MASTERPLAN - TECHNICAL SPECIFICATIONS
HIGH LEVEL ARCHITECTURE

Introduction

This document defines the high-level architecture – technical specifications[1] - for a federated network of platforms approach. This EC DTLF approach aims to enable any participant in the logistic chain to act as a decentralized node, enabling data sharing at source in a secure way with any other.[2] Thereto the FEDeRATED has issued in a wide range of documents stretching from a Vision document to an overarching Reference Architecture report elaborated in detailed technical descriptions featuring IT components.

The four technical specifications a node should comply with are:

1. Semantics,
2. Identification, Authentication, and Authorization (IAA)
3. Service Registry
4. Index

In this document, firstly semantics – constituting the core of the FEDeRATED architecture – is described, followed by its implementation in the Service Registry (second) and Index (third). Fourthly, Identification, Authentication, and Authorization (IAA) are touched upon. On all four technical specifications separate documents are available. In addition, fifthly, this document introduces a semantic adapter. Although it is not listed as a separate technical specification, a **semantic adapter** is required by any node to adapt existing IT systems to semantics.

In this document, the maximum and minimum functionality of the Service Registry, Index and Semantic Adapter are listed. Minimal functionality' focusses on the need to fit into existing processes and IT systems that support part of the other functionality e.g., implementing framework contracts and/or (unstructured) person-to-person communication.

Organizations will always have a different speed of implementation and adoption of the concept. This is to be elaborated in an implementation strategy, not in the least to ensure interoperability. The adoption of the proposed architecture concept depends on the digital readiness obtained and the sense of urgences felt by the many organizations participating in the EU supply chain. The competitive and green agenda value of federated data sharing is elaborated in a great variety of EU and EU Member States policy documents.

1. Semantics

The essence of semantics is to specify machine-readable data to enable stakeholders in multimodal supply chains to exchange information digitally (data sharing in paperless transport). This relates to business transactions as well as compliance to regulations.

[1] Being identified as IT tools by some

[2] The FEDeRATED Vision is to develop an federated network of platform approach as an infrastructure provision containing a set of arrangements and technical applications to enable data in existing IT systems (platforms) of companies and public organizations to become available to users through a pull mechanism

Machine readable specifications of data are by applying semantic web standards like Ontology Web Language (OWL), Resource Description Framework (RDF), and SHApe Constraint Language (SHACL).

This chapter introduces the position of semantics with respect to other developments and presents the structure of semantics, whereby its role in the architecture is provided. The overall view of semantics is shown in the following figure and will be detailed hereafter.

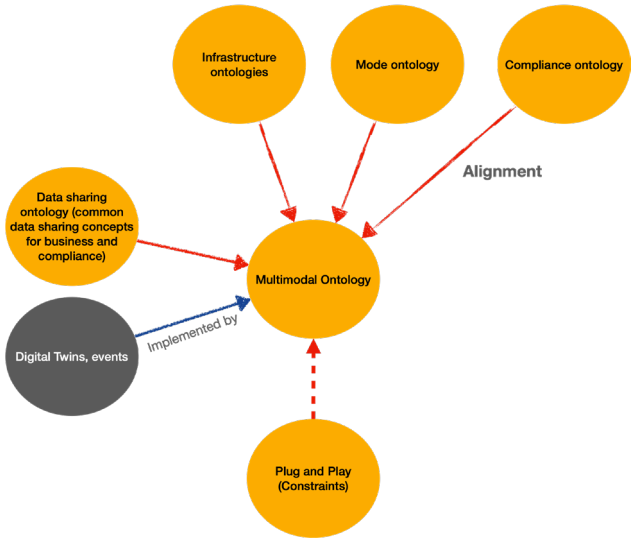

Figure 1 – overview the ontology components and structure

1.1 Context of a multimodal data sharing model

This paragraph identifies the position of semantics with respect to other developments, thereby also touching on the issue of ontology. Ontologies are used to represent a domain of discourse as a common ground for encoding content meaning and user interests.[3]

The federated network of platforms approach focusses on seamless multimodal freight transport. Thereto, the EC has identified 4 DTLF building blocks[4] that must be met. Plug and play' is an important building block that can be applied by communities in their domain when these communities develop an ontology i.e., re-using their existing standards and the multimodal ontology. To realize the federated network of platform approach organizations must apply the multimodal ontology.

[3] Taxonomies lend themselves as natural starting points in explaining ontologies as they can be easily conceived as a sort of lightweight ontology

[4] 'Plug and Play', 'Federation', 'Technology Independent Services', and 'Safe, Secure and Trust'

Multimodal ontology is – and can be realized through - the alignment of existing ontologies and/or the representation of concepts and properties of existing standards[5]. Multimodal ontology can also be used to further specialize.[6]

The following figure visualizes the context of the multimodal data sharing ontology:

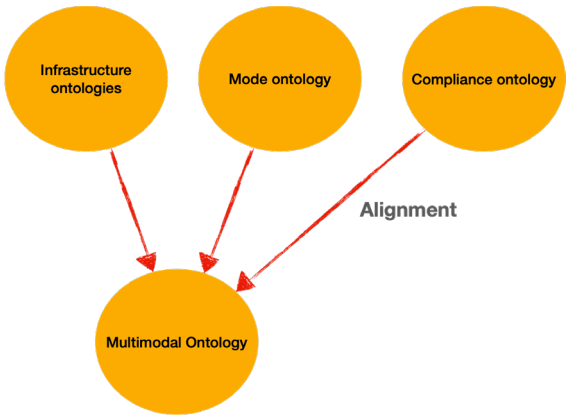

Figure 2 – the multimodal ontology: alignment of existing initiatives

Genuinely, each community (or data ecosystem or 'data space', like a port community, industry associations, infrastructure managers, and regulators) can develop its own ontology and have it aligned with the multimodal ontology. The multimodal ontology can be applied by these communities by specializing it for their members, i.e. including additional details in their ontology. In case such details can be applied in two or more communities, they may be considered to become part of the multimodal ontology. For instance, generic concepts like 'legal entity' and 'organizational roles' can be applicable across several communities.

1.2 The structure of semantics - functionality of the multimodal ontology

Since all communities will have different (implementation guides of) standards with different structures, the multimodal ontology provides an alignment framework consisting of 'Digital Twin' and 'event':

- Digital Twin is a taxonomy of real-world objects (container, truck, barge, etc.) and infrastructure.
- Event is the association between at least two Digital Twins in time and space (past, present, and future, where future is 'expected', 'planned/estimated', and 'required').

The structure of semantics relates to connectivity or rather linkage – i.e., linked data. Business and compliance choreographies. Choreography[7] enables modelling data sharing between any two stakeholders as event types. An interaction or a business document is for

[5] A standard sets out an agreed way of doing something. It's a statement of good practice, designed to make things better, safer, and more efficient. Standards can cover a huge range of organizational activities, from making a product to delivering a service or creating a process. A standard is a collective work.
[6] These specializations may become part of the multimodal ontology via a change procedure.

[7] A choreography specifies the synchronization of business processes of two stakeholders, without modelling these business processes. Each stakeholder models and implements its own business processes supporting a choreography.

instance represented by a subtype of event associating Digital Twins, locations, and organizations for a business activity. The data sharing concepts are also an ontology. Each legal or natural person has its business services (or goals) related to the data sharing concepts.

Any constraints between subtypes of Digital Twins are formulated on 'event' level and specified in SHACL (see page 2). One of the basic constraints is 'cargo', like a container that can be cargo for a trailer and a trailer that can be cargo for a ferry.

The following figure visualizes the result of decomposition, including 'legal entity' and relevant financial and compliance concepts.

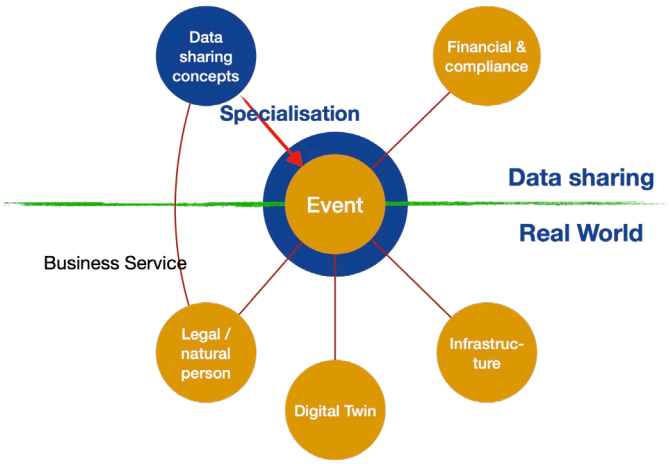

Figure 3 – the multimodal ontology with details of business data sharing concepts

2. Service registry and the multimodal ontology

A Service Registry of an organization supports its **discoverability** and its **data sharing** capabilities. Each organization thus has its own Service Registry. The Service Registry enables:

- any organization to specify its data requirements and
- any other organization to define the business services it wants to provide.

In case the data sharing concepts ontology defines the data structure of the Service Registry (blue circle around Event in figure 3), the Service Registry can be applied in two ways:

- **Design** – to specify business activities and their choreographies. Industry associations and communities can do design.
- **Configuration** – to specify an organizational profile by searching and:
  - selecting those parts of a design that are relevant to an organization,
  - specifying its business services and electing the various lower layer protocols it supports (including endpoints). Business activities and their choreography define constraints to the multimodal ontology.

  In addition, a user selects the constraints applicable for its organization i.e., selecting the relevant logistic Digital Twins applicable to its organization.

The configurations also specify **access control**: each organization is only able to provide access to data that is stored. Access can of course only be given to data of which links are shared by events.

The Service Registry must be able to select a technical protocol for data sharing like openAPIs and XSDs. There will be many deployments of the Service Registry. An interface is required to interconnect.

Discoverability is implemented at two levels, based on known SPARQL endpoints of Service Registries:

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

Any two stakeholders can share data for those parts of their organizational profile that are common. They can do business digitally (and be compliant) if goals and business services can be matched. The more stakeholders implement of a maximal design (see hereafter), the more their business can be supported digital according to plug and play.

The endpoints of Service Registries must be trusted. They are subject to Identification and Authentication.

Each Service Registry contains (or has access to) the complete multimodal ontology, including those that are aligned with that ontology, and has a discovery function for re-use of business activities and their choreographies.

The Service Registry functionality:

- The **minimal** functionality combines design and configuration in generating and publishing an openAPI with an endpoint for a single interaction, like a transport order, a business document, or a visibility event, including a connectivity protocol like CEF eDelivery over TLS (Transport Link Security).
- The **maximum** functionality is a separation of design and configuration where at design time the complete data sharing ontology is applied as input for configuration (plug and play). Data sharing is implemented with semantic technology, only a (semantic) endpoint is specified.

A first iteration Service Registry will come with at least one business activity (‘transport’) and a supporting choreography.

3. The Index

An index of an organization contains all events (with links to data) send as data holder with other organizations and received as data user from data holders. An index shares and stores events between a data holder and -user and supports a data user to formulate queries based on links received via events and share these with a data holder.

The functionality of an index:

- **The minimal** functionality is to share visibility events with no link to additional data. This is only about progress validating the quality of event data[8].
- **The maximum** functionality[9] is to support:
  - data quality validation (correctness and completeness of event data and query (results)),
  - event logic (validating the sequence of events),
  - event distribution (sharing an event with the proper data holder(s)),
  - enable access for replying to data users queries (link-based access control), and
  - query federation (data provenance).

Events with links to data are shared in a commercial – or legal relationship between any two stakeholders. Some events, like an order event, can have links to different types of data like parties involved including their role (shipper, carrier, forwarder) whereas others represent visibility of the execution of a business activity (e.g. an ETA event that links to an order event).

Each organization has its own private index that stores all events (with links to data) sent as data holder to data users and received as data user from data holders. An index supports a data user to retrieve additional data via the links it has received and a data holder to provide a query result by validating the link was shared (**link-based** access **control**) and accessing data. The latter is data that is either stored by a data holder itself or another organization (data provenance). In the latter case, a query is federated to the data source (**query federation**).

An index supports **event distribution** (sharing an event with the proper data holder(s)) based on input of a data holder initiating a commercial relation, the existing of a commercial relation (previous events are stored by an Index), or for legal compliance.

A **node** must support **data quality validation** (correctness and completeness of event data and query (results)) and either in its internal IT systems or by its index. Data quality is specified by an organization profile (see Service Registry).

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

4. Identity, Authentication, and Authorization (IAA)

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

[8] This relates to the minimal functionality of the Service Registry.

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

IAA is built upon two pillars[10]:

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

Authorization is internal to each organization and is the basis for access control. Organizations thus do not know authorized users of other organizations; they trust that authorization is properly implemented by others (organizational trust).

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

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

The implementation of such a distributed solution is still under development. The existing standards, and solutions (like OAUTH2.1) can still be applied to create inter-organizational trust (applicable to data of a Service Registry and an Index). This intermediate level requires one or multiple Identity Brokers acting as intermediate Registration Authorities. Preferably, a regulator is a public body.

5. The semantic adapter

Each organization will have its own internal IT systems, each of them with its own data structure and technology. Organizations may also have implemented existing (open or de facto) standards. The semantic adapter supports transformation between internal - and semantic data and semantic data to (open or de facto) standards. The combination of these transformation supports the transformation between internal data and standards, although there are already tools for such a data transformation.

In case an organization implements an index with semantic technology and shares semantic data (RDF, SPARQL), the semantic adapter transforms between internal data and semantic data structured according to the multimodal ontology.

The functionality of a semantic adapter:

- The **minimal** functionality is the support of a JSON file structure that reflects the structure of (a part of) the semantic model. It is up to an organization to interface with the intermediate JSON file structure. This minimal functionality may not yet support a link for querying, since that requires additional mapping functionality.

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

- The **maximal** functionality consists of an ontology alignment and mapping tool for configuring an RDF plugin for a relational database, combined with a process engine in case data is stored in different systems. All types of intermediate versions are foreseen, like the mapping of semantic queries to (open)APIs (which is complex) combined with a prescription of openAPIs for an internal IT system (e.g., an openAPI for all relevant parts of the Digital Twin taxonomy implemented by an organization).

Whenever a designer or configurator chooses to implement an interaction with a (JSON) openAPI, traditional integration technology can be used.