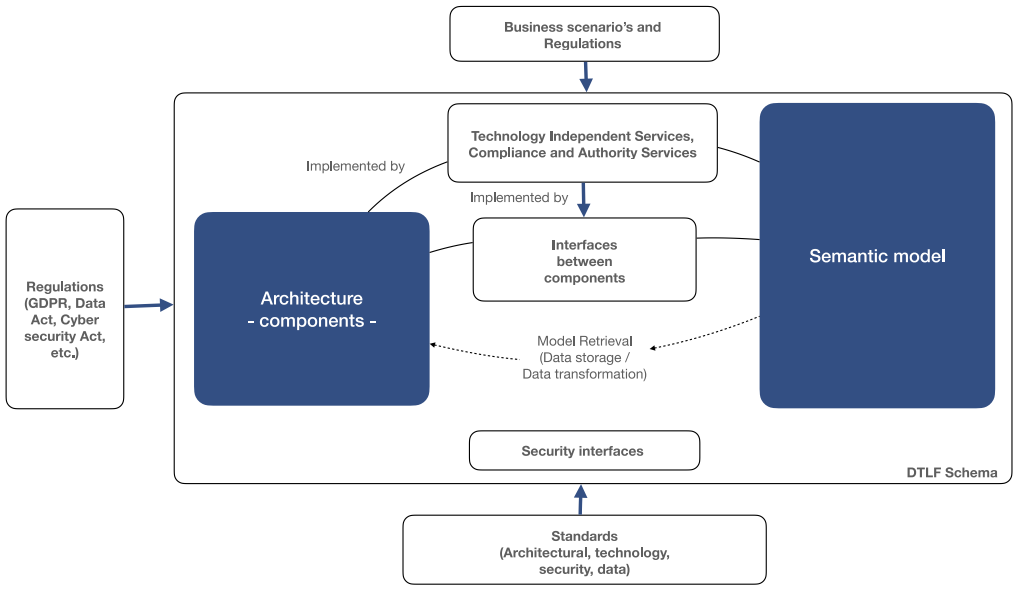


Figure 6 Architectural approach

In this architectural approach, there are three types of inputs for formulating the so-called DTLF Schema (see section 3.3), namely data sharing requirements formulated by business scenarios and Regulations; any other relevant regulations with respect to data sharing and architectural, technology, security and data standards. This approach will be further elaborated in the next sections of this report.

3 Data sharing in the supply chain and logistics sector

Governance rules must be defined for common resources, where transport is an example of a resource. Data sharing provides access to common resources.

3.1 Overview and definitions

In the context of supply chain and logistics, a common resource is a physical resource. Data is shared amongst stakeholders to enable the compliant and efficient utilization of these physical resources. Figure 7 provides guidance on the development of governance.

The following layers are distinguished (bottom up):

1. Physical resources: includes all physical means to perform transport and logistics operations and execute contractually agreed orders: infrastructure, assets and business resources (e.g. personnel). These resources have harmonized definitions and characteristics that allow their owners/managers to offer them as a (per use) service and users to match them to specific operations and assignments. Each resource has capacity, e.g. to transport containers via sea. This layer has to provide details of its performance via for instance sensors, to its owners, where these owners may decide to use platforms to integrate this sensor data with their internal resources (e.g. AIS (Automatic Identification System) – and OBU (On Board Unit) service – and data providers).
2. Physical end-to-end transport process flow: order–ship–pay. The actual business process steps supporting operational execution of the various forms of contract (e.g. framework contracts and spot market), utilizing the digital end-to-end transport process flow layer for data sharing between any two organizations.
3. Digital end-to-end transport process flow: the reference process “order-ship-pay” is a basis of the functionality provided by this layer. The necessary data sharing infrastructure consisting of (community and commercial) platforms and peer-to-peer connectors that allow stakeholders to perform processes described in layer 2 and use of resources described in a layer 1.
4. Business services/business relation layer: the set of business services (e.g. transport, transhipment, storage, including how they are published like timetables and business services) provided by Logistics Service Providers (LSPs) and required by their customers. Authorities are also included in this layer, based on data requirements formulation for Regulations. The previous layer should support this layer by offering Technology Independent Services for the physical end-to-end transport process flow.

Freight Data Space - layering

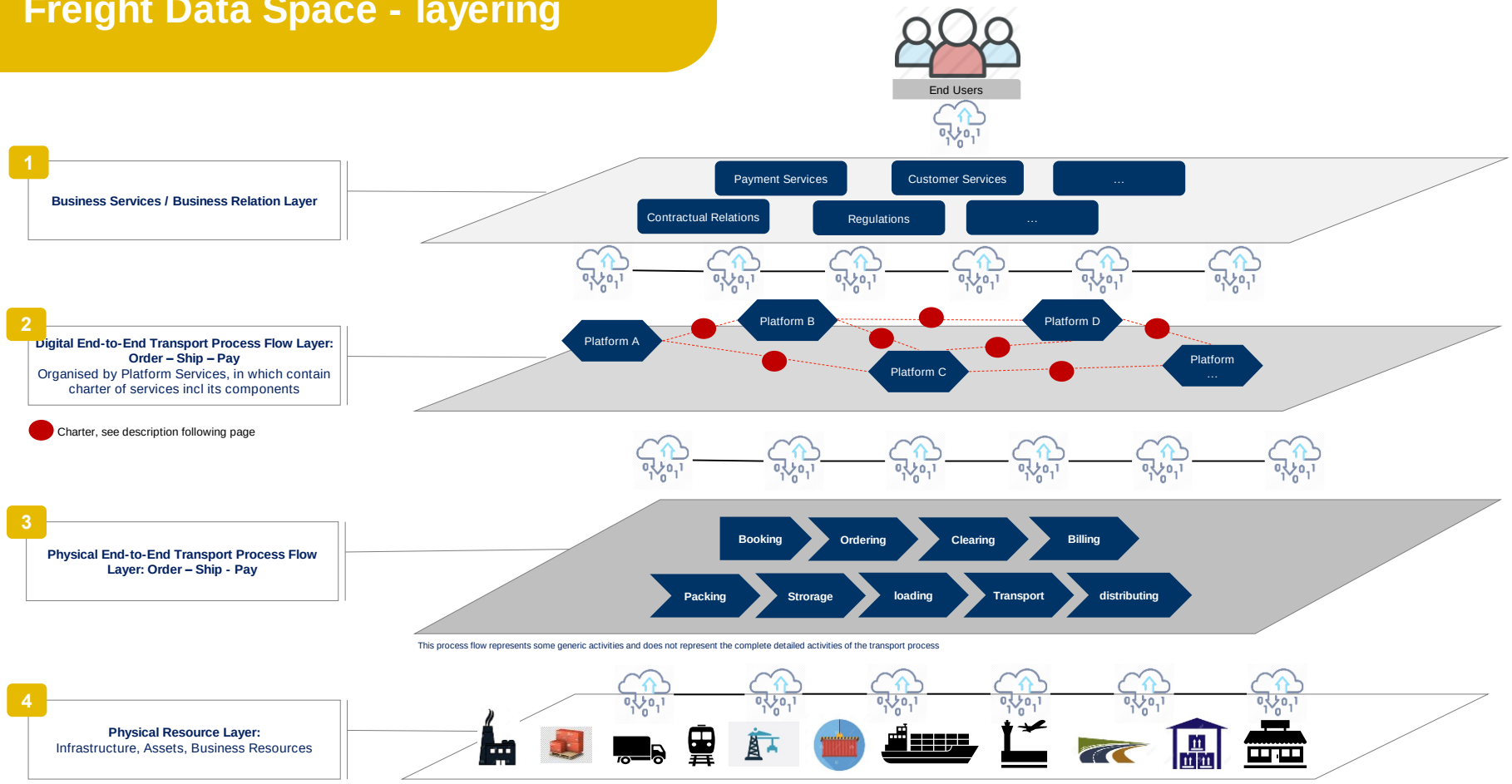


Figure 7 Layers of functionality

3.2 Roles

The roles identified in this document are based on the roles given in the Data Governance Act[1]. These roles are related to existing roles like platform service provider and business roles. Figure 7 presents an example of the various roles and their relations in the context of a specific solution, namely Visibility Platform. This same figure will be applicable for other platforms. The various roles are explained in this section.

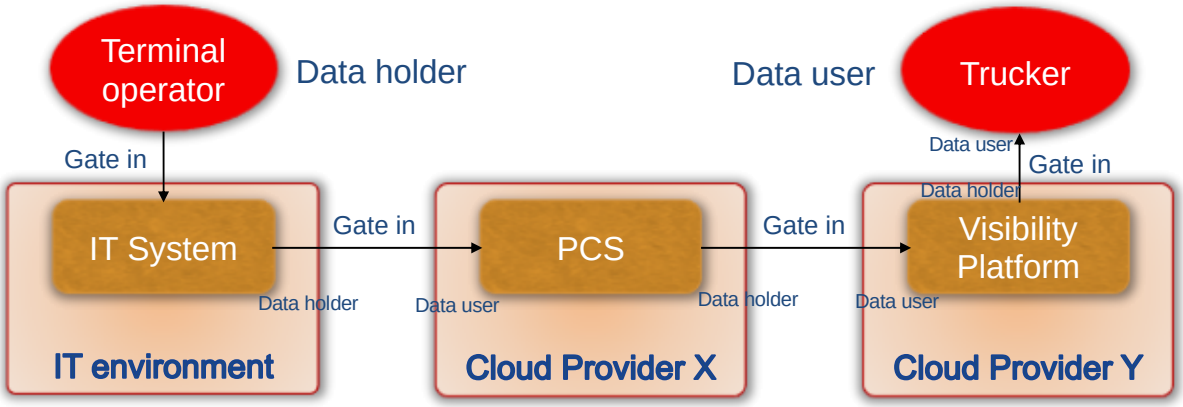


Figure 8 Illustration of data holder and data user roles (Port Community System (PCS) functionality and Visibility Platform)

3.2.1 Data Holder

A data holder is a legal person or data subject who, in accordance with applicable Union or national law, has the right to grant access to or to share certain personal or non-personal data under its control. (source: Data Governance Act). A data holder is the same as a data provider in the context of data sharing.

A data owner is authorized to create, delete, change, and share data with or provide access to data to other actors. When an organization is receiving data from another organization, the data provider(=owner) agrees with the data sharing rules the receiver operates under, and the original data provider remains owner. For example, Customs would receive data (in this case mandated and further use of the data would be regulated by law) on an import declaration and has the right to use that data for specific purposes. Customs as data receiver becomes custodian. Data provenance may be part of custodianship. If, for example, Customs transforms the data into new information ('red' or 'green' risk assessed declaration) then this a new piece of data, and that piece is fully owned by Customs. Similar examples can be given for other stakeholders.

A data steward is responsible for utilizing an organization's data governance processes to ensure that data elements are fit for purpose with regard to both the content and metadata.

A data custodian is responsible for software and hardware integrity and – availability to support a data steward. It is the computing centre used by an organization, which can be a cloud service provider or internal service to that organization.

[1] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0767

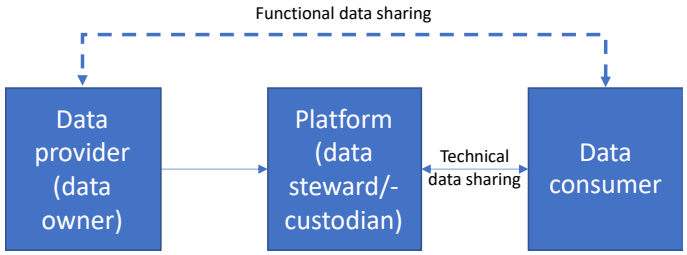


Figure 9 The different roles in perspective

A platform provider acts as a data steward, and potentially also data custodian, for multiple data consumers/data owners (generic data platform) and/or customers/service providers (supply chain and logistics platform). A platform provider may offer its (proprietary) services and make these compliant with the Technology Independent Services, or may offer (parts of) the Technology Independent Services (e.g. booking and ordering, visibility, etc.). Where its services comply with the Technology Independent Services, it can be a member of the federated network of platforms.

A data consumer – and data steward – can outsource data processing to cloud service provider(s). Figure 9 shows the chain, where a data consumer and data provider functionally share data, whilst technically the data is retrieved from a platform that functions as data steward/data custodian. A visibility platform is an example of a solution that is used by enterprises and can provide data to Customs in its role as data consumer.

3.2.2 Data user

A data user is a natural or legal person who has lawful access to certain personal or non-personal data and is authorized to use that data for commercial or non-commercial purposes (source: Data Governance Act).

A data consumer is (a synonym of) also a data user.

3.2.3 Schema owner

In this respect, the DTLF Schema - in simple terms, the architecture and the Semantic Model(s) - is defined in section 3.3.1. Two roles are identified in the context of schema ownership, namely schema manager and schema implementer. The latter may be a schema manager for services shown in the example. There are also other permutations feasible, with schema managers identified for different parts of the model and associations acting as schema managers for their members or as schema implementers.

One can have for instance a schema manager of the road infrastructure and traffic management system, another schema manager for dangerous goods classifications, a third representing data sharing requirements of a sector utilizing transport and logistics, and a fourth representing a particular modality. These schema managers can all link and re-use the DTLF Schema.

In principle, each participant or platform complying with the DTLF Schema (section 3.3.1) will have its own schema, including a schema manager. According to the DTLF Architecture, each participant or platform should support data transformation between their internal schema and the part of the DTLF Schema being accessed.

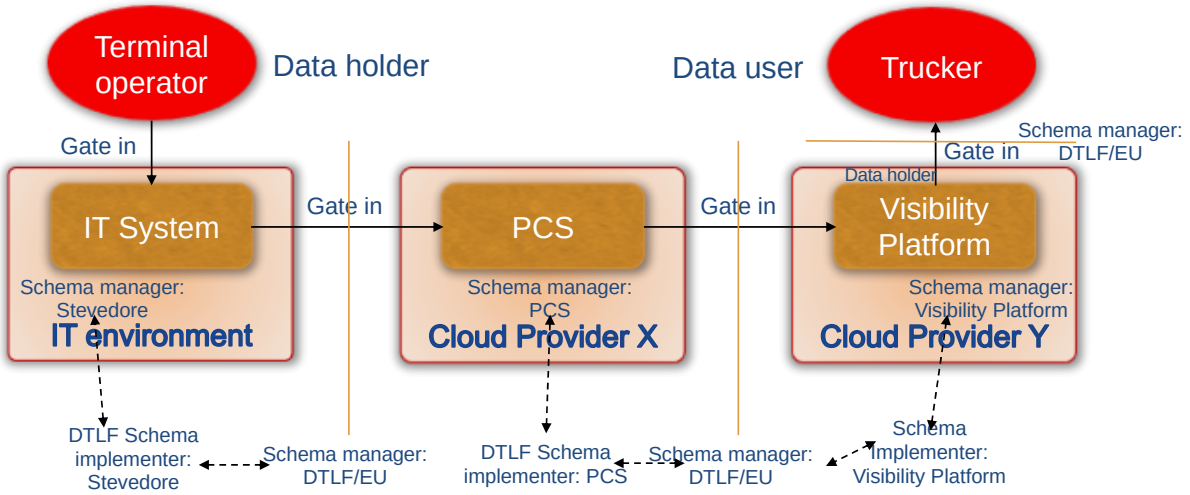


Figure 10 Example of schema manager and schema implementer roles (PCS functionality and Visibility Platform)

3.2.3.1 Schema manager and scheme implementer

A Schema Manager is the data steward role responsible for maintenance of the DTLF Schema guided by the design choices.

A Schema Implementer is the data steward role implementing (a subset of) the DTLF Schema for one or more organizations. An association might be a schema implementer, but individual organizations can also act in this role (e.g. with regard to plug and play). A schema implementer can provide feedback to the schema manager and/or extend the model.

3.3 Proposed DTLF Charter

A Charter sets out the principles and best practice for a particular purpose, in this case the DTLF. The DTLF Data Charter comprises all relevant aspects required for data sharing in the supply chain and logistics sector. The DTLF Data Charter builds upon the Data Governance Act.

3.3.1 DTLF Schema - semantic model(s)

The DTLF Schema is represented by a semantic model(s) for freight data sharing. There is one (set of comprehensive) semantic model(s) specifying the data that can be shared according to the content, with the requirement that the semantic model is able to support future data sharing requirements. The semantic model, and thus the DTLF Schema, covers:

a) All contextual data requirements (enterprises and authorities).
b) The ability of organizations (enterprises and authorities) to integrate with their internal processes and systems (plug and play; DTLF II Sub-Group 2 Team 1).
c) Data sharing between enterprises for business process integration in the context of supply chain and logistics operations (Technology Independent Services and Federation of platforms, DTLF II Sub-Group 2 Team 2).
d) All interfaces between relevant components of the DTLF architecture that refer to supply chain and logistics data sharing (DTLF II Sub-Group 2 team 3).
e) Refinement of data sharing in supply chain and logistics operations for specific scenarios (current foreseen and future ones).

The semantic model is based on an open world assumption, meaning that it will evolve over time by including any future data sharing requirements. It allows for the construction of an adaptive, evolving data sharing infrastructure. The open world assumption also implies linking to (and thus re-use of) other semantic models, e.g. models developed to represent physical infrastructures.

3.3.2 Design choices

The semantic model is based on several design choices+:

a) Digital Twin: all real-world objects have a digital representation called a digital twin. These are: products, cargo, equipment, transport means, locations/infrastructural objects, and persons (natural persons and organisations).
b) Business services and transactions: a customer/service provider relation by which the service provider performs a logistics activity based on a business service it offers to its customer.
c) Event: any association between two Digital Twins that represent a real-world state or an expected state change induced by a business transaction. For instance: container loaded on vessel, container to be discharged at a port.

3.3.3 Data quality

Data quality is defined as:

- Correctness: all data values that are shared adhere to the DTLF Schema.
- Completeness: all mandatory data values are given as specified by (a component of) the DTLF Schema.
- Consistency: data shared at different times by the same data owner must be consistent.

Data quality must be validated before submission of the data based on the DTLF Schema.

3.3.4 Data provenance

Data provenance is defined as the (chronology of the) origins, custody, and ownership of data. It requires metadata added to the data. Data provenance indicates how data is handled by the various stakeholders.

3.3.5 Data sovereignty

For any shared data, the scope of usage by a Data Consumer must be defined and agreed by a Data Owner (data owner and data consumer roles were described in the previous section). This is the generic principle that must be applied to data sharing in the supply chain and logistics sector. It can be more complex, due to the fact that data ownership and risks may be transferred if products and/or risks transfer ownership. Furthermore, complexity is introduced where data of a physical object like a container is re-used upstream within supply chain and logistics operations.

For example: where a truck is used to transport a container with its content to a port, then the truck, container, and the goods within the container are represented by a digital twin. The truck and container are owned and provided by an enterprise, as are the goods or products within the container. The ownership of the products is transferred during the process of shipping that container by a vessel from one port to another. Whenever the container is shipped by a vessel, access to the container and its contents is transferred at port of discharge according to the INCOTERMS.

Data relating to the container is shared, for instance, via a booking or transport order from a freight forwarder to a shipping line. This data set contains information about the container and details of the

requested transport services, such as place of receipt, port of loading, port of discharge and place of delivery and associated times as requested by the forwarder to the shipping line. The shipping line needs to know which entity to inform of the arrival of the container at the place of delivery or port of discharge, since responsibility is taken over by another enterprise at destination. For carrier haulage, the shipping line is the owner of the data set representing the container; the forwarder will include relevant details for transportation of the container content but remains data owner of the goods. The provenance of the goods is mostly implemented by referring to the parties involved like the shipper and consignee. This can be considered as implementation of data provenance reflecting goods provenance.

This example indicates certain aspects of complexity. In general, data sovereignty is governed by:

- Business contracts: data must be shared as to which physical objects are subject to for instance a transportation service.
- Regulation: compliance with regulation requires data to be shared and/or accessible to governing authorities, also in the context of safe and secure use of an infrastructure.
- Safety and security regarding use of infrastructure: infrastructure usage might require publishing data as open, i.e. publicly available to everyone. This might be aggregated data of the past, present, and future (predicted) state of (a part of) an infrastructure, e.g. traffic and water depth predictions.

This might result in a data classification distinguishing for instance between open data and data shared as part of a business contract. Business services, time schedules, and available capacity are examples of open data. Whether or not particular data needs to be open or not, might need to be regulated. In the event where data is open, regulations/conventions/treaties should be adhered to in order that the organization providing the data adequately addresses liability aspects. In other cases, business practices will address liability like the Hague-Visby rules for sea transport and CMR convention for road transport.

As also illustrated by the example, data ownership refers to ownership of assets represented by the concept of 'Digital Twin' and associations between these physical objects represented by 'event' (see design choices). Of course, assets can also experience a change in ownership. In most existing cases, data is duplicated from one logistics stakeholder to another and a recipient claims ownership of data. However, there are also legal and liability aspects concerning data ownership, e.g. knowing the exact content of a container makes a carrier liable for that content and makes the container prone to theft. Rotterdam – and the Hague-Visby rules[2] govern this type of data ownership; INCOTERMS are another mechanisms that govern risk and cost distribution that may impact data sharing.

The following roles can be distinguished in the context of data sovereignty:

- OEM (Original Equipment Manufacturer): any actor that produces an asset. Many OEMs intend to provide remote maintenance services.
- Asset Owner: any actor that is the legal owner of an asset such as equipment, transport means, location, or infrastructural object. Examples of asset owner are leasing companies, transport

[2] See also https://www.jus.uio.no/lm/sea.carriage.hague.visby.rules.1968/doc.html and http://www.admiraltylawguide.com/conven/sdrprotocol1979.html

companies, terminal operators, port operators, etc. An infrastructure manager is also an owner. An asset owner can also be a service provider.

- Service provider: any actor that provides a (logistics) service. A service provider can also be an asset owner.

Data sovereignty in supply chain and logistics is defined as follows:

- A service provider is the data owner of the data related to the service provided (e.g. its associations with other Digital Twins represented by events) with input data received from its customers, for instance:
    - Any details of providing a Logistics Service to a customer.
    - Any relevant information with other service providers (subcontractors) required for synchronisation of logistics services.
    - Any itinerary and condition details that an infrastructure manager requires for safe and optimal use of an infrastructure.
    - Any relevant data that needs to be provided to governing authorities for compliance.
- A service provider can act as data holder of any data provided by its customer and is allowed to re-use this data for providing its logistics service(s), e.g. by providing (access to) the data to another service provider for outsourcing.
- An asset owner can provide the use of an asset for a given period to an OEM in the context of a (remote) maintenance agreement. This can also be expressed in aggregated details provided by a service provider like total average gross weight during transport, mileage of a transport means.

There is one addition to these rules. Packaged cargo is not an asset, although the packaging material like boxes or bottles can be reused and then can become a common asset. There are cases where a customer pays for receiving packaging material and gets refunded when returning the same packaging material. The best examples are bottles or containers used by retail. Pallets, containers and tanks are assets since these can be rented by a service provider. Where there is no refunding mechanism for return, an owner will have to implement a tracking and positioning system for returnable packaging.

3.3.6 Content – context and boundaries

The content of the DTLF Schema is specified by business collaboration, regulatory requirements (compliance for safety, security, and fiscal matters), and optimization of capacity utilization. The latter is from a business – and infrastructure – perspective, resulting in for instance PortCDM (Collaborative Decision Making), path optimization, and corridor management.

3.3.7 Commercial Model of a platform

Any data sharing service provider has its own business and governance model, implementing the Technology Independent Services (TIS).

3.3.8 Cybersecurity

All stakeholders must comply with the EU Cybersecurity Act by implementing sufficient mechanisms for unauthorized access and usage of data, software, and hardware, including access to premises. With respect to data sharing independent of any legislation framework, the following common rules must be followed:

a) Authentication: the identity of a person, organization or system can be authenticated by a certified Identity Provider.
b) Data confidentiality: data is not disclosed to unauthorized users. Data confidentiality relates to data ownership and data sovereignty.
c) Data integrity: the data that is received is identical to the data that has been sent and has not been altered. Since DTLF focusses only on data sharing, data integrity during internal processing is outside scope.
d) Non-repudiation: the immutable proof of shared data in the event of a dispute. Each enterprise or authority operating in the supply and logistics domain can implement or outsource mechanisms for non-repudiation.

These requirements are supported by components of the DTLF architecture.

3.3.9 Technology Independent Services

Technology Independent Services (TIS) is the set of IT services for business process integration and compliance with respect to data sharing within the supply chain and logistics sector, supported by some form of technology. The set can be disaggregated as follows (DTLF II Sub-Group 2 Team 2):

a) Search and find: business services that can be found to match customer requirements, supported by for instance matching algorithms and simulation.
b) Booking: reservation of logistics capacity according to agreed prices and conditions.
c) Ordering: actual commitment of a customer to perform according to the outcome of booking, which might be a framework contract.
d) Visibility: real-time tracking, prediction and execution of a logistics activity as indicated by ordering supporting real-time planning and chain management.
e) Agility: the ability to cope with unforeseen changes and (predicted) delays during execution.

The relation between - and application of these elements of TIS is specified by (a) business process choreography(-ies), developed by DTLF II Sub-Group 2 Team 2.

3.3.10 Open and neutral

The Digital end-to-end transport process flow layer must be open and neutral, implying that all relevant enterprises and authorities are able to connect to this layer and share data end-to-end with all others connected to the layer. It is up to each enterprise and authority to select its platform or solution of choice to participate in the layer.

3.3.11 Level playing field

The Digital end-to-end transport process flow layer must be impartial towards the business interests of any of its users (enterprises and authorities). This has three implications.

Firstly, it implies that all service providers must be able to publish their business services and a customer retrieves all relevant business services to meet its business goal. The latter is called precision and recall: precision – only business services should be retrieved that meet a goal; recall – all relevant business services published on the digital end-to-end transport process flow layer are retrieved.

Secondly, it implies that all authorities must be able to access business data based on their authorizations for their goals as formulated by the regulations they implement (goal binding or fit for purpose). This assures compliance of business to regulations based on a pull mechanism. It requires a

data index by which authorities can find the proper data sets in data sources. In those cases where regulations have to be supported by a pull mechanism, a transformation between pull and push can be applied (see further).

Thirdly, authorities must publish their data requirements for applicable regulations (in a structured manner) to enable enterprises to provide this data upon request or according to regulation. In case the data must be available upon request, the data requirements are the basis for an access control mechanism. Enterprises must be able to be compliant with regulation for providing the proper data at the proper time.

3.3.12 DTLF Schema compliance

All data owners must implement (a relevant subset of) the DTLF Schema supporting their data sharing requirements, either by themselves or via a platform provider acting as data steward on their behalf. There are various options here, for instance:

- A customer and service provider may have a framework contract and implement ordering and visibility of the required data. In that case, ordering must have a reference to the framework contract, or to the relevant legislation that is addressed by this ordering service.
- A customer and service provider only implement that part of the schema (semantic model(s)) that is relevant for their business goals/ business services. For example, if a service provider only provides maritime container transport services, it does not implement transport services for other types of cargo and other modalities. The same is applicable for a platform provider.
- A platform provides supply chain visibility. A data owner and data user can only share visibility data based on their service provider – customer relation.
- A platform complying to the TIS may include additional functionality based on data services, e.g. data transformation and Estimated Time of Arrival (ETA) prediction. As such a platform may distinguish itself from a competitor.

The technology used for implementing the (subset of the) DTLF Schema may differ per participant. Any technology choices of one participant must be decoupled from choices made by other participants; they are able to share data according to the (subset of the) DTFL Schema. Each participant can discover the relevant subset of the DTLF Schema of other participants and share data accordingly.

3.3.13 Compliance with the Data Governance Act

The shared data must be compliant with the data protection legislation framework where the data owner has granted the access to his data.

3.4 Next steps

Whereas the roles already provide an indication of a governance structure, still the various rights and objects for governance must be addressed. Potentially, there are several objects for governance, namely an architecture; its components, interfaces, and implementation of the architecture. This would look like the structure of the Internet, where a distinction between standards and implementation of these standards is made.

Rights can be on different levels:

- Constitutional rights: who may or may not participate in making collective choices.

- Collective choice rights: rights concerning users and components within the information system.
- Operational rights: rights related to access to the information system and to reading and adding data.

Further consideration of these aspects will be required.

4 Business and authority collaboration

This section addresses the various aspects of data sharing within the context of business collaboration, including sharing data between businesses and authorities, and authorities and businesses. These formulate the functional requirements for data sharing in terms of semantics and architecture. Semantics and architecture will be elaborated in next sections.

4.1 Business collaboration

Business collaboration can be addressed at two levels, namely a (supply or logistics) chain of more than two collaborating enterprises and any two collaborating enterprises in multiple chains. The level of individual chains will not be addressed, as each enterprise will have its own outsourcing policies and strategies related to its business model. An example is that a logistics service provider may decide to bundle goods flows of two or more customers or a shipper does not want to bundle his goods flow with that of his competitor(s). Those policies and strategies are considered outside the scope of DTLF. There are two constraints in this respect:

- Leg synchronization: any intermediary should synchronize different transport legs and transhipment of cargo from one transport means to another.
- Visibility: any intermediary should inform its customers of the progress of the shipment of their cargo, if required by a customer at individual leg level.

Business collaboration between any two enterprises relates to a quotation/contract view for service delivery, for instance of transport, between a customer and a service provider. This will be described first. Secondly, the implementation of this contract view will be considered and thirdly, an example of (visibility) milestones is given.

4.1.1 Contract view

The contract view is further disaggregated into various steps, each step providing input to the next step, all having their own data requirements and providing their data to other parties in the chain (both upstream and downstream). Figure 10 sets out the steps; they will be further disaggregated into a business process choreography (see also DTLF I Sub-Group 2 final report).

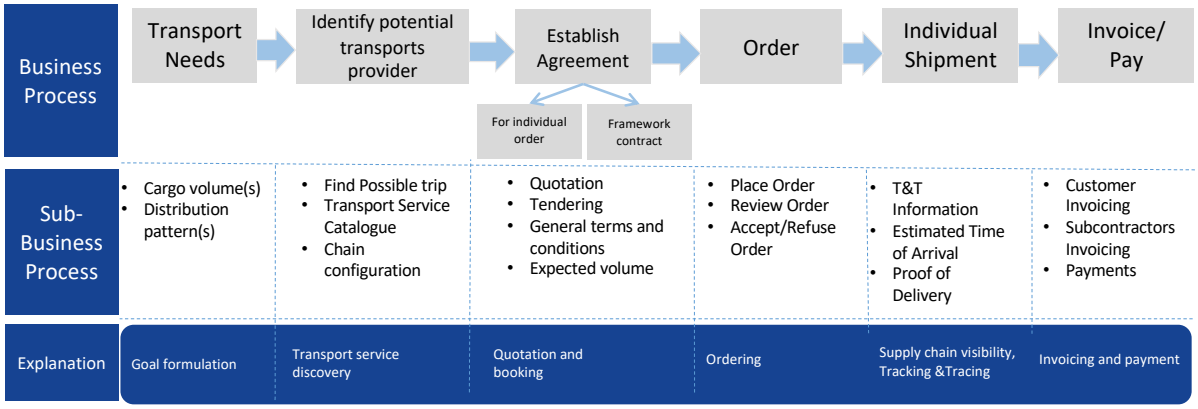


Figure 11 Transport contract view

These process steps can be refined by adding sub-processes such as assigning equipment to an order, bundling orders to trips and assigning a truck-driver, planning of timeslots, weighing, traffic and

corridor management, etc. Each of these sub-processes is either internal to an entity or can be provided as an external business service via subcontracting.

Service requirements, as in transport service requirements, are formulated in terms of cargo volume(s) and their type to be shipped between different locations (distribution pattern(s)) in time periods). These patterns are the basis for finding service providers and a trip(s), or even to compose a chain for a particular distribution pattern. All based on a transport service catalogue, potentially supported by logistics marketplaces.

Next step is to come to an agreement. Quotation, booking, and ordering processes take place. Whenever transport orders are agreed upon, also based on planning information provided by a service provider to its customer, supply chain visibility is required on shipments and cargoes. Track and trace information and estimated/actual time of arrival at the intended destination are required. Cargo will be formally accepted by a service provider according to the agreed terms and conditions. Proof of Delivery will provide details on the state of the cargo (e.g. damage remarks) and is the basis for invoicing and payment.

4.1.2 Implementation variants of a contract

The previous contract view represents the case where a customer and service provider come to an agreement and execute per consignment or shipment. There are variants of this process that can be applied in practice for different types of cargo and modalities, like:

- Framework contracts - in the event of quotation and tendering, a framework contract for a longer time period would most likely be concluded based on expected volumes. Within this framework contract, orders will be placed for individual shipments of cargo. General terms, prices, and conditions are agreed upon within the framework contract;
- Non-negotiable prices – there are situations where prices and conditions are standard and non-negotiable, like for parcel delivery & like for eCommerce. However, an eCommerce provider with high-volume of shipments will also have a framework contract with a parcel delivery service provider.
- Payment before delivery – small shipments based on individual orders may have to be paid for before they are transported. It can also be the case that goods are only handed over to the next stakeholder in a chain, after payment (like for instance according to the Incoterms ‘free on board’ (FOB)).
- Change of ownership – cargo can change ownership during transport (negotiable B/L). This might influence payment and financing of logistics operations.

Any combinations of the contract processes, including regulatory and financial aspects, will be considered.

4.1.3 Examples of visibility milestones

Especially in multi-modal transport, there are various points at which cargo is handed over from one carrier to another. Each carrier will report on its progress, having its own milestones to report to for instance an intermediary that handles those for one or more customers (see Figure 12).

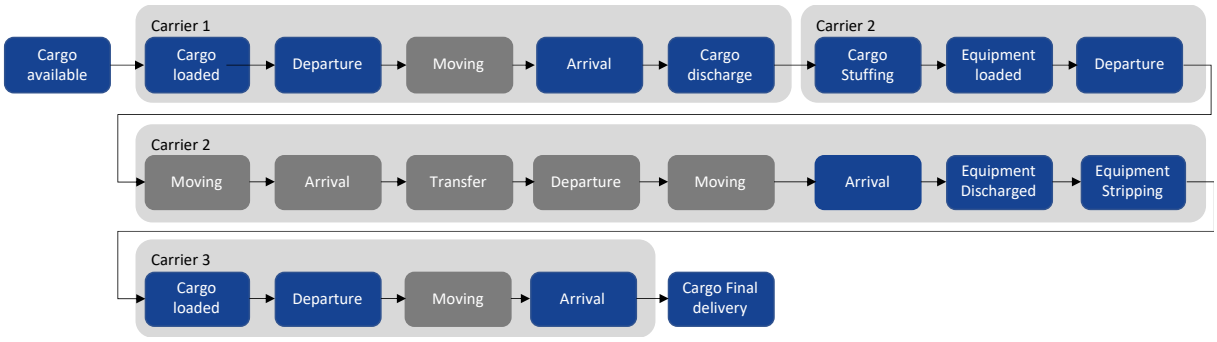


Figure 12 Milestones in multimodal transport

Figure 12 indicates the milestones that are relevant to multiple stakeholders, such as 'cargo loaded' (blue) and internal milestones (grey). It visualizes three modalities with different carriers involved, where the second carrier uses equipment such as containers for transportation and transfer between different means of transport. The transfer between different transport means of the same modality may also be relevant to the customer that offered the cargo.

4.2 Regulations

Enterprises require what can be called 'a seamless goods flow', compliant with regulations. The latter means that they will have to provide data to Member State authorities which implement and monitor the compliance with these regulations. There are two aspects relevant in this context, namely data sharing between an enterprise and an authority and the European (and/or global) context.

There are basically two ways of sharing data between business and authorities, namely:

- **Push:** data is submitted by business to authorities. Customs declarations are an example of a push. Authorities will define procedures of providing regular updates by business (if required) and timing, i.e. the moment at which particular data needs to be 'pushed' to the authority domain.
- **Pull**: data is available for access by authorities. In this case, authorities need to know that data is accessible and where to access the data. It requires identification and authentication of authorities by enterprises to prevent any unauthorized access, since data is commercially sensitive.

The implementation of these mechanisms depends on choices made in regulation, which may relate to any risk assessment procedures implemented by governing authorities.

The second aspect that needs consideration is that of the European context. Cargo and transport movements are executed within the European Union (EU): within a Member State or between Member States (*intra*-Community *transport)*, incoming to the EU, crossing a border from a non-EU country to a Member State, and exiting the EU crossing a Member State border into a non-EU country. These movements are applicable to any transport modality. Different regulations such as the European Single Maritime Window environment (EMSWe), eFTI, and Union Customs Code (UCC) cover cross-border data sharing, where a choice of push or pull is made.

4.3 Towards plug and play

In the context of business collaborations and regulations, organizations need to implement data requirements. Ideally, logistics goals and – services and compliance to regulations specify data requirements of various stakeholders. A stepwise approach can be taken by these stakeholders to gradually implement plug and play. The following steps are identified, they can be taken separately or in combination:

- Document data sets – this is about digitization of existing business documents like electronic CMR, - Bill of Lading (eB/L), and – Airway Bill (eAWB). Data can be made available according the architecture (section 6.2).
- Visibility data (events) – these represent the past, present, and foreseen progress of transport orders. They may refer to document data sets. Examples are already shown before.
- Compliance to regulations – data will have to be available to an authority, based on a (national implementation of a) EU Regulation. Please note that compliance can be based on digitization of documents.
- Logistics services – this includes digitization of all relevant individual process steps in the so-called transport contract view.

It is up to each stakeholder to select its way forward. However, one of the aspects of plug and play is data sharing with any other stakeholder, without the requirement of prior agreements. It implies that someone implementing logistics goals or – services can only share document data sets with those that have implemented at least the capability to share document data sets. These types of rules will be detailed.

5 Modelling

This section presents design concepts that are the basis for the DTLF conceptual model and details on its content and structure. It should be noted that some of these concepts reflect the evolution of approaches to data exchange and these will evolve. Not every stakeholder will be familiar with these and some effort will need to be made to ensure that stakeholders within the transport and logistics industry have the opportunity to evolve and adapt their systems over time.

The DTLF semantic model presented in this section will constantly evolve, based on all requirements of various use cases that deploy the model with a particular technology (see section 6). Governance needs to assure how to cater with this evolution. Governance will be addressed in the final report of DTLF SG2.

5.1 Design concept

The basic design concepts are 'Digital Twin', 'Event', and 'Business & Compliance', as shown in Figure 13 (Digital representation). These will be described in more detail in this section. In summary, a Digital Twin is the data representation of any real-world object, for instance containers, trucks, airplanes, vessels, infrastructure, locations, etc. A business transaction represents the commercial relation between any two enterprises in a chain based on business services; An 'Event' represents the relations between Digital Twins in the context of a business transaction[3]. The latter is not a definition of 'Event' as can be found in literature but is proposed to serve as a basis for sharing any information of the (planned/expected/actual) progress of physical activities like transport and transhipment; also in the context of administrative actions like Customs declarations.

Figure 13 presents this design approach for the semantic model. It separates the transport contract view (organization network) from the digital representation of physical world objects, where the transport contract view will refer to this digital representation. The organization network shows that any two stakeholders sharing data always have a 'context' of a goal and a business service, supported by a 'choreography'.

[3] See also glossary.

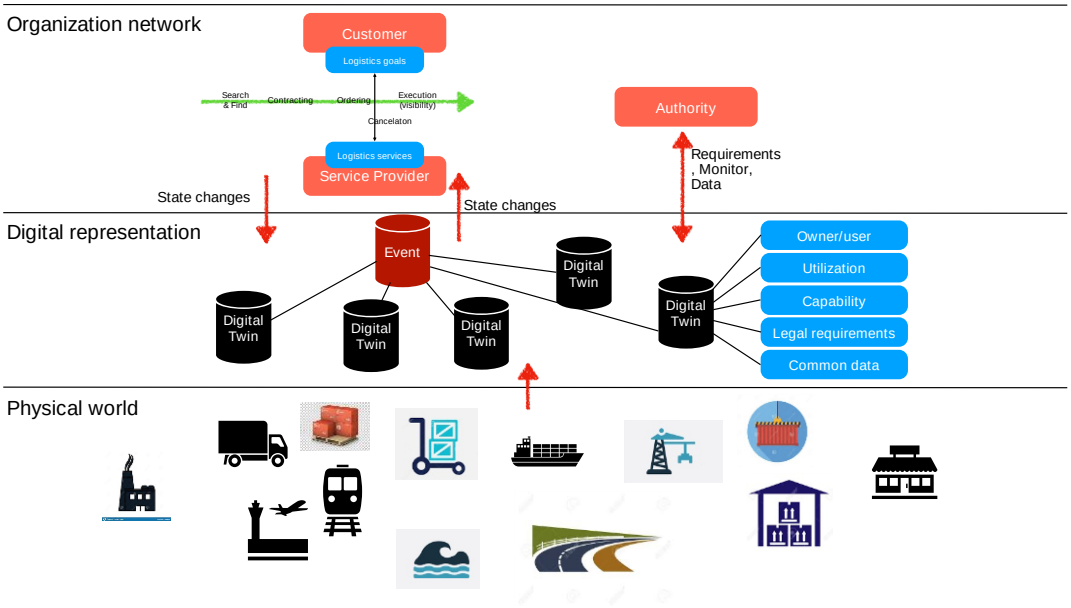


Figure 13 Layering design choices

5.2 Overview of the Conceptual Model

Based on these basic design choices, a top-level Reference Model has been constructed. Figure 14 shows the main concepts.

NOTE:

The Conceptual Model will be refined by the pilots and Living Labs of the DTLF consortia FENIX and FEDeRATED. In line with the guiding principles outlined in section 3.3, the final Conceptual Model will align with existing well-established semantic models, ontologies and data models such as those offered by UN/CEFACT, GS1, UBL and others.

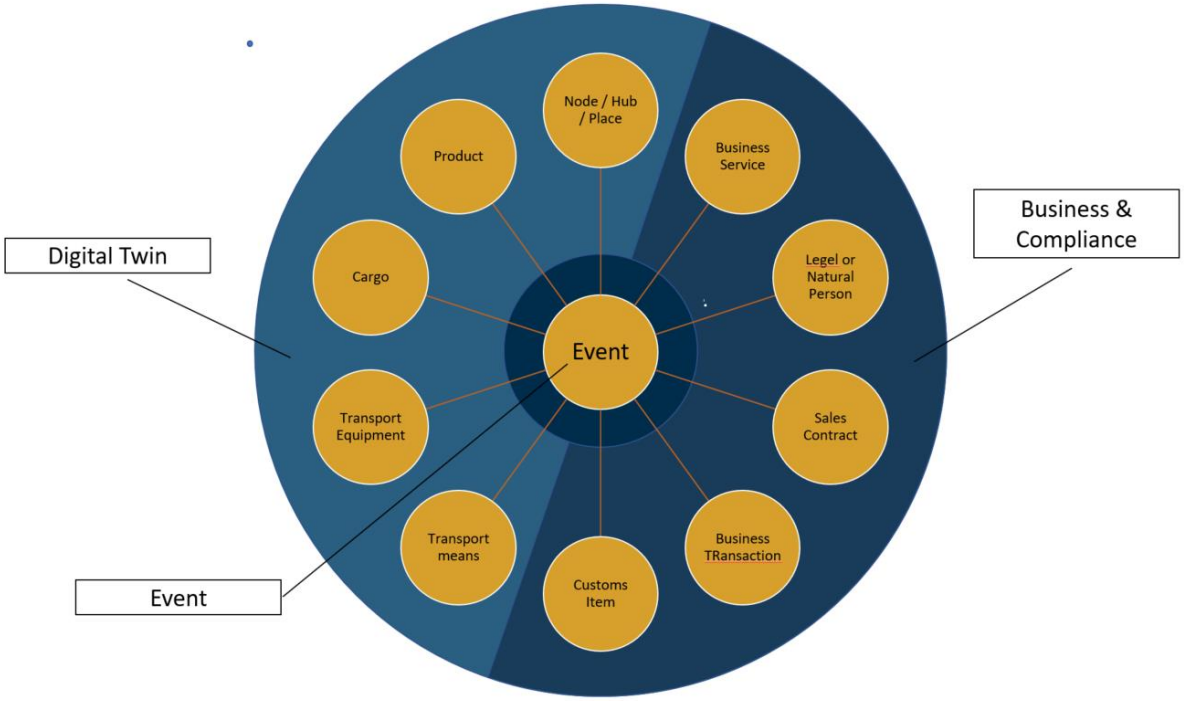


Figure 14 The main outline of the Conceptual Model
(source: adapted from FEDeRATED Semantic Modelling Group)

The transport concepts presented in Figure 14 will be further refined in cooperation with the DTLF consortiums FENIX and FEDeRATED and input from other parties as necessary. A short description of concepts and key elements follows:

- Event: the concept of an 'event' is at the core of the model and represents actions, milestones, transactions, or any other real-world activity that has occurred or is intended to occur at a specific moment in time. These will typically involve one or more interactions or associations between location, business service, transport means, transport equipment and cargo. For instance, a vessel is expected to arrive at a given time in a port. Similarly, for a container, which will be loaded on and discharged from a vessel in a port.
- Area of interest (Node/Hub/Place): this concept represents physical infrastructure and may refer to a terminal, location, port, etc., where physical activities related to cargo, like transhipment or storage can take place. It can also be a border crossing, infrastructures including locks or bridges, or a city centre with certain access restrictions. A node will play a role (often reflected in the name used amongst stakeholders) within the context of a particular transport activity. Some examples are Port of Call for a vessel, a Port of Loading to indicate where a container is (to be) loaded onto a vessel, or a Place of Acceptance, i.e. where responsibility for the freight is transferred.[4]
- Business services: a business and compliance concept, this represents the commercial relations between any two enterprises in a chain based on business services. A business service is the basis for several related business transactions, for instance a shipper will have several

[4] The place of acceptance is known as the place where a carrier or forwarder takes over responsibility of the cargo from a shipper.

business transactions with a forwarder over time. Data that is shared may form the basis of documents that are required for legal reasons. Providing data regarding traffic volumes or corridor management is also considered as a business service, be it a business service provided by an infrastructure manager.

- Legal or Physical Person: a business and compliance concept, this represents a business entity or even any individual e.g. a crew on board a transport means (e.g. ship, train, etc). Physical persons have roles, which also relate to qualifications. Hence, a person can have a role, of e.g. a truck driver or captain.
- Sales contract: a business and compliance concept, this represents a sales contract, contract of sale, sales order, or contract for sale: the legal contract for the purchase of assets by a buyer from a seller in which the seller relinquishes ownership of the assets to the buyer in exchange for some monetary compensation. Sales contracts also specify the terms of sale, which may take any of several forms.
- Business transaction: a business and compliance concept, this represents a single uniquely identifiable arrangement within the context of business services for the execution of services agreed. In this paper dealing only with transport and logistics, the sales contract between the seller and buyer of the goods is not a business transaction.
- Customs item: a business and compliance concept, this represents the classification of products or cargo for customs purposes according to the Harmonised System codes (HS). One can basically have three classifications: export, import, and incoming/transit. Export and import relate to products and incoming/transit to cargo.
- Transport means: in the group of Digital Twins concepts, this represents the vehicles that transport the cargo, such as trucks, vessels, trains, airplanes, barges etc. Each transport means has a specific transport mode; some might have more than one mode.
- Transport Equipment: in the group of Digital Twins concepts, this represents any asset used to facilitate transport and handling of cargo and are not able to move on their own (e.g. trailer, container, push barge, rail wagon, …).
- Products: in the group of Digital Twins concepts, this represents the actual objects that change ownership between a seller and a buyer. When products are transported, they are referred to as cargo by the logistics and transport partners. For transportation purposes, products may be packaged (containerised cargo) in transport units that may be identified uniquely, or they may be transported directly in transport means (bulk cargo).
- Cargo: in the group of Digital Twins concepts, this represents the goods that are transported from origin to destination. Cargo may be bulk or containerised. Cargo is defined by generic descriptions of what the product is, for example fruit, textiles, etc.. Cargo can be packed, repacked, consolidated, reconsolidated etc. Transport means (truck) and transport equipment (trailer) can also be cargo itself, for example on a vessel or ferry (roll-on/roll-off).

The presented model as explained above, represents a very high-level grouping of the top-level concepts. Therefore, depending on the business case, the transport model may be refined further on the various main concepts. These refinements may imply additional business constraints. For example, containers can be transported via the maritime mode only by vessels equipped for container transport.

5.3 Using the Conceptual Model

The Conceptual Model may be viewed from different perspectives, based on evaluating an event association between collaborating roles (e.g. customer and service provider) and the concepts related. Examples are:

- Consignment data set: any data set shared between a customer and service provider providing details of cargo to be transported from one location to another at the same time. This is essentially the transport order.
- Document data sets: data that is normally contained by a particular document relevant to a consignment, e.g. a business document like a CMR, cleaning document, or a document related to the shipment for example a Certificate of Origin issued by an authority for goods included in the Sales Contract. This data set may refer to other data sets like cargo and transport means.
- Itinerary data set: a data set combining operations on cargo and a transport means at locations. An itinerary refers to cargo data, transport means, nodes, transport equipment, and transport means operator (driver) data; It may have a unique identification stored by the event connecting transport means and locations.
- Route data set: any physical route of a transport means during an itinerary. A route relates to a physical infrastructure, for instance by means of physical coordinates or identification of a road.
- Reporting data sets: reporting data sets like FAL messages (Facilitation of International Maritime Traffic[5]) and future reporting data sets like eFTI and eMSW are shared as a set of references to one or more of the other data sets if available, e.g. a reference to the cargo loaded at a node and (to be) discharged at another node and the crew of a means of transport.
- Sales contract data set: covering the subset of sales related data that is of interest for transportation and cross-border procedures e.g. total value of the sale drives (potentially including transport, insurance, and other relevant costs), customs procedures in many countries of the world.[6]

One can also consider including nodes or hubs as specific data sets, where enterprises operating these nodes provide services, e.g. a stevedore enterprise providing transhipment services at a terminal. Other types of nodes might contain storage and service facilities (e.g. warehouse or distribution centre, tanker cleaning station, container depot).[7]

It is important to note the data models deal with semantics that should be independent of any specific kind of technology. Semantics expressed as ontologies (owl turtle, rdf, json-LD) may be implemented in different syntaxes (e.g. XML, JSON, or even CSV) and transferred among partners using a wide range of protocols/mechanisms and different forms of synchronous or asynchronous connectivity, (e.g.

[5] https://www.imo.org/en/OurWork/Facilitation/Pages/FALCommitee-default.aspx

[6] E.g. new EU VAT Ecommerce regulation coming into effect 1st July 2021. All imports of sales value below EUR 150 are subject to the new regulation, which makes electronic submission of data to Customs mandatory.

[7] GS1 is currently running a programme to define exactly these types of data set for the physical location aspect as well as for the legal entity and services aspect.

RESTful API, FTP, EDI-networks, Access Points like in PEPPOL, etc.). The FENIX and FEDeRATED consortia will provide solutions in the areas of syntaxes and transfer mechanisms.

The various data sets mentioned above can be a basis for developing message structures and Application Programming Interfaces (APIs): these will be elaborated as part of specification of the Technology Independent Services at a later stage.

Some examples are:

- An API for booking of a transport means that can be specified based on location, cargo and the transport means required.
- An API query endpoint (e.g REST, GraphQL, SPARQL) that supports the route and the cargo carried by a transport means.
- An eFTI API will be able to retrieve data that is required by authorities. This excludes data concerning business-to-business information. These are limited subsets from transport documents like waybills such as CMR.
- An eMSW API will comprise the means of transport and its relevant port call, information regarding the cargo on-board and cargo to be loaded/discharged in the port, as well as an overview of passengers and crew. Where the cargo is containerized, and the port is the first port of call in the EU the eMSW API may also provide the content of containers by their TARIC code. In the latter case, one container will have one TARIC code, meaning that there will be one Customs item for each container; a container can carry multiple goods with different TARIC codes.

These APIs can be predefined at different levels and deployed by various platforms. For instance, each stakeholder requiring access to data can define and publish its query for eFTI and eMSW queries that are being formulated at EU level.

Again, we need to stress that the definition of content and structure of these APIs should be independent of transfer mechanisms that will be implemented as part of the eFTI physical infrastructure.

5.4 Events in the reference model

'Events' is an important concept. Collectively, events form the lifetime history of interactions and relationships. For example, the relationship between two Digital Twins concepts, a container and a vessel is created after loading the container on a vessel and ends after its discharge. These interactions and relationships are activities with start and end milestones. Events represent a logical relationship that is more comprehensive than the usual meaning of the word 'event'. The concept should also not be confused by the term 'milestone'.

High-level milestones are:

Generic milestone	Description
Start	Creation of a relationship between any two transport concepts as shown in Figure 12
End	Termination of a relationship between these transport concepts

These milestones are combined with date/time, representing the lifetime of the specific association between two Digital Twins concepts. A milestone linked to the specific lifetime of an association may have **one** of the following properties:

- Estimated: the time provided by a specific stakeholder (e.g. customer, service provider) related to a specific event e.g. ETA.
- Planned: the time at which a service provider plans to execute a task or service. For instance, when using a voyage scheme (shipping schedule), it is the time of call of a vessel in the Port of Loading according to the schedule.
- Requested: the time at which one stakeholder can perform a logistics activity and requests another stakeholder desiring that activity to be available at the requested time. E.g. a port terminal may request a vessel wanting to be unloaded to arrive at the terminal on a specific date and at a specific time.
- Actual: the time at which a relevant milestone took place, e.g. actual loading of an intermodal container on an inland waterway barge. This is important for keeping track of where transport equipment is and by implication where consignments and shipments are.

The first three milestone types are all in the future (future milestones). Other types of milestones may be used to specify the earliest and latest date and time (e.g. for pick-up or delivery). Future milestones may be updated at any time driven by business requirements and changing situations in the execution of transportation.

'Actual' milestones are exchanged only after the event that they relate to has occurred. The Actual milestone exchange may be done in near real-time or with some delay. Clearly, 'actual' milestones should not be changed once recorded (unless they were recorded in error; this refers to data integrity, see section 7).

Each (combination of) milestone(s), its lifecycle, and association(s) can be used to express user-specific events such as 'container loaded' or 'gate-in'. The latter event can be expressed by the value 'start' of a milestone between a truck and a terminal at a location. Other examples of user-specific events are 'accept', 'load', 'depart', 'arrival', 'unload', 'deliver', 'pick-up', and 'drop-off'. The list is quite extensive.[8]

5.5 Towards the DTLF Semantic Model

The model is based on an open world assumption: the first version will not include all functionality and is developed to futureproof the incorporation of any (future) required additional functionality. The model will be modality independent, thus able to provide cross-modal interoperability. To achieve this, design concepts must be considered (see previous section). All underlying details of how the model is composed, including details of the application of this model, require elaboration.

The following aspects are relevant for constructing the model.

[8] The ISO/IEC 19987 & 19988 - EPCIS and Core Business Vocabulary (maintained by GS1) provide for a large number of "standard" events. The Port Information Manual provides further standard events. GS1 has started a mission specific workgroup to standards even more milestones to achieve "Integrated Track & Trace for Multi Modal Transportation".

5.5.1 Standards perspective

The DTLF semantic model must be able to integrate various modalities, cargo types, and hubs. Therefore, it is based on existing standards, obligatory from a regulatory perspective or optional. In this context, DTLF suggest three perspectives towards standards, namely:

- To construct the basis for the Semantic Data model. Relevant standards of each modality are used for this purpose, such as the UN CEFACT Multimodal transport Reference Data Model and the Dutch standard Open Trip Model (OTM – version 5) for road, IATA ONE Record for air, IMO, International Taskforce Port Call Optimization (ITPCO) and PortCDM (Port Collaborative Decision Making) standards for sea, TAF/TSI of rail, and RIS standards for inland navigation. Customs standards for incoming cargo (Entry Summary Declaration and CUSCAR/SAL) and import declarations will also be used to develop the model (considering the semantics of what might seem identical data properties). Additionally, available interfaces with systems and platforms are used like Tradelens, Container Status Message, and shipping instruction of a platform. The model is also to align with the latest developments such as the eFTI data model and eMSW .
- To identify 'business constraints' that are standardized and adopted by global trade data sharing. These are amongst others relevant UNECE Recommendations (like for packaging and transport modes), ISO standards (e.g. country, and container size and type), the Harmonised System Code (HS-code, of which the first six digits are harmonized globally), TAF/TSI specific codes, commercial (defacto) standards, and codes for dangerous cargo. The list of relevant standards of constraints will be extended when the FEDeRATED semantic model will be applied on a broader scale.
- To use the semantic model – as per its alignment to existing trade and logistics standards - as a basis for the implementation of various data sharing interfaces. The semantic model will specify all types of data sharing interfaces like APIs and express all standards. Applying the model will lead to potential extensions of that model, including the format constraints.

5.5.2 Separation of supply and logistics ontology concepts and formal constraints

The semantic model specifies all semantic details, classifications (e.g. container size and type code sets and UN Locodes), and relevant constraints (e.g. format constraints) for multimodal supply and logistics chains. These are all represented as separate modules, thereby improving the maintenance, applicability, and extendibility of the model.

5.5.3 Expressing common logistics concepts

Various logistics concepts such as Trip, Voyage, Route, Logistics chain, etc. can be expressed in the context of the logistics concepts of the Digital Twins (see Figure 14). These constructs are considered as 'standard' view on the model and will be included in the semantic model (the Knowledge Graph).

5.5.4 Context-specific terminology

This is where the approach addresses three roles: customer, service provider, and authority: This includes roles such as shipper/consignor, `consignee, forwarder, carrier, Customs, etc. These roles will all be expressed in terms of the semantic model and included in the model (the Knowledge Graph). The same is applicable to events that can be mode and/or platform specific. These context-specific events are expressed in standard milestones of events in the model.

5.6 Knowledge Graph and Shape Graph

The semantic model needs to address two aspects, namely the logistics concepts of the Reference Framework and its constraints, such as cardinality, format, and values from restricted value sets (for example code lists). Although these are already available, they need to be integrated in the semantic model for completeness and to support all types of data transformations and alignments.

The objective of the semantic model is to support a wide variety of use cases (see also section 1.3): the semantic model should be shared and thus be represented by open standards. It should for instance support regulators in rapid development of consistent and complete specifications of data requirements in the context of a regulation. It should also support supply chain stakeholders in digitalization of their supply chains and user groups in replacing business documents with exchange of required data electronically, all resulting in consistent and complete specifications that are fully aligned with data requirements of regulators.

To achieve this objective, separation of concern is applied by constructing the semantic model. Where many stakeholders are familiar with logistics concepts, experts must be consulted to include all details like code lists and particular format constraints. Thus, the DTLF distinguishes:

- The **DTLF Knowledge Graph**: a representation of the Reference Model (Digital Twins, business transactions between organisations, and their associations modelled via 'event') with their semantics and properties of these concepts, represented as an ontology.
  A knowledge graph may be seen as a network of things and their relationships, hence the term 'Graph'. For sharing, it can be technically represented in RDF (Resource Description Framework) or any of the popular syntaxes such as Turtle file format (Turtle – Terse RDF Triple Language), JSON-LD, XML, OWL etc. The Knowledge Graph should be usable by business experts to formulate their data requirements, supported by easy-to-use tools.
- The **DTLF Shapes Graph:** any format constraints specified by open and de-facto standards, technically represented as so-called shapes (SHACL – Shape Constraint Language, W3C standard recommendation). Examples are units of measure (including transformations between different units), currency codes, country codes, packaging types like bags and boxes, and UN LOCODE for locations.

The Shape Graph automatically adds all details to the data requirements for sharing data between information systems of different stakeholders. Where multiple options are feasible, e.g. the ISO country code list distinguishes a numeric and an alpha representation, a selection should be made. Such a selection is specified for a particular use case and part of a shape graph of that use case.

The Reference model can be quite abstract to business experts. They are used to express data requirements in logistics concepts and events that they are familiar with. Therefore, three aspects need to be expressed by the Knowledge Graph, namely **roles**, **logistics concepts** that can be derived from the model and **user-specific events**.

Both the Knowledge Graph and Shape Graph can be detailed for a given industry by the relevant industry body, as will be addressed by governance. This is for further work.

5.7 Modelling challenges

Where the Reference Model is based on the concept of Digital Twin representing business transactions between a customer and a service provider, frequently in practice, some other terminology may need to be modelled and applied. Roles and generic concepts are amongst those concepts where such variations may occur. The challenge is to match and to express these concepts in terms of what should be called the core Reference Model. These challenges will be discussed in the following.

Many open and de-facto standards contain **roles** of organizations and locations, such as shipper, consignee, notify, place of acceptance and port of loading. These roles are contextually defined, where the context is given by the structure of a logistics chain with its business transactions. Adding this context information to the Reference Model will add complexity to the model and limit its applicability to business scenarios. It will restrict the support of any not yet foreseen application such as truck platooning or automated transport. Adding these future applications requires changes to the Reference Framework and potentially expanding the semantic model. Thus, to keep the semantic model, DTLF defines *roles in existing logistics concepts* of the Semantic Model, as part of the DTLF Knowledge Graph.

Secondly, the semantic model needs to cover **generally applied logistics concepts** that can be defined in terms of existing ones. In terms of data modelling, they can be considered as 'queries' on the model[9]. Route, voyage, trip, flight, container track, train composition, and logistics chain are examples of such concepts.

A trip and voyage, which are similar and can be formulated as **'itinerary'**, are for instance expressed as the 'sequence in time of places of call of a transport means (truck for trip and vessel for voyage) where cargo is loaded and/or discharged.' In this example, the *event association* between a transport means and a location and that between a cargo and a location need to be sharing the same location to be valid, for an operation to load that cargo. Furthermore, there has to be a business transaction (shipment order) with the carrier expressing that the cargo needs to be loaded at that location.

Thus, although formulating these logistics concepts in terms of Digital Twins, business transactions, and events can be complex, they can be expressed by the model. Formulating newly derived concepts requires knowledge of the Reference Framework, and this can potentially be done by the modelling experts.

The last that will have to be expressed, are user-specific events, such as a 'container loaded on vessel' or a 'vessel departed'. These user events can be expressed in the event associations of the model. For instance, 'container loaded on vessel' creates the event association between a container and a vessel and indicates that the event association between this container and location has ended (the container is not at the location, but on the vessel). After receiving 'vessel departed' the semantic model can infer that all containers loaded on this specific vessel have departed (been disassociated) from that location.

Another user-specific event is 'cargo pick-up at location'. This event can be formulated as a combined event, where the cargo is loaded on for instance a truck at a given location and the truck is departed. Thus, similar events can be expressed on the model.

[9] For instance https://sparxsystems.com/enterprise_architect_user_guide/14.0/model_domains/dbsqlquery.html

Figure 15 shows the proposal to express these aspects in the DTLF Knowledge – and Shape Graph.

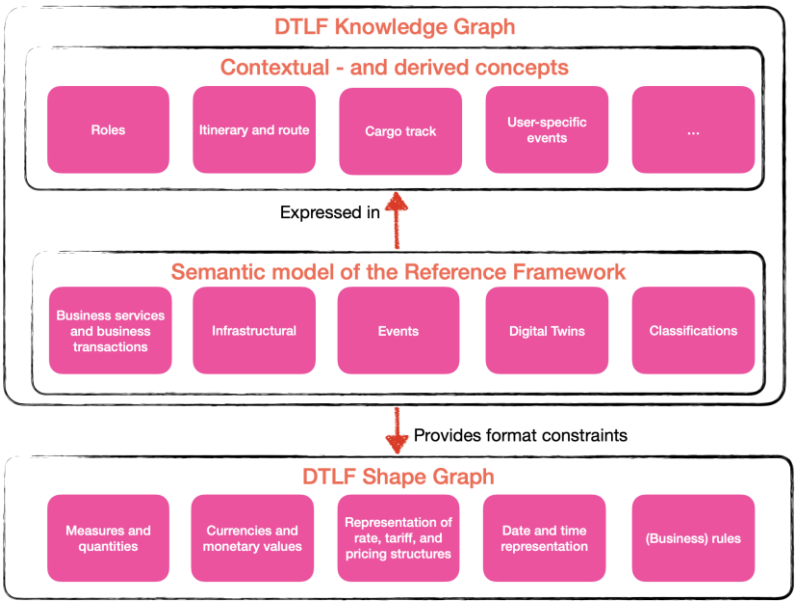


Figure 15 FEDeRATED Knowledge – and Shape Graph

Thus, the **DTLF Knowledge Graph** contains (potentially to be expanded):

- The DTLF Reference Model represented as a semantic model (ontology). It consists of:
    - Digital Twins: all identified Digital Twins and their data properties are defined by the Knowledge Graph.
    - Event: event constructs the associations between any two Digital Twins and has the properties identified for the Reference Framework (see before). An association may have additional properties such as the role of a person in its association to a transport means, e.g. a truck driver.
    - Business services and business transactions model data shared in a commercial relation between a customer and service provider.
    - Infrastructural aspects like locations, infrastructures, addresses, geo-references, areas, etc.
    - Classifications: all types of code lists that are applicable, e.g. ISO country codes, HS classifications, Dangerous goods classifications, and UN codes for packaging types.
- Any roles of a Digital Twin are expressed in their context of business transactions in supply and logistics chains. This especially concerns roles of organisations and locations.
- Any combinations of Digital Twins and events, structured for instance according to a time sequence, are expressed separately. In traditional technology, these are queries of the model that can be standardized. The most common used are for instance (the list can be extended):
    - Itinerary: a timed sequence of transhipment locations or hubs passed by a transport means to load and/or discharge cargo. An itinerary can have a scope in time (e.g. an itinerary of a truck can be defined by the time that passes between when it has left its main location, will return there, and the driver will return home), geography (e.g. an itinerary of a vessel, called 'voyage', is defined by its direction: Far East – Europe is one itinerary and Europe – Far East is another), and frequencies (e.g. an itinerary of a plane, called 'flight' can be daily

at a particular time, except on Saturday, Sunday, and bank holidays). An itinerary can have a separate ID (e.g. flight or voyage number).
    - Route: the use of the infrastructure (road, rail, etc.) taken by a transport means between any two locations of its itinerary.
    - Transport means track: the actual route and itinerary of a transport means for a given period. The period can be linked to one itinerary, like for a particular voyage number.
    - Transport leg: any two adjacent (in time) locations and transport means used to transport cargo from between these locations.
    - Cargo (container) Track: timed sequence of transport legs for cargo.
    - Logistics Chain: the outsourced business transactions provided by a service provider for the execution of a customer order, that includes a cargo track and for each business transaction its service provider linked to a transport leg within that cargo track. Each transport leg can be further decomposed into a logistics chain.
    - Customs Item: a sort operation on HS-codes (or another applicable customs classification) of cargo (incoming/transit) or products (import/export).
- User-specific events: a user (group) may be used to names for events, milestones, dates and times, and associated Digital Twins. Examples are 'load', 'discharge', and 'arrive'. These will be expressed in terms of the standard events, milestones, dates and times, and Digital Twins of the Knowledge Graph to provide a mapping between any two (different) user perspectives.

The DTLF Shapes Graph[10] can contain all cardinality and format constraints identified by the Reference Framework. Examples are:

- Various representations of a Digital Twin: one can have different representations of 'organization', 'person', and 'area of interest'. A location can for instance be expressed by its UN LOCODE, its coordinates, and a geo coordinate. A shape graph specifies the representation that is applicable to a particular data property.
- Standard structures for representing values (similar to the Core Components): this is applicable to measures and weights, currencies, date and time (that has to include a time zone), and prices, rates, etc. Additionally, these constraints may include the use of data types like integers and reals with a maximum length and decimal point (e.g. the accuracy of a weight in kg).
- Data constraints: any restricted values (code lists) of a data property. The same data property may have two or more lists of restricted values (e.g. different ways to represent a country code). The value restrictions can also be given by a taxonomy (e.g. HS-code). These data format constraints can also be complex, e.g. the structure of an ISO (sea) container identification or the postal code for a specific country.
- Rules: any constructs that specify dependencies between values of different data properties.

An initial draft of the DTLF Knowledge Graph has been developed based on existing open standards used in the various modalities. Primary focus of the draft was the construction of the Reference Model as Knowledge Graph. Extensions are currently made by including the concept of business services and support of an initial version of the business process choreography (to be included at a later stage).

[10] Shape graphs are technically represented by SHACL – Shape constraint language, a W3C (World Wide Web Consortium) standard.

These will lead to including code lists and any other format constraints.

Whenever applicable, available (open) semantic models must/will be included. Relevant concepts of these models will be imported and are the basis for DTLF concepts and/or data properties. An example is the geonames ontology that contains validation rules for postal codes of countries. It also contains geographical representations of for instance countries, provinces, municipalities, etc. These could be combined with for instance UN Location codes applied in (global) trade.

5.8 Next steps

Validation of the DTLF semantic model is initially based on a validation sheet. After having processed the comments of the FEDeRATED Semantic Team, a second validation can then be done by DTLF Sub-Group 2 team 2. Experience in alignment (interoperability and compliance) of data models from different transport modes and different standards shows that the development and validation is iterative and an ongoing process.

Secondly, the application of the semantic model needs further elaboration, it could have several uses. This also relates to an architecture (DTLF SG2 team 3) and the choreography. For instance, the model can be used to formulate the eFTI (Regulation EU 2020/1056) data requirements, but it could also support supply chain digitization like the Living Labs of FEDeRATED and pilots of FENIX. Further application of the model to support 'plug and play' is also required (DTLF SG2 team 1). It can support organisations in specifying their data requirements based on the business services they support, compliant with regulations.

Thus, there are many potential uses and electronic business documents that will have to be supported. This requires governance of the model and its utilization by associations, regulators, and enterprises. Governance will imply metadata representing a data set for each use/scenario.

Finally, the model needs to be aligned with existing semantic models like IATA One Record for air transport, the maritime single window SafeSeaNet and standards such as UN CEFACT BuyShipPay (MMT – Multi Modal Transport model), and WCO Data Model. This requires the construction of mappings between different environments and the support of various syntactical representations, e.g. the mapping of SHACL to XML or JSON. There also needs to be a relation between the model for supporting of deployment of a use case by for instance generating openAPIs. Currently, FEDeRATED is performing this analysis in cooperation with these organization, where they are willing to support.

According to planning, the next step will be the specification of the choreography as a basis for developing the Technology Independent Services.

6 Technical (data sharing) perspective

This section presents a data sharing architecture. It is based on compliance with the Technology Independent Services that are under development by DTLF II Sub-Group 2 Team 2. The architecture serves as a basis for authorities to access data stored by (trusted) platforms or IT systems of supply and logistics stakeholders for their task. It also includes public infrastructure managers that provide services to users of that infrastructure.

There are basically two classes of user scenarios underlying the architecture. The first class is that of business process integration between enterprises. It specifies the Technology Independent Services (TIS) of those enterprises (DTLF II Sub-Group 2 Team 2). The second class is that of authorities both requiring data of and providing data to enterprises. The latter is in their role of infrastructure manager; the first (subclass) of authority use consists for instance of eFTI support, but also of others such as the data pipeline concept developed by HMRC (Her Majesties Revenue and Customs) administration, the Dutch Customs Administration (DCA) and various projects such as EC FP7 CORE, further elaborated in H2020 PROFILE with DCA and Belgium Customs Administration (BCA).

Firstly, the concept of TIS compliance and authority access is explained (access will be governed by security). Secondly, a proposed architecture and its components are described and thirdly implementation variants are given.

6.1 DTLF Schema Compliance and authority access

6.1.1 DTLF Schema – and TIS compliance

Supply - and logistics enterprises **MUST** comply with (a subset of) the **DTLF Schema** as specified before. It means that (a subset of) the DTLF Architecture must be implemented, enterprises MUST be **TIS compliant** (Technology Independent Services**)**, authorities **SHOULD** have (controlled) access to (pull) and/or receive data (push - declarations) from those enterprises based on the goal binding principle, and enterprises **MUST** provide data accordingly to authorities.

Conceptually, the following figure represents this proposal, where TIS has been expanded to its various parts (see DTLF II SG2 Team 3):

- Publish, search, and find of logistics services, timetables, available capacity (transport means, infrastructure, etc.), etc.
- Book - agree on conditions and prices for performance of particular logistics activities. These might be for individual transactions or longer time periods (i.e. framework contracts).
- Order and plan the execution of logistics activities with reference to agreements made in the booking.
- Visibility – provide progress of the execution of planned logistics activities to customers.
- Cancellation for agility - adjust execution to changing conditions.

The list only states data requirements for each part of the TIS. Enterprises may decide on how to apply them. Data sovereignty and regulatory compliance determine who will be able to access which data, supported by security mechanisms (see section 7).

These are shown by the following figure. The figure also shows negotiable and non-negotiable services: for some logistics services, prices and conditions cannot be negotiated.

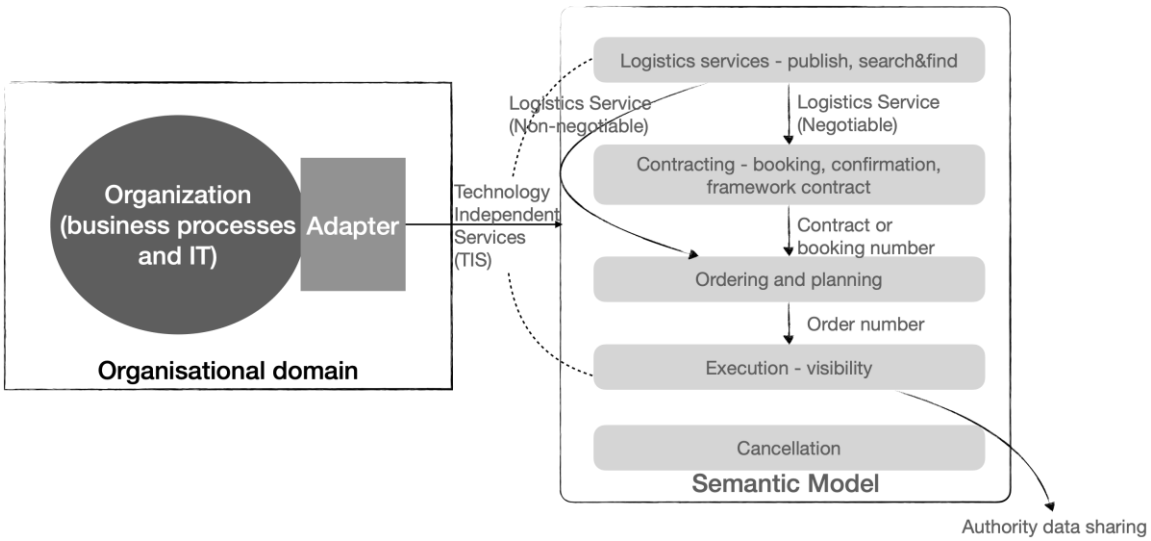


Figure 16 The concept of TIS compliance

Compulsory reporting to authorities (data push, see before) is based on data sets that are available during execution. Potentially, they need to be complemented with additional concepts like a customs item is required in customs declarations. Other regulations may support the pull mechanism.

The previous figure shows that authorities require access to visibility data, potentially linking with additional data represented by orders (consignments, shipments).

In this respect, the following three topics must be addressed: data sharing requirements or leading principles, implementation variants, and the role of authorities. These are the basis for specifying the functionality of what is a conceptual adapter shown in the previous figure.

A TIS consumer **MUST** implement any coordination mechanisms by its business processes and IT systems in case it consumes TIS of different enterprises. The reasoning is that a TIS implementation (e.g. provided by a platform) might be mode-specific, whereas a TIS consumer coordinates a multi-modal chain involving different TIS providers, each with their implementation.

Technology Independent Service comprise different features, namely (see also the European Interoperability Framework and its various aspects):

- **Functionality** - which part of the business process choreography do they support? The following functionality is currently identified:
    - Publication, search & find
    - Booking
    - Ordering according to agreements achieved during booking (e.g. framework contract based)
    - Visibility based on the existence of a commercial contract (order).
- **Data semantics** - what is the semantics of the data shared by the services?
- **Technical protocols** - which technical protocols are applied, e.g. messaging, REST APIs, including the data syntax (XML, RDF, JSON(-LD),...).

A subset of TIS implemented by an enterprise (via one of the options) **MUST** support the part of the TIS functionality and/or business services of that enterprise. This refers to an enterprise only

implementing for instance visibility or the relevant part of the semantic model fitting the business services (e.g. container transport by sea).

In case a TIS provider implements the subset of TIS functionality by a different implementation option, these implementations **MUST** share data. This rule is applicable to integrate for instance a platform supporting ordering with another one providing visibility. The following rules **SHOULD** be implemented in those solutions:

- Order - Contract: any order **MUST** refer to a contract concluded by booking either by having the order number equal to the booking number or containing the contract (booking) number and a reference (URI) to contractual agreements.
- Visibility - Order: visibility details **SHOULD** be provided by a reference to an order between a service provider and customer and its relevant details (order number and/or URI of the order).

Each TIS implementation **MUST** comply with the EU Cybersecurity Act. Compliance will be identified by Team 4 and elaborated by Team 1 of DTLF II SG2.

6.1.2 Data sharing requirements or leading principles

DTLF I SG2, FEDeRATED, and FENIX have developed several data sharing requirements and leading principles. Furthermore, there is a note on security in data sharing as a basis for security mechanisms to be implemented.

The main requirement to the architecture is: it needs to fit with existing knowledge and expertise, processes, and procedures, as much as possible. This main requirement eases adoption.

The leading principles are the following (see also DTLF SG2 Team 4):

- **Data sovereignty** – each participant in data sharing can act autonomous considering common agreements.
- **Data quality** – data that is shared must adhere to agreements made by the semantic model.
- **Trust** amongst participants – different levels of trust can be identified, namely commercial (related to logistics service provision), data sharing environment (trust in the technical solution), and trust from a cyber security perspective (e.g. Identity and Authentication using a trusted Identity Provider/Certification Authority).
- **Open and neutral** – all organizations must be able to share data, SMEs, large enterprises, and authorities.
- **Plug and play** – joining and departing the federated network of platforms at a single point of entry, without actions to be taken by other users of the federated network of platforms. This should include technical protocols, identifiers, and all relevant aspects identified by DTLF II SG2 team 1.
- **Level playing field** – from a logistics service perspective, all enterprises (SMEs and large ones) should have the same opportunities.
- **Rapid deployment of new TIS** – data sharing will evolve; a solution should be able to rapidly deploy any new data sharing services.

These leading principles are supported by the TIS and their semantic model, but they also must address mechanisms like Identity, Authentication, and Authorization (IAA), non-repudiation, encryption, and other cyber-security aspects (see EU Cyber-Security Regulation).

This requires addition of the following functionality:

- **Identity Provider** – any identities used by individuals must be certified by trusted Identity Providers. This must address trust procedures and certification of Identity Providers.
- **System (PKI) certificates** – systems have to be trusted based on certified PKI-certificates (e.g. eIDAS certified).
- **Non-repudiation** – logging and audit trail functionality must be implemented.
- **Encryption** – key distribution mechanisms must be defined, including agreement of the encryption algorithms that are applicable. Encryption needs to have agreement on various encryption levels like link encryption and end-to-end encryption in case various third-party components and platforms are used for data sharing.
- **Access Policies** – although enterprises are in control of their data, they will have to provide access to (trusted) authorities based on access policies. Access policies are provided by authorities and can be implemented by enterprises in their systems. Access policies will be expressed in the DTLF II SG2 team 2 semantic model.

The EU Interoperability Framework (EIF) is the basis for defining the various aspects that need to be addressed in data sharing, namely: technical protocols (including syntax), data semantics, organizational interoperability, and legal aspects. DTLF II SG2 team 2 addresses data semantics and organizational interoperability; team 4 addresses trust, data sovereignty, and governance for meeting 'open, neutral, and providing a level playing field'. Team 1 includes capabilities of creating a level playing field on business level. Team 3 should focus on technical protocols and should also consider the feasibility of supporting the requirements.

6.1.3 Authority data requirements

Authorities can implement push (declaration-based), pull, or a combination of both, for instance to retrieve commercial data additional to a declaration. In case of a pull mechanism, an authority needs to know from which IT system data can be pulled. Any IT system, e.g. enterprise system or platform supporting different enterprises, needs to identify the authority, the goal for which data will be pulled by the authority, and the data that is required by that authority based on the goal. The latter are the access policies of an authority.

An authority access policy will be based on a Regulation, is based on a data set for that Regulation, and optional a one-time admission must be provided (based on for instance a court order).

There have been and still are a number of (EU - and otherwise funded) projects and Living Labs (e.g. the Dutch Living Lab in FEDeRATED) that are exploring and developing the following view: data access to enterprise data is based on supply chain visibility. The progress of logistics activities of a transport means provides the actual state of the real, physical world. This state can have a reference to additional data like transport order data and – documents. An event-based interface between enterprises and authorities is required, where a set of events represents an itinerary of a transport means and updates of individual events the actual progress. This aspect is further elaborated by team 2.

6.2 Architecture

6.2.1 Data sharing options – messaging and events

The proposed approach of DTLF is to prevent data duplication, thus reducing errors during processing and storage by various stakeholders and increasing data quality. It implies associations between various Digital Twins of the DTLF semantic model will be shared with reference(s) to additional data. However, many applications are still message based, meaning data is pushed from one application to another. However, there will also be (future) applications that can be based on only pull, limiting the amount of data that is shared and stored (which contributes to sustainability of IT). Pull and push will therefore both exist and need further elaboration.

Push and pull can be described as:

- **Push** – data is duplicated from one stakeholder to another, applying messaging. Many data sharing applications are based on the push mechanism. Push decouples IT systems of data holder and – user. It enables a data holder to make data available at the time he wants, a data user to process the data at the time required and does not require data access facilities of a data holder. A data user needs to have data storage capacity (of course also relevant regulations and acts like GDPR, cyber-security - and data governance act.
- **Pull** – a data user is informed of the existence of new data with a link to that data, sharing event data. Data is retrieved by a data user at the time he needs it. A data holder must provide sufficient IT capabilities to provide the data and can implement (local) access control for data sovereignty. Communication capabilities (e.g. Internet) need to be available to access the data.

To support push and pull in one environment, the following functionality needs to be implemented:

- **Event extraction** (from push to pull). Whenever a data holder is not able to extract event data out of its data sets, i.e. it supports a push (messaging) mechanism, there has to be functionality for event extraction. This is implemented by a data holder.
- **Link (URI) evaluation** (from pull to push). Whenever a data user requires all data of various events, it will have to evaluate all links at the same time and construct the proper data set. This might be required in case there is no guarantee that a URI will be accessible at all relevant times, for instance for constructing a Master Airway Bill (MAWB) and its House-AWBs (HAWB) for an airplane when it is taking off or landing.

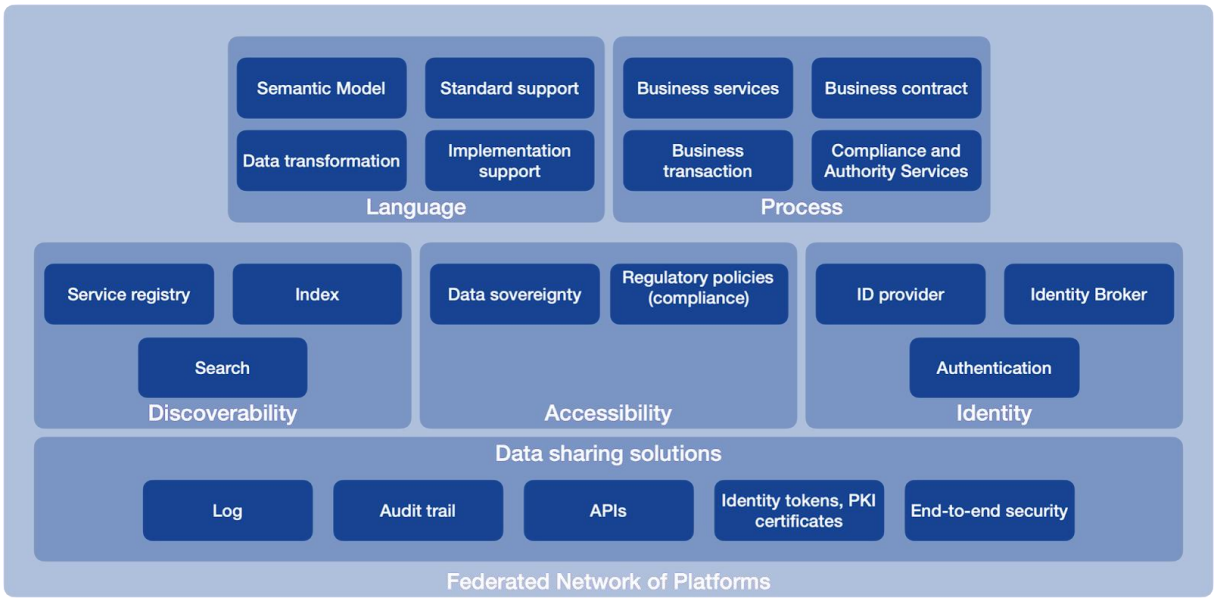
To be able to perform event extraction and link evaluation, data needs to have a common semantics. Thus, **data transformation** needs to be supported, where the input and output of data transformation can be based on existing (implementations of) open/defacto standards and/or internal data formats by organizations. Whenever additional data is retrieved by evaluating a URI, it will have to be transformed to the semantic model, structured in a syntax, and a data user will be able to transform it to a data structure and -syntax of its choice using the same semantic model. Data transformation configuration is part of plug and play.

The data structure of ‘event’ complemented with TIS functionality is specified in more detail.

6.2.2 Overview of functionality

Since data sharing will be implemented as a federated network of platforms, various technologies will have to be supported. To facilitate the implementation, technology independent data sharing

functionality will be specified. These are the so-called architectural elements, as shown in the following figure.



The architectural elements are composed of three groups:

- Conceptual – the specifications for a seamless goods flow. These are addressed by **process** – the integration of business processes of organizations for data sharing, supported by a set of technology independent IT services (interpretability and re-usability) – and **language** – semantics of (meta)data supporting existing standards and integrating with IT systems of organizations.
- Functional – the functional elements that support data sharing. These are **identity** – the ability to authenticate an identity -, **accessibility** – support of data sovereignty, compliant with regulatory data policies – and **discoverability** – functionality to find the proper data holder, based on metadata (service register) and visibility data (index).
- Data sharing solutions – generic functionality for non-repudiation and security, supported by Application Programming Interfaces.

Each stakeholder, either a data holder/-user and an IT (platform) service provider, decides on a specific implementation of the functionality, best serving its business model and governance structure. Such an implementation needs to comply with the architectural elements, including any standardized interfaces between the functional elements.

The functionality is defined as follows (see next table).

Table 1 – overview of components and interfaces

Component	Brief description
Model Management (language and process)	Model management supports development and maintenance of semantic model, utilizing existing models. There might be a model management component providing complete functionality and an easy-

Component		Brief description
		to-use version for constructing views on a model. Model management is decomposed into additional components specifying functionality.
	Language	All relevant functionality (components and models) for configuring data sharing. Tools are required to express data requirements and access policies in the semantic model; mappings to standard must be constructed. This is all configuration to data transformation.
	Process	Process is about integration of business processes of two logistics enterprises and the formulation of data requirements for regulations, including data sharing with those authorities.
Discoverability – the ability to share (publish, search, find) logistics state information.		
	Service Register	A Service Register stores two types of data: • Business services of a service provider. • Data that can be provided by a data holder (or a platform on behalf of more than one data holder). • Technical capabilities (e.g. supported technical protocols) The business services and data have a URI fitting with a URL of a service provider and/or data holder/platform. The business services are used in the context of a logistics contract. The data of a data holder can be used in progress of executing a chain. The technical capabilities are for instance the data standards and – syntax, and the technical protocols supported by an organization or platform.
	Index	An index stores all events representing the past, present, and future state of a relevant part of a logistics (sub)system. An event can refer to multiple data sets via a URI. The index may also contain a(n immutable) hash of a data set that can be pulled out of IT systems of a data holder. The future state relates to a business contract, e.g. orders, plans, and updated estimates related to a plan.
	Search	A search engine queries the index to produce relevant information out of event data. The queries can be specific to a data user. The find services interface can be applied to identify the type of data that can be provided by a data holder for an event.
Accessibility – access to data controlled by a data holder		
	Access Control – data sovereignty	Access management retrieves a regulatory access policy in case an authority requires data access and retrieves the appropriate data for the pull URI according to the access policy from an internal IT system (inhouse data). Access policies are regulatory – and internal access policies.
	Regulatory Policies Management	Regulatory Policy Management provides authorities with the functionality to formulate its data access policies, based on (a subset of) the semantic model. eFTI provides for instance a subset.

Component		Brief description
Identity – functionality for federation of identities		
	Identity Provider	An Identity Provider manages an identity of a user and can provide an access token to that user based on its credentials. The access token is on behalf of an organization that registers the user with its credentials.
	Authenticate	To authenticate the validity of an identity token.
	Identity Broker	An Identity Broker can authenticate an access token submitted by registered Identity Providers or another Identity Broker.
Data sharing solutions – generic data sharing functionality		
	Log and Audit Trail	A log contains the actual data that is shared and an audit trail its metadata like timestamp and organization activating a service. This type of data is relevant for non-repudiation and is input to the business model of a data sharing solution.
	Data exchange / APIs	Support of a technical protocol for data sharing.
	PKI certificates/ identity tokens	Used in the context of system-to-system data sharing with a technical protocol
	End-to-end security	A data sharing solution only has sufficient data for routing a data set to a proper data user, where all other data can only be read by the data holder and – user that share that data.
Additional functionality – all functionality for complying with the architecture		
	Data transformation	Data transformation transforms inhouse data to the required output syntax according to the semantic model. Data transformation uses model retrieval, used during data transformation configuration.
	URI evaluation	A function to compose a data set by pulling relevant data from data holders based on URIs provided in one or more events. The function may validate the (immutable) hash of the extracted data set(s) in case it is provided by the index. This function can be triggered from the event extracted by the search engine from the index.
	Event extraction	An optional function to extract event data out of internal data sets. These might be simple events like the existence of a type of document related to a business transaction including its URI or more complex events like a (partial) container track extracted from a B/L.

6.2.3 Distributed functionality and interfaces

These components are related to each other, based on interfaces. The following figure shows only the relevant interfaces, under the assumption that each role (data holder and -user) implements relevant components.

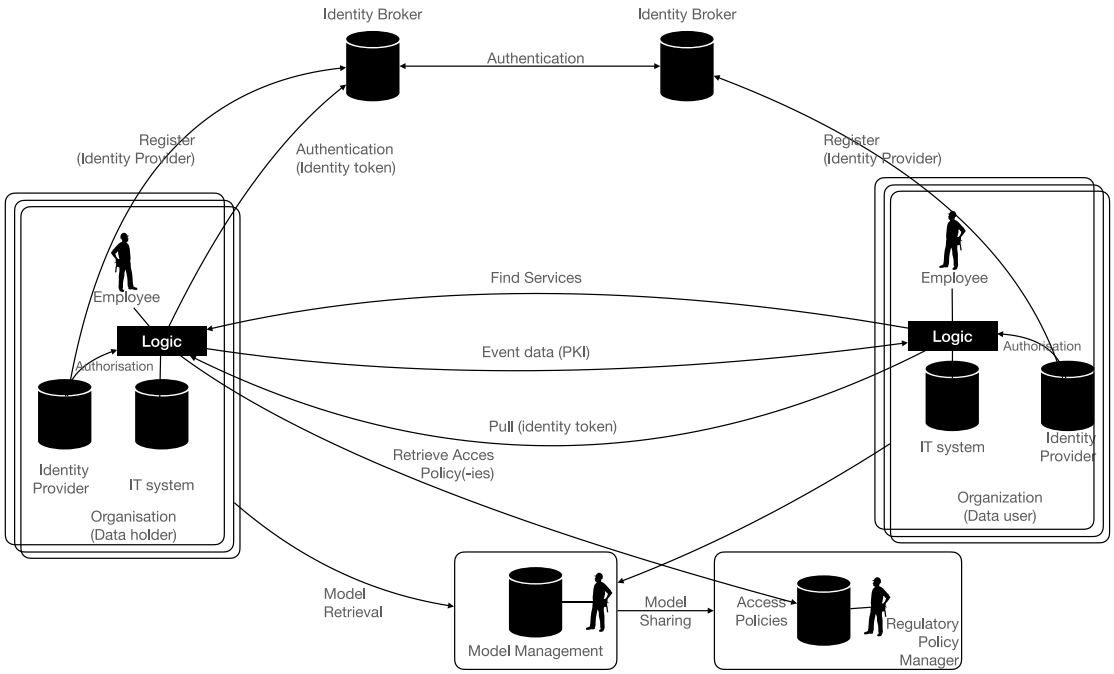


Figure 17 proposed architecture – high level overview

The figure shows several data holders and – users. Each organization can have both roles. In the next figures, only a single data holder and – user will be depicted; there can be multiple.

The proposed architecture is based on sharing links between various organizations in their respective roles (**data pull** with **Linked Data**). These links represent the status with respect to the parts of TIS, e.g. 'order' with expected locations and dates/times for transport of particular goods or 'report' (visibility) with the actual time of for instance discharge of those goods at a location. These are so-called 'events' that can represent associations between the various Digital Twins, locations, and business transactions. They include references (URIs – Uniform Resource Identifiers) to additional (document) data sets.

Conceptually, these events are shared via an index, where this index is fully distributed. Each data holder and – user can have its index with event data. The latter implies that each data holder and – user must have a data custodian for storing event data of the index. Each data holder is in control of distribution of relevant events to data users (data sovereignty). Each data user will be able to implement additional functionality on its index, like a publish/subscribe mechanism informing those new events are available.

The previous figure shows multiple Identity Brokers to support federation of identity and authentication. Each data user will apply its Identity Provider of choice that has to be registered by an Identity Broker. There will be multiple Identity Brokers, both in the public – and the private domain. The public domain Identity Brokers are based on the eIDAS Regulation for Business-to-Administration (B2A). A private domain Identity Broker will support B2B. Since till now A2B based on pull has not yet been considered, a solution must be developed.

These external services can be community, commercial, private, or public owned services. They support authentication, but they also must be trusted.

The proposed architecture works as follows:

- **Set-up phase**. An organization develops logic for interfacing with another organization as data user and/or -holder, referring to (a subset of) the Technology Independent Services, implements the relevant identification, authentication, and authorization (access policies) mechanism, implements the required data transformations for integrating with internal IT systems, and configures the rules for distribution of event data to relevant data users. Access control and distribution considers the implementation of relevant regulations: national, EU, and/or global regulations must be implemented with their access policies. An authority identifies the Regulations it implements, specifies its access policy for those Regulation(s), implements a local search on event data sets, and the services it provides (in the case of an infrastructure manager).
- **Registration phase**. An organization registers the Identity Provider it wants to use with an appropriate Identity Broker. Enterprises also register with the Service Registry, including the (subset of) TIS they support and its syntax for data pull. Their logic will disclose their logistics services and all other parts of the Technology Independent Services.
- **Deployment phase**, which comprises actual data sharing between enterprises, i.e. between logic of individual enterprises according the choreography with distribution of event data with URIs, including sharing this event data with proper authorities in the context of a regulation or on voluntary basis. Authorities can search for Digital Twins on their Index, which may include a URI. In case authorities in a Member State implement one index, that index requires access policies specifying that individual authorities in that Member State can only access data relevant to the Regulation(s) they implement, e.g. a road inspection authority will only access to trucks and not barges for inland waterway transport. Search and pull of data are based on APIs with OAUTH2.0 identity tokens that can be validated by any providing these APIs.

This brief overview of functionality already indicates a choice, namely OAUTH2.0 for identification and authentication. Yet, other interfaces must be specified, both in terms of their semantics and syntax. For instance, the semantics of an ‘event’ must be defined as part of the TIS by Team 2, whereas its syntax must be defined by team 3. The same for the pull interface of an authority, where for road transport the UN CEFACT eCMR data structure might be applicable (depending on the implementation of the eFTI Regulation).

6.2.4 Interfaces between components

There are various interfaces identified between the components of the architecture. These interfaces are specified by the semantic model (see the following figure).

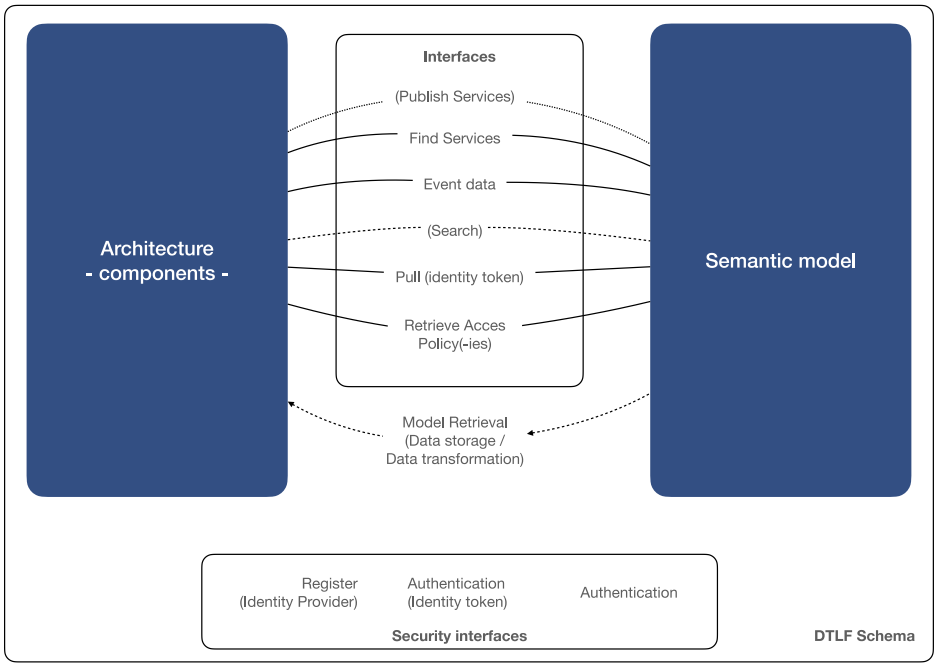


Figure 18 architecture, semantic model, and interfaces

The figure shows that the security interfaces are not specified by the semantic model. These are generic (open) standards. Also, model retrieval is outside scope of the interfaces, it is used to configure data transformation and specify (temporary) data storage. The search interface is in brackets, it is optional as indicated before, depending on the implementation.

According to the previous figure, the following interfaces **MUST** be supported by a data holder and data user:

- **Event interface** – this is the event API where either an enterprise automatically publishes data or a data user collects data ('listener'). It supports the Technology Independent Services as specified by DTLF II SG2 Team 2, both those of enterprises and Infrastructure Managers.
- **Pull** – this is a simple API with a URI, reason or goal (e.g. identification of a Regulation) and an OAUTH2.0 token identifying the authority requesting access to enterprise data. Potentially, it includes a (reference to) a certificate whereby the requester is officially granted access. The result of the pull is based on the configurations provided by the Service Registry; it requires data transformation.
- **Find Services** – the capability to find business services, also in the context of data that is required and/or can be provided by a data holder.
- **Model sharing** – (parts of) the semantic model should be available to authorities for formulating their access policies and to enterprises and authorities to integrate all the search, pull, and TIS interfaces with their internal IT systems.
- **Retrieve access policy** – this is an API call for retrieving access policy(-ies) specified by the semantic model. Its syntax is still to be defined.

The choice of a syntax (and a data structure in that syntax) might also depend on the implementation, see next paragraph. The security interfaces for Identification and Authentication must be included in this list, but these are supported by open standards.

Additionally, two more interfaces may be defined, based on two implementation variants. First, a **search interface** might have to be developed, since it specifies the data requirements of a recipient, especially authorities requiring access to data to support their governance processes. A search interface is a local interface of an authority accessing event data stored by an (local) index. There can be multiple search options; the set of these search option is called the 'metadata search'. It is a search on a Digital Twin, e.g. a vessel, a container, etc. The return of the search might be simple, e.g. provide the content of a truck, or complex, e.g. an itinerary of a transport means or a container track. This search result, expressed in the semantic model, is defined per authority.

Besides TIS implementation for business process integration and compliance with regulations, there **MAY** be various providers of (commercial/paid) information services. Weather forecast services, ETA calculation, data transformation, and many others can be distinguished. These information services also require a form of standardization but can be included in the Technology Independent Services by a platform or any other solution which will allow platform providers to distinguish their services from others whilst still complying with the Technology Independent Services. These are not included in the previous figure but can simply be added. An API Registry can be included in the previous figure for storing and retrieving External (third party) services. These services, expressed in the semantic model, can be applied by any data holder or -user to improve its data quality. An ETA service for road transport is an example; an anomaly detection of a vessel between two ports in an itinerary can be another one.

The focus is currently on TIS for business process integration and regulatory compliance; in future these TIS are extended to cover additional functionality. Any TIS extensions are governed according to the rules defined by Team 4.

6.2.5 Organizational implementation variants - platforms

Implementation variants relate to organizational – and technical choices. Technical choices will be given in the next section.

Organizational choices are made in combining functionality of two or more organizations. According to the architecture, each organization is responsible for implementing its own functionality and the relevant interfaces supporting that functionality. However, organizations can also decide to share functionality, either as community or of a commercial service provider. It concerns the functionality of the Index, the Service Registry, and potentially the data storage function:

- Shared Index – two or more organizations use the same Index. All event data shared by a data holder shared with any of these organizations is stored in the Shared Index. A Shared Index can for instance be used by authorities that need to access similar data of enterprise.
- Shared Service Registry – two or more organizations share a service register, for data – and/or business service brokerage. A Shared Service Register may improve discoverability of data and business services.
- Shared data store – the data that is referred to in an event is stored in a shared store, decoupled from internal IT systems of organizations. A shared data store may be used to ensure performance and availability of data referred to in an event.

These choices will lead to different ways of implementing the interfaces.

6.2.6 Technical implementation variants

The technical implementation variants refer to the technology used for implementation of the required interfaces and functionality by an individual organization. Since platforms relate to organizational choices, these platforms will also provide data sharing functionality. Therefore they are also considered in this section.

DTLF II does not impose any implementation of TIS. Some examples of implementation variants are shown in the document of team 2. The basic implementation variant is the one shown by the architecture. It is a completely distributed implementation, where each organization adapts its logic to be compliant with the DTLF Schema and the TIS and implements the various software components to support the required functionality. DTLF distinguishes the following implementation variants:

- **Organization** – an organization implements the required components itself, either using open source – or Commercial Off The Shelve (COTS) solutions.
- **Platform (shared functionality)** - a community or commercial platform implements (a subset) of the required functionality on behalf of its users, enterprises.
- **Peer-to-peer** – an organization uses standard technology (connector) and services of an infrastructure service provider (broker- and clearing house services) for data sharing as developed by the International Data Space Association (IDSA).
- **FENIX connectors** – a platform or organization implements a FENIX connector for interfacing with others. This is a variety of the previous implementation variants.
- **Distributed, immutable database** (or Blockchain – (BCT)/ Distributed Ledger Technology (DLT)) - each organization has its local interface to a data sharing infrastructure with a BCT network. There are various technologies that can be applied in different ways. One of the choices is to store data on- or off chain; events can for instance be stored on-chain and data referred to via these events off-chain. This latter variant will be shown.
- **eSens e-Delivery** – a connector supporting technical specification and developed as open source for data sharing between organizations.

An organization that is responsible for operating a particular variant, can outsource to a cloud service provider. A platform can for instance be deployed by a cloud service provider; a distributed immutable database can be deployed by more than one cloud service provider, where each cloud service provider operates a node of the network. An organization or platform service provider can also apply managed cloud services for development and deployment of the required functionality, where the cloud service provider provides a variety of software components to rapidly develop and deploy new services.

Besides these variants, particular components and their functionality also need to be implemented as an infrastructure facility. Governance of this infrastructure facility is dealt with separately in another section of this document.

The distribution of functionality is shown in the following table.

Table 2 – distribution of functionality for the implementation variants

Component	(commercial / community) Platform	FENIX connector	IDSA Reference Architecture	Blockchain based applications
Service register	x	x	x	x
Index	x			x
Search engine	x			x
Identity provider	x			
Data transformation	x			
Log and audit trail	x		x	x
Authentication and access control	x	x		x

Not all components are listed; functionality that is not listed in the table is not (yet) implemented by existing solutions. The figure shows a commercial or community platform is able to offer all functionality. However, this functionality and its supporting interfaces may only cover part of the overall functionality of the information space. They are also not open and neutral, and do not provide a level playing field. Another note is that a FENIX connector can be used both for platform interoperability, but also as peer-to-peer implementation.

This table also does not state whether the solutions comply with the interface specifications. One can have for instance two different implementations using the same technology that are not (yet) interoperable.

6.3 Next steps

The next steps are to further specify the various interfaces and provide examples of how the architecture can be applied in practice. The interfaces will be specified in terms of the semantic model, supporting the Technology Independent Services, infrastructure management services, and compliance to regulations.

Additionally, the mapping to existing solutions will be described, for instance the expression of visibility milestones from a user perspective in terms of the semantic model. Another aspect is the exploration of applying the semantic model for transformation and support of existing (implementation guides of) standards.

7 General perspectives

General perspectives describe additional functionality required for a trusted and safe data sharing infrastructure. It considers the perspectives security, which underpins the architecture, non-repudiation/accountability, and finance and payment.

This version addresses the security perspective only. Other perspectives will be included (see next steps).

7.1 Security perspective

DTLF SG2 Team 4, the H2020 PROFILE project, and FEDeRATED elaborated security in the context of data sharing. This is described by the section.

There is a need for a general security framework for data sharing. Such a framework is an addition to a cybersecurity framework (e.g. the EU Cybersecurity Act) that can be applied by any organization in the context of digitization. This memorandum specifies a general security framework for data sharing between organizations (i.e. the interorganizational domain). This data sharing security framework can be applied by Customs authorities in accessing external data sources.

This security framework assumes that each organization that participates in data sharing has sufficient cybersecurity measures implemented, i.e. they are 'trusted' data sources. If they don't have sufficient measures taken, not all measures described in this memo might be that effective.

Any rules and mechanisms described in this memo might be overruled by the implementation of an EU Data Regulation that is under development as part of the EU Data Strategy. This memo not only provides mechanisms that can be implemented, it also relates rules of conduct in the context of data sovereignty and - provenance (data provenance: an audit trail that accounts for the origin of a piece of data (in a database, document or repository) together with an explanation of how and why it got to the present place).

The structure of this section is as follows:

- Background – explaining the organizational and interorganizational domain and various roles for data management relevant to individual organizations.
- Data sharing security objectives – the relevant objectives in the context of data sharing between any two organizations.
- Security mechanisms – mechanisms to deal with the potential risks.
- Implementation levels – implementation of the mechanisms in the context of data sharing via platforms.

7.1.1 Data security domains

In this respect, separation of concerns is applied by distinguishing organizational – and interorganizational domains:

1. Organizational domain – each participating organization is expected to implement security measures according to international open standards and frameworks (references to be included). The assumption is that each organization has implemented the necessary measures

for identification and authentication of its employees (both physical access and access to IT facilities) and to prevent cyber-attacks to data, processes, and technology.

2. Interorganizational domain – additional measures need to be taken to address passive – and active attacks on data when shared. Passive – and active attacks in the interorganizational domain will be discussed in this note.

In the organizational domain, a distinction between data owner, - steward, and - custodian[11] is made, where the role of data steward is also relevant to the interorganizational domain (i.e. platforms can have the role of data steward):

- Data owner: the role that is authorized to create, delete, change, and share or data with or provide access to data to other actors. When an organization is receiving data of another organization, it basically becomes holder of that data. Data provenance might be part of data ownership: reference to an original data source. Customs can for instance become owner of declaration data, where a trader is the source. The role of 'data owner' refers to the business processes of an organization.
- Data steward: the role that is responsible for utilizing an organization's data governance processes to ensure fitness of use of data elements - both the content and metadata.
- Data custodian: the role that is responsible for software and hardware integrity and – availability to support a data steward. It is the computing centre used by an organization, which can be a cloud service provider or internal to that organization. A data custodian also must implement various security measures to prevent unauthorized (physical) access to data processing facilities. This latter is outside scope of this document.

An organization (enterprise/authority) can be a data owner, where a platform acting as data steward, and that platform has outsourced actual data processing and storage to a cloud service provider. Furthermore, a platform may act as data steward for several data owners, where these data owners have agreed on data governance. Data steward and – custodian roles can be combined and are treated as such in the following description. It is assumed that each data custodian implements the necessary measures to prevent unauthorized (physical) access

A data steward and data owner may have separate data policies with respect to making data externally available. The basic assumption is that data access complies with data policies of a data owner that are agreed between the data owner and data steward.

7.1.2 Data sharing security objectives

Data sharing security objectives are relevant to organizations in their role of data owner and should be applied to any (intermediate) platform(s) that act as data steward (and have implemented a data custodian role) during data sharing between those data owners.

[11] The roles of data owner, - steward, and – custodian are a further refinement of the roles data holder and – user, see before. They have to be distinguished from a security perspective to identify security measures that need to be implemented.

It is not the objective of the H2020 PROFILE project to perform a thorough risk analysis. PROFILE Security is much often related to cyber-security with all kinds of threats. Basically, security risks can be classified as:

- Passive attacks – unauthorized access to data and data processing facilities.
- Active attacks – unauthorized changing or creation of data and data processing facilities.

There are many means by which these attacks occur like spoofing, (distributed) denial of service, man in the middle, etc.

With respect to the interorganizational domain, the following goals must be achieved to prevent passive - and active attacks:

- Authentication – is the person, organization, or system really the one that it states it is.
- Data confidentiality – data is not disclosed to unauthorized users, i.e. any user that did not implement the hereafter mentioned security mechanism for data sharing. Data confidentiality is identical to data sovereignty.
- Data integrity – is the data that is received identical to the data that has been sent.
- Non-repudiation – immutable proof of shared data in cases of disputes.

To support these goals, various mechanisms can be implemented. These are discussed hereafter.

7.1.3 Data sharing security mechanisms

Only authorized access to data is required to prevent any risks. This is at the following levels:

- Identity and authentication – any user accessing data via a function (or API) needs to have a verified identity that can be authenticated. The following 'users' are distinguished:
    - Persons that are employed by an organization. Persons have roles and capabilities which grants them access to data. Each organization is responsible for organizing identification and authentication of their employees.
    - Identity providers used by organizations will be part of the infrastructure. Existing mechanisms like Decentralised Identities (DID[12]) and iSHARE[13] are topics for research.
    - Organisations will also have an identity in this infrastructure. This identity is related to employees and systems like indicated, but also relevant for trade. As such, identity is a data property of an organization in the semantic model, where it can be used in data sharing.
    - Systems identifications. This concerns both IT back-office systems of organizations, assets with a sensor (Internet of Things – IoT), or platforms acting as data steward.
- Access control – the rules by which a user can access data. Access control provides data confidentiality or more particular data sovereignty where individual data owners can decide

[12] DIDs are a new type of identifier for any object (e.g. person, organization, cargo, semantic model) that can be validated against agreed credentials (e.g. a Chamber of Commerce registration of an organization). DIDs are verifiable and decentralized.

[13] iSHARE is a uniform set of agreements or scheme for data sharing between organizations. It contains a register of certified Identity Providers and is based on open standards (OAUTH2.0 and XACML), where OAUTH2.0 is extended to support delegation.

themselves to whom they intend to disclose data sets. Access control can be based on a data classification (e.g. open data versus data shared in a commercial relation) and mandatory regulations by which particular data sets need to be accessible to authorities.
Access control can be formulated by for instance XACML (XML Access Control Markup Language), based on the semantic model. There is a processing model underpinning XACML, consisting of for instance policy enforcement points that must enforce access control. A data steward can provide a policy enforcement point supporting data policies of data owner(s).
Access control can be based on roles (RBAC – Role Based Access Control), i.e. Customs always has access to particular container data and attributes (ABAC – Attribute Based Access Control), i.e. Customs only has access to data of movements coming in or going out/passing its territory. As the examples indicate, an RBAC and ABAC mechanisms might have to be combined.
Access control will be based on data policies/- governance. We distinguish two types of data sharing policies:
    - Policies enforced by regulation(s). A data owner must implement these policies. Governance of these policies is implemented by (a) Regulation.
    - Policies of a data owner. A data owner can define its data policies, implementing a data classification. That data owner governance its data policies. These data policies might be agreed amongst data owners and are thus privately governed access control policies (this memo does not focus on data policies for an individual organization, only on data sharing).

  Any data sharing policy needs to comply with existing regulations like GDPR and new regulations.
- Data encryption – the data that is shared cannot be accessed or changed by any unauthorized user during its exchange. A hash calculated on data fields and added to the data before encryption provides data confidentiality and data integrity. Data encryption can also be implemented in such a way that it supports data authentication, data confidentiality, and data integrity (see hereafter).
  The main issue with data encryption is key management: data is encrypted/decrypted by using keys. These keys need to be changed frequently and must be stored in a tamper resistant environment to prevent unauthorized access to these keys.
- Log and audit trail – each data set shared between any two organization is stored in a log, including an audit trail of actions during sharing. A log and audit trail can be implemented by every data provider and – consumer but can also be (immutable) stored in a trusted environment (i.e. a clearing house or a distributed ledger).
- Distributed ledger – a distributed ledger can provide data integrity by immutable storing a hash of the data that is shared. It can also be used as a log and audit trail.

Data encryption and authentication are often combined with the application of asymmetric encryption methods. The data is first encrypted with the public key of the recipient (so only the recipient can access the data) and this encrypted data is again encrypted (or complemented by a hash) by the sender with the sender's private key (so a recipient is able to validate the origin with the public key of the sender). This is a well-known and often implemented mechanism.

The following matrix shows which security goals are covered by measures.

	authentication	confidentiality	integrity	non-repudiation
Identification and authentication	x (data user)			
Access control		x (data user)		
Encryption	x (data holder)	x (data links)	x	
Log and audit trail				x
Distributed ledger			x (immutability)	x

Note that a distributed ledger is a technology that can combine various mechanisms like identification and authentication (permissioned ledger) and encryption, but basically is an immutable log and audit trail providing non-repudiation and data integrity. Data integrity on a ledger can be either integrity of off-chain data (i.e. data that is not stored on the ledger) by storing a hash on the ledger and on-chain data.

7.1.4 Implementation levels of data sharing security mechanisms

Whereas each data owner can implement its data policies by an access control mechanism that can be shared, the other mechanisms can be implemented at two levels:

- End-to-end – solutions are applied over a network of platforms, i.e. from a data owner, acting as data holder, to a data user across at least one platform acting as data steward. Organizations sharing data over the various platforms can authenticate any data owner, independent of the underlying platforms and links.
- Links – each link implements the solutions, meaning that a chain of trust is created. Each link between IT back-office systems and a platform, between platforms, and platforms and Customs authorities need to implement the identified mechanisms. Each system involved will implement the necessary security mechanisms in its role as data custodian (see before).

The following figure relates end-to-end and links, showing where mechanisms can be applied. It also shows the various data holder/-user relationships, as defined by the Data Governance Act.

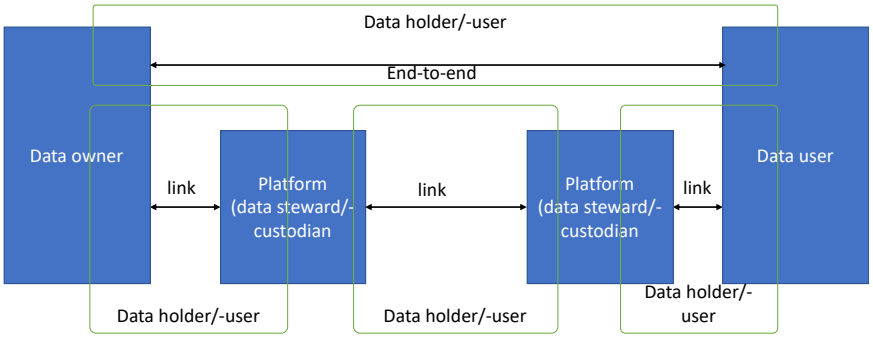


Figure 19 link – and end-to-end encryption

7.1.4.1 Link security

Link security is mandatory. It is about system-to-system data sharing. Link security is supported by PKI[14] (Public Key Infrastructure) certificates (e.g. from an eIDAS[15] (electronic Identification, Authentication, and trust Services) certified organization) over for instance 'https' or other protocols that support encryption. It requires that each platform acting as a hub in the chain of links from data owner to data consumer has implemented the required organizational security mechanisms.

PKI certificates provided by an eIDAS certified organization seem to be sufficient, although solutions like iSHARE also provide a register of participating systems that can be authenticated.

7.1.4.2 End-to-end mechanisms

It is up to users to agree on applying end-to-end mechanisms. End-to-end identification and authentication contribute to trust amongst organizations that do not know each other and needs to be developed to enable logistics innovations like synchromodality and agility. Applying end-to-end mechanisms will impose restrictions on platforms and solutions, like argued hereafter. Alternative solutions are the creation of a 'network of trusted networks' like developed by the IPCSA (International Port Community Systems Association), which may however not be sufficient since IPCSA covers hubs and not complete supply chains.

7.1.4.2.1 End-to-end data encryption, - authentication, and - integrity

In this case, we distinguish between the actual data that is shared end-to-end (the 'payload'), and control information required by a platform to take actions like routing the payload to its proper recipient.

End-to-end data encryption can only be based on an infrastructure that is completely agnostic of the payload or based on a symmetric key mechanism by which a platform is able to decrypt and process the data and encrypt it for the intended the data user. The latter is only required when the payload needs to be transformed by a platform. Asymmetric keys can also be used, but these only cover links like the link between a data source and a platform.

There are various technologies like Distributed Ledger Technology (DLT), connectors of the International Data Space Association (IDSA), a solution of sharing links via triple stores adopted by IATA, and the e-SENS[16] Delivery component, that are perfectly able to implement end-to-end encryption and data authentication. The payload can be encrypted; control data is required for routing the data to the intended recipient(s). It implies that such a data sharing solution does not add

[14] PKI is a framework which supports the identification and distribution of public encryption keys at links between IT systems. The framework provides authentication, data integrity, confidentiality and non-repudiation.

[15] eIDAS is an EU Regulation on electronic identification and trust services for electronic transactions in the European Single Market. It entered force in 2014 and applies from July 1st 2016. Besides support of PKI certification, certified identities for human interaction between citizens and representatives of enterprises with government services are provided, where the attributes of the certificates are harmonized.

[16] The aim of e-SENS is to facilitate the deployment of cross-border digital services through generic and re-usable components, like e-ID, e-Documents, e-Delivery, semantics and e-Signature. These components are basically agnostic of data semantics to make them applicable for all eGovernment services (www.esens.eu).

functionality, e.g. it cannot perform any data transformations of the payload. These solutions are all peer-to-peer solutions; they require link security.

In the same way, end-to-end data authentication and - encryption is feasible on the payload, whilst also safeguarding data integrity. Encrypting the payload with the private key of the sender provides end-to-end authentication. A data user can clearly identify the data holder by applying the providers public key for decrypting the data. Data integrity can be provided by adding a hash to the data based on the private key of a recipient and known data elements (or on the payload itself, see next figure). In case data is transformed into another structure, the values of these data elements must be found in a transformed payload.

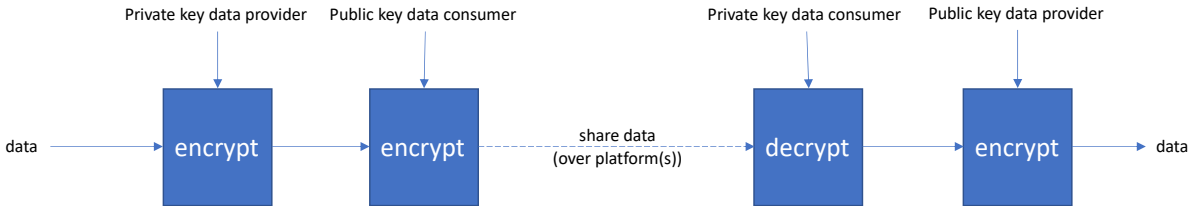


Figure 20 confidentiality and data integrity implemented by asymmetric encryption

The latter solution of hashing is a digital signature to data that refers to a user. The user either signs the data in its own system, which means the proper rights are applied, or the user signs the data in the system of someone else. In the latter case, the user should have access to his/her proper private key and be able to sign the data with that key, without disclosing the key to the platform used.

This relates to identification and authentication of a user. Whenever private keys are used to encrypt or authenticate data, these keys should be stored in a tamper proof environment and never leave this environment. The tamper proof environment, like a password manager or an electronic wallet, should destroy itself and its contents whenever an unauthorized user tries to access it. Access should be controlled by multi-factor authentication (see next text).

7.1.4.2.2 End-to-end identification and authentication

In case a user requires access to data stored in another system, end-to-end identification and authentication is required. Of course, platforms can have their mechanisms for multi-factor authentication based on for instance a card, a PIN or a password, and a form of biometric identification. For instance, electronic banking applications running on smart devices are protected by a PIN, where the device is also protected by (another) PIN. Potentially, also SMS can be applied to share an additional PIN or a barcode could be scanned to access the application.

End-to-end identification and authentication require:

- Each organization is responsible for its internal user management, including rights of these users. The rights can also cover the ability to share data with other organizations.
- It is up to each organization to install its own Identity Provider and Certification Authority or to outsource it to third party. Most organizations with internal IT applications have this functionality installed.
- In case Identity Provision and Authentication are outsourced for data sharing with other organizations and the organization still has its own internal Identity Provider and Certification mechanism, this latter needs to be synchronized with the external one. This is to prevent any

situation where a person is not an employee anymore but is still not deleted as such from the external provider's records.

- Each Identity Provider and Certification Authority has to be accessible via open standards like OAuth2.0. Multi- or two-factor authentication might be applied, where another channel (e.g. SMS) than the one used for data sharing is applied for sharing for instance a PIN.
- There needs to be one or more Registries with known and trusted Identity Providers and Certification Authorities. The Dutch iSHARE organization is providing such a registry.
- A Registry needs to verify the identity of an organization with one or more attestations, e.g. a Chamber of Commerce Registration, LEI (Legal Entity Identifier) registration or an EORI number (implying identity has been verified by a competent (Member State) authority).
- Whenever any two organizations require end-to-end identification and authentication, the registration of both should be based on one or more attestations of each organization that the other one accepts.
- Any APIs (Application Programming Interfaces) require an OAUTH2.0 token that can be authenticated by the relevant Identity Provider. The OAUTH2.0 token maybe have a lifespan, like a period or several times it can be used.

A mechanism for end-to-end identification and authentication, including the construction of distributed registries of trusted Identity Providers is under development via the Dutch iSHARE initiative. Whenever users select end-to-end identification and authentication, they need to select this mechanism. These distributed registries of Identity Providers support a federated Identity and Authentication mechanism.

7.2 Next steps

The next steps are the elaboration of non-repudiation/accountability, clearing and settlement of the use of components of the architecture, and identification of potential other perspectives like monitoring.

Non-repudiation is the capability of providing (immutable) proof that data has been shared. It is supported by an audit trail of actions that have been taken and a log of the data that has been shared during these actions. Basically, each stakeholder should implement an immutable log and audit trail and should provide proof of immutability. In many implementations, a centralised notary function is implemented for this purpose, but ledger technology may also be used to provide integrity of a log and audit trail.

Clearing and settlement relates to the capability to utilize IT services of components provided by different service providers. Since DTLF Schema compliance (section 6.1) is one of the principles for implementation reducing vendor lock-in and implying that any organization can easily use IT services of different providers. Addressing this perspective relates to pricing schemes of service providers and implementation time of IT services of any of these service providers. This refers to both the architecture (team 3) and plug and play (team 1).

The requirements of these various perspectives will be formulated by DTLF II SG2 team 4. Impact on functionality will be elaborated by team 3.

8 Final remarks

This intermediate report contains the main elements for constructing an open and neutral data sharing infrastructure based on a federation of platforms, meeting the requirement for a level playing field. Those main elements are:

- DTLF Schema – the set of interfaces and semantic model required for implementing the federation of platforms. The DTLF schema is the subject of governance and encompasses all relevant elements.
- Roles – the main roles for data sharing and DTLF Schema as a basis for deployment and governance of the DTLF Schema.
- Semantic model – a specification of data shared for freight and logistics as part of the DTLF Schema, where the specification has a technical representation in an open standard.
- Architecture – the architecture with interfaces, where the interfaces that support business process collaboration (i.e. Technology Independent Services) and compliance, are part of the DTLF Schema. An indication of potential implementation(s) is given.
- Security perspective – identification of the various interfaces and functionality required for safe and secure data sharing, supported by open standards.

Next steps are the proposal for a governance structure and -procedures, specification of the Technology Independent Services supporting business and authority data sharing (section 4), detailed specifications of the various interfaces and the role of the semantic model in data sharing, and construction of plug and play functionality. Furthermore, implementation variants will be elaborated and, when necessary, additional perspectives will be developed. Support of open standards and relevant models will need further elaboration, including validating that the semantic model can support the integration with existing solutions and platforms. All these actions are specified in the DTLF II SG2 work plan.