IPIC 2019 | 6th International Physical Internet Conference | London

# Maintaining control over sensitive data in the Physical Internet: Towards an open, service oriented, network-model for infrastructural data sovereignty

S. (Simon) Dalmolen MSc[1,2] , H.J.M. (Harrie) Bastiaansen PhD[2], E.J.J. (Erwin) Somers BSc[2]
S. (Somayeh) Djafari LL.M.[2], M. (Maarten) Kollenstart MSc[2,] M. (Matthijs) Punter MSc[2]

1. University of Twente, Enschede, The Netherlands
2. TNO, The Hague, The Netherlands

Corresponding author: S.Dalmolen@Utwente.nl

***Abstract:*** *Changing market dynamics force organizations evermore to share data in supply chains. Misuse of the data shared can cause major damage to the business and reputation of organizations. Hence, being in control over the terms-of-use for sharing data (i.e., data sovereignty) is a key prerequisite for sharing potentially sensitive data. This, however, provides a major challenge as data sovereignty concepts are currently mainly provided in communities with their own specific data sovereignty solutions. This faces data providers with a threat of lock-in and major integration efforts in case of data sharing with a multitude of data consumers. As alternative, a network-model approach for providing generic infrastructural data sovereignty can overcome these challenges. Its technical concepts are currently maturing. Its business and service concepts however are still under development. This paper proposes an open, service-oriented, network-model approach for infrastructural data sovereignty. The goal is to support a broad variety of end-user and service provider options for maintaining sovereignty in the data sharing processes. It uses an illustrative and representative logistics scenario and describes how infrastructural data sovereignty may stimulate adoption of sharing of (potentially sensitive) operational data as required for realizing the physical Internet.*

***Keywords:*** *Data Sovereignty, Network-Model, Service-Orientation, Open, Metadata, Terms-of-Use*

## 1 Introduction

Digitization is fundamentally changing supply chain collaborations, business strategies, business processes, firm capabilities, products and services (Bharadwaj et al. 2013). Organizations are increasingly working together to serve consumers through mutually dependent and co-operative supply chains. Improving the agility and flexibility of (supply) chain collaboration offers potentially major benefits but also poses major challenges, both from an organizational and a technical/IT perspective (Luftman et al. 2017).

In transitioning towards more advanced forms of supply chain collaboration, organizations are faced with a dichotomy. On the one hand, they are becoming ever more aware that data is a valuable asset in the emerging data economy and should be handled by the organizations as such (Gunasekaran et al. 2017), (Marinagi et al. 2015). On the other hand, they require trust that the organization's data is handled in a controlled and secure way as a prerequisite sine qua non the organization may not be prepared to share its data. Consequently, there is a growing need for a 'data-centric' foundation provided by an (open) infrastructure for multi-lateral sharing (Nicolaou et al. 2013), which enables organizations to be in control over the conditions and terms-of-use under which their potentially sensitive data is shared. This is referred to as data sovereignty.

For logistics companies being data providers in Physical Internet supply chains maintaining data sovereignty over their sensitive data applies to a multitude of data consumers, e.g. other logistics companies, logistics service providers, authorities. This, however, provides a major challenge as data sovereignty concepts are currently mainly provided by (closed) communities with their own specific solutions. Consequently, the data provider is faced with both a threat of consumer lock-in by their community providers and with major integration efforts on defining managing and enforcing data sovereignty requirements for a multitude of data sharing relationships with different data consumers. Hence, for such multi-lateral data sharing whilst maintaining data sovereignty over sensitive data, a single-entry point for the data provider may give clear operational advantages in agility, reduced complexity, improved efficiency and lower costs. Generic and re-usable capabilities for defining and enforcing terms-of-use based on standardized protocols may yield major benefits. An open network-model approach for infrastructural data sovereignty for multi-lateral data sharing can offer such capabilities.

The technological concepts and components for the network-model approach are currently maturing. However, this is not (yet) the case for its business and service concepts, aimed at supporting a broad variety of end-user and service provider options for infrastructural data sovereignty within the network-model approach. To overcome this lack of maturity, the research question that we address in this paper is how to design an overarching technical, service and business architecture for a network-model approach for infrastructural data sovereignty. The novelty and main contribution as put forward in this paper is on service-oriented business architecture for data sharing support processes that allows data providers to maintain sovereignty over the sensitive metadata that is generated and managed in a network-model approach for infrastructural data sovereignty. The approach as proposed in this paper can contribute to (internationally accepted and standardized) development and deployment of such a network-model for infrastructural data sovereignty, which is key for wide-scale adoption.

The structure of this paper is as follows. Section 2 provides a representative logistics scenario illustrating the growing need for infrastructural data sovereignty functionality. Subsequently, the topic of data sovereignty as key prerequisite for sharing sensitive data is described in a generic and implementation-independent manner in section 3. The following section 4 addresses the benefits and potential of a network-model approach for infrastructural data sovereignty and presents current initiatives working towards that goal. A service and technical architecture for an open, service-oriented, network-model for infrastructural data sovereignty for multi-lateral sharing of sensitive data is elaborated in section 5 and section 6, in which the former addresses the service-oriented network-model architecture approach whereas the latter more specifically elaborates the service portfolio to be provided. The final section 7 and section 8 provide a discussion and the conclusions, respectively.

## 2 Illustrative logistics collaboration scenarios

To illustrate the growing need for infrastructural data sovereignty functionality, this section provides a representative logistics case on the minimization of the number of transport movements, governed by internal (business) and external (regulatory) policies. For transporters, minimization of number of transport movements may lead to efficiency and cost optimization. For society, potential benefits are in lower CO2 emission, less traffic jams and higher safety in traffic-intensive areas.

Minimization of the number of transport movements requires supply chain collaboration. Two fundamental business capabilities to be supported in such supply chain collaboration (Dalmolen et al. 2012) are:

- *Supply Chain Composition,* also referred to as 'goal matching', in which shippers, Logistic Service Providers (LSPs) and transporters match demand for and supply of transport capacity.
- *Supply Chain Visibility,* also referred to as 'situational awareness', in which overview is provided over the full supply chain and context, e.g. through track and trace functions for shippers on the status and location of the loadings under transport.

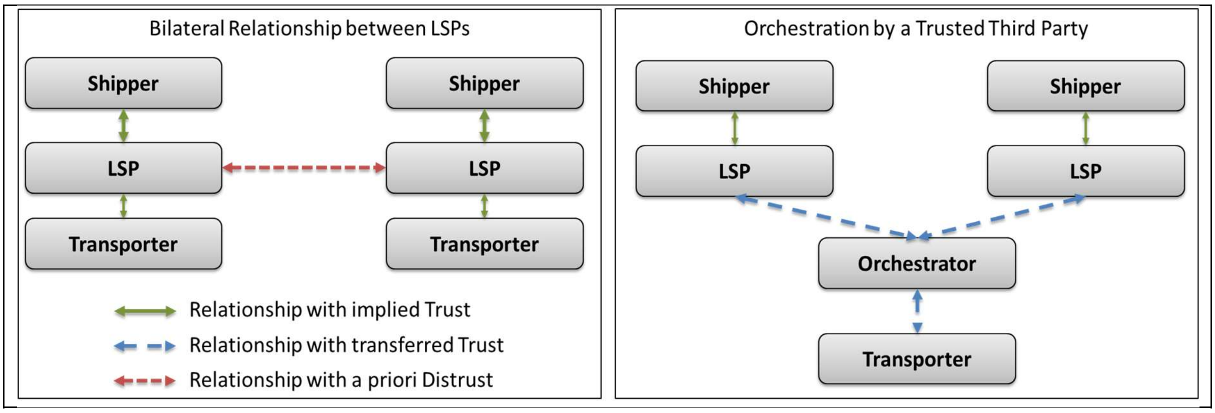

Bilateral Relationship between LSPs
Shipper
Shipper
LSP
LSP
Transporter
Transporter
Relationship with implied Trust
Relationship with transferred Trust
Relationship with a priori Distrust
Orchestration by a Trusted Third Party
Shipper
Shipper
LSP
LSP
Orchestrator
Transporter

*Figure 1: Trust Relationships for Typical Collaboration Scenarios: Bilateral Relationships between LSPs (l) and Orchestration by a Trusted Third Party (r).*

Realization of these business capabilities may be achieved by typical collaboration scenarios, with their own type of trust relationships, as illustrated in Figure 1 for:

- *Bilateral relationship between LSPs.*

  This collaboration scenario requires the sharing between LSPs of (potentially sensitive) business data, e.g. on current and scheduled transport capacity. To protect this (potentially sensitive) data, the providing LSP will impose strict terms-of-use (in the form of (enforceable) access and usage control policies) for the shared data on the data consuming LSP.

  This collaboration scenario may for instance be applicable to long-distance road transport, e.g. between cities and/or internationally, by transporters under free and competitive market conditions, with cost efficiency as a major key for success. Minimization of the number of transport movements is the responsibility of the transporters themselves.

- *Orchestration by a Trusted Third Party.*

  This collaboration scenario involves an independent and trustworthy intermediary orchestrator role that operates a (fair and independent). An intermediary orchestration role may either be imposed by regulations aimed at the societal needs for minimized transport movements or be established by (competitive) transporters for jointly optimizing their shared business goals on (e.g.) efficiency and environmental sustainability.

  This collaboration scenario may for instance be applicable to local transport, e.g. for urban area parcel deliveries and/or inner cities shop supplies. Under influence of the explosive growth of number of parcel deliveries due to online shopping, there is an increasing demand by society to minimize the number of transport movements in urban areas and inner cities.

Figure 1 also shows several types of trust relationships for data sharing between the supply chain participants in the collaboration scenarios, which are respectively referred to as:

- *Data sharing relationship with implied Trust:* Data is (for instance) shared between stakeholders that are not direct competitors, but rather are supply chain partners with mutual benefit for collaborating. There is no direct motivation for / threat of misuse of the (potentially sensitive) data provided.
- *Data sharing relationship with transferred Trust:* Data is (for instance) shared with a supply chain partner, that also has a data sharing relationship with a possible competitor for which it must be prevented that he gets access to the (potentially sensitive) data provided. Hence, trust is to be established that sufficient mitigation measures are applied to prevent the supply chain partner from sharing the (potentially sensitive) data with the possible competitor.
- *Data sharing relationship with a priori Distrust*: Data is (for instance) directly shared with a possible competitor with the mutual goal to optimize the operational processes. There is a relationship of a priori distrust between the possible competitors. Hence, strong mitigation measures are required to enforce the required trust levels prior to sharing data.

The required capabilities for maintaining data sovereignty for the various types of trust relationships in the typical collaboration scenarios will differ. Moreover, they are currently mostly provided within closed communities, with their own specific solutions. This is referred to as the ‘hub-model’ approach. As described in the introduction, the hub-model faces data providers with a threat of consumer lock-in and major integration efforts in case of data sharing with multiple data consumers. An open network-model approach for infrastructural data sovereignty provides an attractive alternative. Its design principles, a service-oriented approach and the International Data Spaces (IDS) initiative as proponents of the network-model approach are described in the following sections.

# 3 Data sovereignty: key enabler for sharing sensitive data

Data sovereignty can formally be defined as *‘a natural person’s or corporate entity’s capability of being entirely self-determined with regard to its data’* (Otto et al. 2019), i.e. allowing a legal person to exclusively and sovereignly decide concerning the usage of data as an economic asset.

## 3.1 Data sovereignty over both primary data and metadata

Clearly, maintaining sovereignty by the data provider applies to the primary, potentially sensitive, data that is shared between data provider and data consumer. However, maintaining sovereignty also applies to the secondary, derived, information on the data sharing transactions, referred to as ‘metadata’. The metadata for data sharing stems from the required support processes for managing data sharing agreements and transactions at the various stages of their life-cycle. The goal of the support processes for data sharing is to prevent misuse of the data shared by a data provider. They include the processes for the data provider and consumers to comply with internal policies (e.g. on terms-of-use, access and usage control) and with external policies (e.g. on regulations). Table 1 lists the main data sharing support processes, categorized according to the subsequent life-cycle stages for data sharing agreements and the associated data sharing transactions, together with the main metadata artefacts generated by these support processes for data sharing.

*Table 1: Support processes for data sharing and metadata artefacts.*

| Support processes for data sharing | Metadata artefacts |
|---|---|
| **Definition and exposure of an available data set**<br>• Definition and publication of a data set<br>• Definition of a data sharing profile<br>• Publication of a data sharing profile | • Data descriptor<br>• Data transaction<br>• Data request<br>• Data response<br>• Data sharing agreement<br>• Access control policy<br>• Usage control policy<br>• Security profile policy<br>• Service level<br>• Terms-of-use<br>• Commercial conditions<br>• Juridical conditions<br>• Contractual conditions |
| **Making a data sharing agreement.**<br>• Definition of terms-of-use, incl. usage and access control policies<br>• Definition of the commercial and juridical conditions<br>• Negotiation, acceptance and signing of a data sharing agreement | |
| **Performing a data sharing transaction.**<br>• Clearing of data sharing transactions, including non-repudiation<br>• Data sharing, including binding of the transactions to an agreement<br>• Settlement and discharging of data sharing transaction | |
| **Logging, provenance and reporting.**<br>• Logging and binding of data transactions to data sharing agreements<br>• Tracking, monitoring and reporting of data transactions to stakeholders<br>• Auditing, billing and conflict resolution | |

The data sharing support processes as listed in Table 1 require and generate metadata. On the one hand, the data sharing agreements are metadata in themselves. On the other hand, the management, control and administration processes over their associated data sharing transactions are a major source of metadata. These metadata artefacts as generated by the support processes are listed in the right column of Table 1.

## 3.2 Data sovereignty maintaining capabilities

Maintaining data sovereignty and preventing misuse of shared data implies providing a data provider with the enabling capabilities to be in control over who is allowed access to his data, for which purposes and under which usage control conditions, i.e. the terms-of-use, in compliance with their internal (business) policies and with external policies (e.g. on regulations), and consisting of:

- *Procedural data sovereignty maintaining capabilities:* these include administrative capabilities such as data sharing agreements (terms-of-use and conditions), certification and attestation, logging and data provenance, reporting and accountability.
- *Technical data sovereignty maintaining capabilities:* these include technical capabilities such as peer-to-peer data sharing, encryption and key management for data in transfer and in storage, sandboxing and containerization and policy-based admission control (Yavatkar et al. 1999) and enforcement and blockchains.

The procedural and technical data sovereignty enabling capabilities are closely related to the concepts of legal enforceability and technical enforceability of data sharing agreements, respectively. Legal enforceability ensures that by means of automation generated digital data sharing agreements and their associated data sharing transactions are correct and acceptable in legal procedures. Technical enforceability ensures for the data provider that the agreed-upon conditions under which data is shared are (securely) implemented in the open infrastructure for multi-lateral data sharing.

The pivotal concept in both the procedural and technical data sovereignty maintaining capabilities are the terms-of-use. These are expressed as a combination of applicable access control policies and usage control policies. Usage control is a generalization of access control that also addresses how data is used after it is released. Table 2 provides a list examples of (classes) of access and usage restrictions.

*Table 2: Examples of (classes) of access and usage restrictions.*

| **Access control restrictions (access control policy)**<br>*Stating which individuals, roles or systems are allowed access to the data provided.* | **Usage control restrictions (usage control policy)**<br>*Stating (limitations on) how data may be used after it has been shared.* |
|---|---|
| • Provide or restrict data access to specific users<br>• Provide or restrict data access for specific systems<br>• Allow access to data<br>• Inhibit access to data | • Provide or restrict data access for specific purposes<br>• Delete data after X days/months<br>• Use data not more than N times<br>• Use data in a specific time interval<br>• Log data access information<br>• Share data only if it is encrypted<br>• Control printing shared data |

## 4 Infrastructural data sovereignty: the network-model approach

The capabilities for maintaining data sovereignty (as described in a generic, implementation-independent, manner in the previous section) are currently mostly provided by (closed) communities for trusted data sharing, with their own specific solutions. This is referred to as the 'hub-model' approach and is commonly applied for sector specific, closed, communities. As described in the introduction, it faces data providers with a threat of consumer lock-in and major integration efforts in case of data sharing with multiple data consumers. This section describes the open network-model approach for infrastructural data sovereignty, as opposed to the solution specific hub-model approach.

### 4.1 An open network-model for infrastructural data sovereignty

Figure 2 illustrates the transition from a solution specific hub-model approach towards an open network-model approach for infrastructural data sovereignty (Liezenberg et al. 2018).

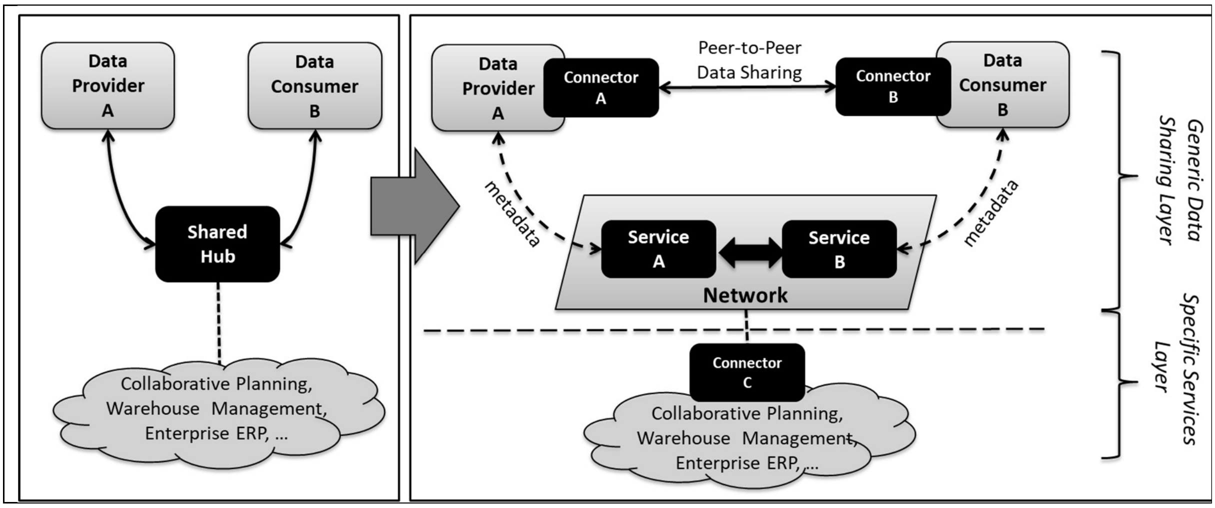

Data Provider A
Data Consumer B
Shared Hub
Collaborative Planning, Warehouse Management, Enterprise ERP, ...
Data Provider A
Connector A
Peer-to-Peer Data Sharing
Connector B
Data Consumer B
metadata
metadata
Service A
Service B
Network
Connector C
Collaborative Planning, Warehouse Management, Enterprise ERP, ...
Generic Data Sharing Layer
Specific Services Layer

*Figure 2: Transition from a Hub-Model Approach (l) to an Open Network-Model Approach (r) for Data Sharing (Liezenberg et al. 2018).*

The upper part of Figure 2 depicts the ‘Generic Data Sharing Layer’ providing infrastructural data sovereignty capabilities, the lower part of Figure 2 depicts the ‘Specific Services Layer’ in which a multitude of specific value adding services can be supported.

The network-model approach is currently attracting major attention in overcoming the challenges associated to the hub-model. It provides generic infrastructural data sovereignty capabilities, enabling a single-entry point for the data provider with common and agreed upon protocols for defining and enforcing terms-of-use for data sharing. A network-model approach has previously been successfully developed and realized for infrastructural service provisioning in the banking and telecommunications sector.

For trusted data sharing using the network-model approach, the right part of Figure 2 shows three main leading architectural principles that are currently gaining acceptance for maintaining data sovereignty:

- *Peer-to-Peer data sharing.* To maintain data sovereignty by the data provider, local data processing is used in combination with peer-to-peer sharing of potentially sensitive data between a provider and consumer. For sharing the data, it is not stored in a centralized data base or controlled, forwarded or processed by an intermediary organization. As such, it prevents data providers from having to rely on intermediate external (trusted) organizations and from relinquishing full self-control over their potentially sensitive (meta)data to be shared.
- *Distributed infrastructure for support services.* Peer-to-peer data sharing as described in the previous bullet point has to be enabled by the support processes as listed in Table 1. These support processes will have to be implemented in a highly distributed infrastructure. In this infrastructure data providers and data consumers are subscribed to their own set of intermediary service providers providing their own portfolio of data sharing support services.
- *Openness for wide-scale adoption.* This network-model approach should be open to enable wide scale adoption and lower the barriers to participate. It has to be noted that for the various stakeholders in the distributed infrastructure ‘openness’ has its specific meaning (National Research Committee 1994):
  - *Open to end-users:* it does not force end-users into closed groups or deny access to any sectors of society but permits universal connectivity. This is also referred to as creating a ‘level playing field’.
  - *Open to solution providers:* it allows any solution provider to meet the requirements to provide enabling components in the distributed and open data sharing infrastructure under competitive conditions.
  - *Open to service providers and to innovation:* it provides an open and accessible environment for service providers to join and for new applications and services to be introduced.

In the network-model approach, data sharing is done on a peer-to-peer basis according to the first leading architectural principles. Nevertheless, this peer-to-peer data sharing may be used to populate a centralized data lake as part of a value adding service in the specific services layer, e.g. for logistics collaborative planning, warehouse management or enterprise ERP, as depicted in Figure 2. This may seem contradictory and may seem to make the generic data sharing layer in the upper part superfluous. It is noted, however, that also in these cases there is added value in the generic data sharing layer: (1) in the aligned and standardized mechanisms of communicating from data provider to service provider the terms-of-use under which the data is shared, (2) in the enforcement thereof in the domain of the service provider, and (3) in the added value of providing supporting functions for data sharing by external trusted roles as independent party.

In the distributed, business architecture of the open network-model, multiple and independent participants provide and govern their own services and solutions (Heikkilä et al. 2008), (Nicolaou et al. 2013). Nevertheless, they will have to be seamlessly interoperable in realizing and providing the overarching trust and data sovereignty enabling capabilities. To enable wide-scale adoption with low barriers to participate, they have a joint interest in defining and adhering to an agreed-upon reference architecture, ensuring the specific functions and business interests of each participant and supported by well-defined standards for interoperability. Such an open, service-oriented, business architecture for an open network-model approach will avoid strong monolithic implementations and prevent 'lock-in', by service providers. The following subsection describes current initiatives pursuing such an open, service-oriented, business architecture.

## 4.2 Initiatives on the open network-model for infrastructural data sovereignty

The open network model for maintaining data sovereignty is currently gaining major interest. This is reflected in both policy making and infrastructure development initiatives as listed in Table 3.

*Table 3: Overview of policy making and infrastructure development initiatives.*

| **National and European policy making initiatives** |
|---|
| **ETP ALICE:** European Technology Platform 'Alliance for Logistics Innovation through Collaboration in Europe'<br>ETP ALICE assists the implementation of the EU Horizon 2020 research program. It is based on the need for an overarching view on logistics and supply chain planning and control for efficient logistics and supply chain operations. Its Systems & Technologies for Interconnected Logistics research and innovation roadmap identifies the need for new business models and data governance approaches with collaboration to enable trust and data sovereignty (ALICE 2018). |
| **DTLF**: Digital Transport & Logistics Forum<br>The DTLF is a group of experts that brings together stakeholders from different transport and logistics communities, with a view to build a common vision and roadmap for digital transport and logistics. In its report (DTLF 2018), the DTLF Subgroup 2 'Corridor Information Systems', identifies the drivers for creating a common data sharing commodity and outlines the basic supporting roles for supporting such a common data sharing commodity. |
| **National initiatives**<br>For instance, to support data sharing in and over economic sectors and society, the Dutch government has recently released several direction setting policies (Dutch Ministry of EA&CP 2018), (Dutch Ministry of EA&CP 2019) in which the economic value of data sharing is outlined with the importance of an adequate data sharing infrastructure as a key enabler. |
| **Infrastructure development initiatives on an open, network-model approach.** |
| **iShare**<br>The Dutch iShare initiative for the logistics sector realizes a uniform set of agreements for identification, authentication and authorization, such that organizations can share logistics data in a simple and controlled way, including with new and previously unknown partners (NLIP 2019). |
| **AMDEX** - Amsterdam Data Exchange<br>AMDEX is an initiative of the Amsterdam Economic Board to facilitate, local, European or international cooperation in a transparent open data market (Amsterdam Economic Board 2019). It |

will offer infrastructure and common rules to secure a trusted and safe environment that interested partners can join to create platforms for real-time data-driven cooperation.

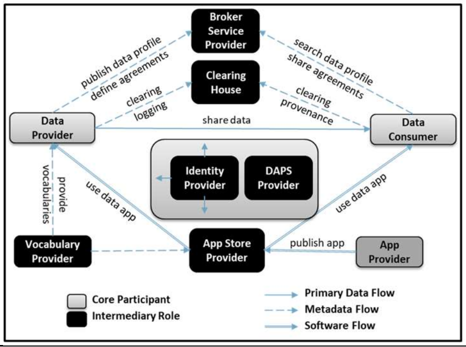
**FIWARE** – Future Internet WARE

FIWARE is a framework of open source platform components providing a set of APIs for the development of (smart) applications in multiple vertical sectors. An open source reference implementation of each of the FIWARE components is publicly available for fast and low-cost deployment. FIWARE was funded by the Seventh Framework programme of the European Community for research and technological development.

In the following subsection, the IDS initiative as listed in the table will be further described.

## 4.3 The International Data Spaces (IDS) initiative

IDS is currently gaining major international traction for realizing an open infrastructure for trusted, multi-lateral, data sharing. The IDS reference architecture (Otto et al. 2019) is aimed at enabling the trusted sharing of (potentially sensitive) data, whilst maintaining sovereignty. It can be considered an architectural elaboration of the Trusted Multi-Tenant Infrastructure (Trusted Computing Group 2013). Figure 3 depicts and describes the main roles as distinguished in IDS.



| IDS Intermediary Trusted Roles | |
|---|---|
| **Role** | **Function** |
| **Identity Provider** | - Manage Identities of Participants<br>- Validate Identities of Participants |
| **DAPS Provider** | - Provide attribute information on participants and connectors<br>- Manage and register security profiles |
| **Clearing House** | - Clearing and settlement of data transaction<br>- Logging of data agreements and transactions<br>- Auditing, reporting and conflict resolution |
| **Broker Service Provider** | - Registration and publication of data profiles<br>- Definition, negotiation and signing of data sharing agreements |
| **App Store Provider** | - Publication and provoisioning of data apps |
| **App Provider** | - Development and maintenance of data apps |
| **Vocabulary Provider** | - Managing and offering of (domain specific) vocabularies |

*Figure 3: Roles in the IDS Reference Architecture (l) (Otto et al. 2019), together with a Functional Description for the Intermediary Roles (r) (Dalmolen et al. 2018).*

The 'Intermediary Roles' in the IDS reference architecture act as trusted entities provided by trusted third parties (TTPs). In addition to the roles as depicted in the figure, the IDS distinguish roles for providing certification and (remote) attestation functions. IDS adhere to the leading architectural principles as described in subsection 4.1.

# 5 Service-oriented infrastructural data sovereignty

Maintaining sovereignty by the data provider over the metadata associated to his data sharing activities gives rise to operational challenges. An area of tension exists. On the one hand the stringent data sovereignty requirements ask organizations to keep the control over this metadata by storing and processing it as much as possible within their own (security and trust) domain. On the other hand the manageability and cost-efficiency thereof tend organizations to transfer the management and storage of metadata to external and specialized organizations such as an (IDS) Identity Provider, Broker Service Provider and Clearing House. As such, service-oriented infrastructural data sovereignty addresses the topic designing and managing the data support processes and their associated metadata in Table 1.

As described in the previous section, an open, service-oriented network-model provides major advantages to support the large variety of 'intermediate' architectural options that may be commercially and technically viable between the extremes of on the one hand full self-control and on the other hand outsourcing of the supporting data processes and their associated metadata. This section describes how interaction patterns between the various roles in such an open network-model for multi-lateral sharing of sensitive data may be realized.

## 5.1 Metadata interaction patterns for infrastructural data sovereignty

Figure 4 is an elaboration of the network-model approach depicted in Figure 3. It shows multiple instances of the intermediary roles that will coexist with a data provider and data consumer in general being subscribed to different instances, i.e. their ‘home’ intermediary roles.

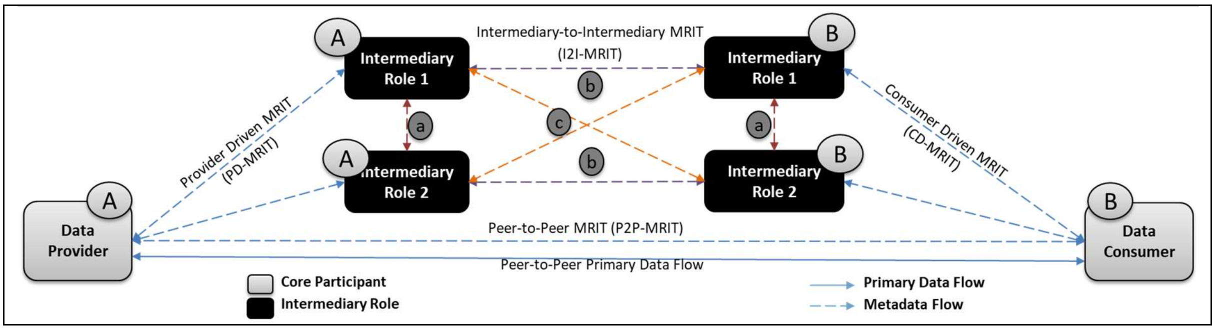


*Figure 4: Metadata Interaction Patterns within the Generic Data Sharing Layer in a Distributed Open Network-Model Approach.*

The figure shows the various metadata interaction patterns within the generic data sharing layer, distinguishing:

- *Provider and consumer driven metadata interaction patterns,* in which the data provider and consumer orchestrate the sharing of metadata with the intermediary roles they have subscribed to.
- *Intermediary-to-Intermediary metadata interaction patterns,* in which the intermediary roles of the various data providers and consumers orchestrate the sharing of metadata amongst them.

The suitability of both types of metadata interaction patterns for sharing metadata between roles in the distributed open infrastructure for multilateral data sharing is evaluated on the following criteria:

- *Maintaining sovereignty over the metadata by the data provider and consumer*

  Maintaining sovereignty over their metadata makes it essential for data providers to be in control over the proliferation chain of his metadata. Proliferation along a chain of interconnected intermediary roles by means of Intermediary-to-Intermediary metadata interaction patterns implies loss of such control and having to trust and rely on intermediary roles that are potentially not even known to the data provider.

- *Complexity of the overarching interoperability architecture*

  The widescale adoption of agreed-upon (and preferably standardized) interaction protocols strongly depends on the complexity and number of standardized interconnections to be realized by the various intermediary roles in the overarching role model. Having to implement and adhere to standardized interaction protocols for a multitude of types and instances of intermediary-to-intermediary metadata interaction patterns may be (too) complex, both from the development and deployment perspective.

  It is to be noted, that this complexity may be technically overcome as has been demonstrated in the ‘old-school’ world of pre-divestiture telecommunications at the end of the previous millennium. In their regulated environment, a limited number of (mostly non-competitive) major telcoes had a common interest in closely collaborating in developing standards for interoperability to achieve globally interoperable services. In the current liberalized situation for data services however, such a centrally governed development and deployment process is non-existent. Hence, definition and adoption of agreed-upon intermediary-to-intermediary interoperability protocols are a far less viable option.

On these criteria, the provider driven, and consumer driven metadata interaction patterns are to be preferred over the intermediary-to-intermediary interaction patterns. They give the data provider and consumer with the required control for maintaining sovereignty over their metadata. No direct intermediary-to-intermediary metadata sharing beyond the direct control of the data provider and consumer are required, preventing them from having to rely on trusted third parties.

The following subsection describes how the preferred provider driven and consumer driven metadata interaction patterns may be realized in an open, distributed, architecture for multi-lateral data sharing.

## 5.2 Policy enforcement framework for data sovereignty

The combination of the procedural and technical data sovereignty enabling capabilities (as described in subsection 2.1) constitute a data sovereignty framework for the supporting life-cycle processes for data sharing. They will have to be technically implemented by means of a data sharing connector as shown in Figure 5.

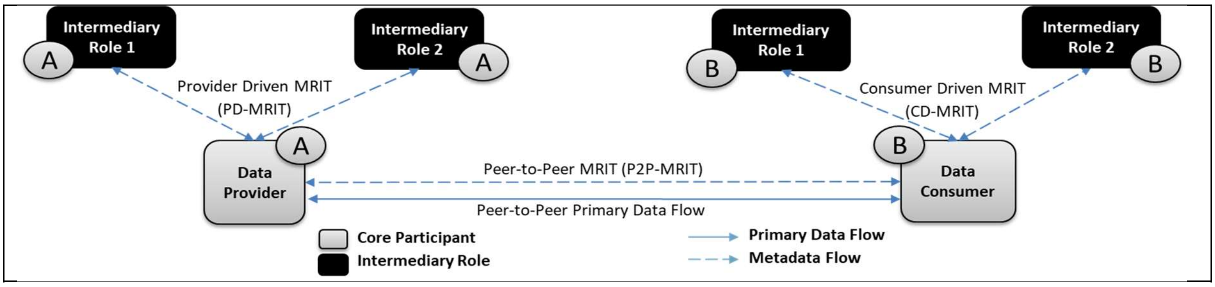


*Figure 5: Runtime Environment for Metadata Flow Control based on Data Sharing Connectors.*

As figure 5 shows, a data sharing connector consists of an app execution environment in combination with a policy execution framework:

- The *App Execution Environment (AEE)* runs a set of containerized apps of which the input and output data flows are being controlled by the associated PEF. These could be the apps of the intermediary roles. Typically, the apps in the AEE provide the procedural data sovereignty capabilities for legal enforceability.
- The *Policy Execution Framework (PEF)* includes capabilities for technical enforceability of the agreed-upon terms-of-use, access control policies and usage control policies and collaboration of the PEF-instances in the connectors of the local and remote data sharing endpoints. Typically, the PEF provides the technical data sovereignty capabilities for technical enforceability.

For the IDS reference architecture as described in the previous section, the data sharing connector is referred to as an ‘IDS connector’, currently being standardized under the terminology of ‘Security Gateway’ (DIN SPEC 27070). It consists of an execution core container, with the AEE and PEF, that is able to retrieve certified data apps from intermediary roles from an app store. The execution core container has a data router for routing incoming and outgoing messages through the correct data apps. Furthermore, it is enabled to enforce access and usage control policies.

## 6 Service elaboration for infrastructural data sovereignty

As described in the previous sections, an open service-oriented business architecture provides major advantages to support the large variety of 'intermediate' architectural options that may be commercially and technically viable. The following subsections describe how this can be realized for the basic main functions of processing and logging of sensitive (meta)data.

### 6.1 Services for processing of sensitive metadata

Utilizing the capability of the AEE in the runtime environment of the data sharing connector (as depicted in Figure 5) enables the intermediary roles to provide their supporting services for the data sharing support processes as listed in Table 1 by means of apps executing locally within a secure, containerized, connector. This way the data provider maintains sovereignty over the associated metadata as it does not (have to) leave the local data provider's or data consumer's connector.

An illustrative and representative use case entails the supporting subprocess on the definition of terms-of-use, including the usage and access control policies as provided by an intermediary and trusted broker service provider role. A main added value and distinguishing factor for a specific broker service provider can be in minimizing the complexity for defining and configuring the applicable terms-of-use for their subscribed data providers, thus minimizing the required skills and IT-savviness for the data provider, lowering the barriers of adoption. In this scenario, the broker service provider offers a set of terms-of-use templates to be used by its subscribed data providers. The quality and ease-of-use of the templates will be a main competitive advantage. The templates are provided by means of a data brokering app running in the data provider's trusted data sharing connector. This data brokering app fulfils the role of the delegated data brokering service (including negotiation and signing), executing locally in the data provider's trusted data sharing connector, i.e. within the data providers domain and under control of its local policy enforcement framework. As part of the delegated data brokering app installation and configuration process, its associated access and usage control policies are provided and instantiated within the data provider's policy enforcement framework, preventing from misuse or data leakage of the associated metadata.

### 6.2 Services for logging of sensitive metadata

For the supporting life-cycle sub-processes for 'logging, provenance and reporting', a broad variety of options may be supported in an open, distributed, architecture. Such supporting sub-processes are typically enabled by services provided by a clearing house intermediary role. Similar as for the illustrative and representative use case described in the previous paragraph, the service of the clearing house intermediary role may be provided by means of an app of the clearing housed executing locally within the secure data sharing connector of the data provider or consumer. This approach enables various service alternatives to be supported with respect to locally (i.e. within the domain of the data provider or data consumer) versus centrally (i.e. within the domain of the clearing house) logging of metadata:

- *No centrally logging of metadata.* This reflects the strictest approach to maintaining data sovereignty in which the data provider keeps the data sharing support processes and associated metadata under his own full-control and within his own (security and trust) domain.
- *Centrally logging of hashed metadata.* In this approach, the data provider keeps the data transaction metadata within his own (security and trust) domain, whilst providing hashed metadata to an intermediary role for logging, i.e. his subscribed clearing house. In case

of conflict resolution, the clearing house can act as trusted third party by verifying the validity of the logged data by the data provider by means of the hashes.

- *Centrally logging of encrypted metadata.* In this case, the data providers metadata does not log metadata in his own (security and trust) domain. The metadata is logged by his subscribed trusted third-party clearing house, preferably in an encrypted format. Management of the encryption keys may remain under control of the data provider.

Illustrative examples for which these various forms of logging of metadata apply are e.g. for the data provider for logging metadata on the data provided for the case of conflict resolution, and for the data consumer for logging metadata and data provenance to report on compliance to the agreed upon terms-of-use.

## 7 Discussion

Data is crucial for companies and their daily operations, as well as for the longer term strategies. The data sovereign is important to be in operation in the future. Sharing data is essential to achieve operational excellence. These two opposing forces therefore make it challenging. In Chapter 2 a logistic example is given with different variants of trust between the partners themselves.

Especially in the orchestration scenario, in which the trusted third party plays a crucial role, data sovereignty is essential. Currently you see in daily practice that a port community system fills in this functionality. Unfortunately, the shipper is not often in control of his own data. An additional side effect that you often find in practice is a vendor lock-in in terms of software and business functionality. The current proposal can address the challenges described above.

To implement these business scenarios forces the stakeholders (shippers, LSP's, transporters) to share sensitive data in the logistic value chain, requiring a trusted multi-lateral data sharing infrastructure to spur their willingness to share such data. As such, they give rise to new challenges: (1) on compliance to internal business policies for trusted data sharing with stakeholders that could potentially be competitors and compliance to external regulatory policies, (2) on privacy regulations such as General Data Protection Regulation (GDPR) and competition laws, as applicable to this specific scenario. The use of data sharing agreements that are enforceable, both legally and technically (as described in the previous section), may provide the means to the stakeholders to gain the required level of trust for being willing to share their (potentially sensitive) data in in support of these business scenarios.

Implementation of these business scenarios requires that shippers, LSP's, transporters and other service providers in the logistic value chain share (potentially sensitive) business and operations data. As such, they give rise to new challenges:

- *Compliance to internal business policies for trusted data sharing:* to reap the indicated benefits of exchanging data, operational data which may be valuable and business-sensitive has to be shared with stakeholders that could potentially be competitors. A trustworthy infrastructure based on solid agreements and contracts and a technical secure data sharing infrastructure are a prerequisite for convincing stakeholders to exchange such data, i.e. an interoperable, multi-lateral, trusted data sharing infrastructure.
- *Compliance to external regulatory policies:* to share data, different regulations are introduced by European law makers. Notwithstanding the inherent complex role of data and algorithms, an increased understanding is needed about how data regulation should be applied in case of data platforms. The following high-level challenges have been addressed in the literature. For example, (BDVA 2019) emphasizes from practical point of view that there are questions on how "to incorporate and adjust for the effects of the

regulatory landscape in data market e.g. how to be compliant, when, where and which regulation comes into effect, how to gather knowledge on implementing the regulation etc.". Another challenge that has been addressed by law scholar and the European Commission's future approach towards sharing data in competition policies. They expect that Commission's approach will likely be speedier enforcement through complementary regulatory measures and adjusted rules in merger control and online vertical restrictions.

## 8 Conclusions

A primary objective of this paper has been to describe the need and architectural approach for infrastructural data sovereignty for multi-lateral sharing of sensitive data in an open network-model. A technical and service perspective has been proposed for transferring (outsourcing) data sharing support processes and their associated metadata to external, trusted, and specialized organizations. The expectation is that the concepts as described in this paper will improve the data provider's sovereignty and control over both their sensitive primary and secondary metadata, in a world that is ever more realizing that data is a valuable asset to be protected and exploited. It may lower the barriers for organizations for sharing their data in the transition towards a data-centric global information society.

As the technical components of the data sharing concepts as described in this paper become more and more available, adequate governance needs major attention to stimulate wide scale adoption and prevent from a lack of uptake. This applies to both governance of its development and deployment. Openness and interoperability through standardization are major enablers for success.

Standardization must focus on interoperability of the data provider and consumer with the supporting intermediary roles in an open network-model. At the same time, to optimally support service-orientation for infrastructural data sovereignty, standardization should not be (too) prescriptive with respect to various service options that may be provided by these intermediary roles. As such, conforming to the architectural considerations as described in this paper, standardization should focus on and be limited to standardization of the (information models for the) metadata artefacts and the interaction messages and protocols for conveying them between the various roles in the open network-model approach. It is to be noted that the main concepts of the IDS architecture and their interoperability requirements as described in this paper are currently being standardized as DIN SPEC standards, (DIN SPEC 27070), (DIN SPEC 16593-1).

Leaving the uptake to individual commercial users or sectors may not be an adequate approach to wide-scale adoption, as it may not be contributing to their core business, vision and ambition. Public-private cooperation may provide a better option to success. Support by governments and authorities in jointly developing the data sharing architecture into a broadly available public utility may be envisioned, supported by adequate commercial implementations and marketing power to develop, deploy and exploit the open infrastructure, e.g. by independent service providers or telecommunication operators.

## 9 Acknowledgement

The work as presented in this paper has been supported and co-financed by the Dutch Top consortia for Knowledge and Innovation 'Institute for Advanced Logistics' (TKI Dinalog, www.dinalog.nl) of the Ministry of Economy and Environment in The Netherlands.

In addition, this paper builds upon the work done within the Dutch NWO Research project 'Data Logistics for Logistics Data' (DL4LD, www.dl4ld.net), supported by TKI Dinalog and the Dutch Commit-to-Data initiative (https://commit2data.nl/).

## References

- ALICE (2018): *The Digital Transport and Logistics Forum - Research & Innovation Roadmap*.
- AMDEX (2019): *Amsterdam Data Exchange.* https://www.amsterdameconomicboard.com/ initiatief/amdex.
- Bharadwaj A., O.A. El Sawy, P.A. Pavlou, N. Venkatraman (2013): *Digital Business Strategy: Toward a next Generation of Insights.* MIS Quarterly (37:2), pp. 471–482.
- Otto B., S. Steinbuß, A. Teuscher, S. Lohmann (2019): International Data Spaces: Reference Architecture Model Version 3. (https://www.internationaldataspaces.org/wp-content/uploads/2019/03/IDS-Reference-Architecture-Model-3.0.pdf).
- Liezenberg C., D. Lycklama, S. Nijland (2018). *Everyting transaction*. LannooCampus.
- National Research Council (1994): Realizing the Information Future: The Internet and Beyond. National Academies Press.
- Dalmolen S., H.J.M. Bastiaansen, H. Moonen, W. Hofman, M. Punter, E. Cornelisse (2018): *Trust in a Multi-Tenant, Logistics, Data Sharing Infrastructure: Opportunities for Blockchain Technology.* 5th International Physical Internet Conference (IPIC) 2018, University of Groningen, pp. 299–309.
- Dalmolen S., E. Cornelisse, A. Stoter, W. Hofman, H.J.M. Bastiaansen, M. Punter, F. Knoors (2012): *Improving Sustainability through Intelligent Cargo and Adaptive Decision Making.* E-Freight Conference. (http://www.efreightconference.com/uploadfiles/papers/efreight2012_submission_25.pdf)
- DIN SPEC 16593-1. *RM-SA - Reference Model for Industry 4.0 Service Architectures - Part 1: Basic Concepts of an Interaction-Based Architecture.*
- DIN SPEC 27070: *Reference Architecture for a Security Gateway for Sharing Industry Data and Services.*
- Digital Transport and Logistics Forum DTLF Subgroup 2 - Corridor Information Systems (2018): *Enabling organisations to reap the benefts of data sharing in logistics and supply chain.* http://www.dtlf.eu/sites/default/files/public/uploads/fields/page/field_file/executive_summary2_reading__0.pdf.
- Gunasekaran A., T. Papadopoulos, R. Dubey, S.F. Wamba, S. J., Childe, B. Hazen, S. Akter (2017): *Big Data and Predictive Analytics for Supply Chain and Organizational Performance.* Journal of Business Research (70), pp. 308–317.
- Heikkilä J., M. Heikkilä, S. Pekkola, (2008): *Coordinating and Boundary Spanning Roles of Business Networks.* In Vervest, P., Van Heck, E. & Preiss, K.,(Eds.): Smart Business Networks a New Business Paradigm.
- Dutch Neutral Logistics Information Platform – NLIP (2019). *iShare Data Sharing Initiative.* https://Www.Ishareworks.Org/En/.
- Luftman J., K. Lyytinen, T. ben Zvi, (2017): *Enhancing the Measurement of Information Technology (IT) Business Alignment and Its Influence on Company Performance.* Journal of Information Technology (32:1), pp. 26–46.
- Marinagi C., P. Trivellas, P. Reklitis, (2015): *Information Quality and Supply Chain Performance: The Mediating Role of Information Sharing*. Procedia-Social and Behavioral Sciences (175), pp. 473–479.
- Dutch Ministry of Economic Affairs and Climate Policy (2018). *Generiek afsprakenstelsel voor datadeelinitiatieven als basis van de digitale economie.* (https://www.rijksoverheid.nl/documenten/rapporten/2018/12/30/generiek-afsprakenstelsel-voor-datadeelinitiatieven-als-basis-van-de-digitale-economie).

- Dutch Ministry of Economic Affairs and Climate Policy (2019): *Nederland Digitaal - De Nederlandse visie op datadeling tussen bedrijven.* (https://www.rijksoverheid.nl/documenten/publicaties/2019/02/20/nederland-digitaal---de-nederlandse-visie-op-datadeling-tussen-bedrijven).
- Nicolaou A.I., M. Ibrahim, E. van Heck (2013): *Information Quality, Trust, and Risk Perceptions in Electronic Data Exchanges*. Decision Support Systems (54:2), pp. 986–996. (https://doi.org/10.1016/j.dss.2012.10.024).
- Big Data Value Association – BDVA (2019): *Towards a European Data Sharing Space – Enabling data exchange and unlocking AI potential*. BDVA Position Paper (http://bdva.eu/node/1277).
- Trusted Computing Group (2013): *Trusted Multi-Tenant Infrastructure Work Group - Reference Framework.*
- Yavatkar, R., Pendarakis, D., and Guerin, R. (2000). *A Framework for Policy-Based Admission Control*. IETF RFC 2753.