

Towards a Federated Infrastructure for the Global Data Pipeline

Wout Hofman^(✉)

TNO, The Hague, The Netherlands
Wout.hofman@tno.nl

Abstract. Interoperability in logistics is a prerequisite for realizing data pipelines and the Physical Internet. Forecasting data, real time data, and actual positions of shipments, containers, and transport means shared via events have to be harmonized and are expected to improve all types of processes, support synchro-modal planning, and improve risk analysis from a compliance and resilience perspective. Technically, several solutions are implemented by organizations and innovations have been validated in so-called Living Labs or demonstrators in various projects. These solutions do not yet provide open systems required for a (global) data pipeline. A federation of solutions is required to construct data pipelines and to support sustainable development of applications on smart devices allowing Small and Medium sized Enterprises to collaborate. This paper proposes a set of platform services and so-called platform protocols to allow interoperability of different platforms for constructing a data pipeline. The proposed services and protocols further extend existing interoperability solutions and services for supply and logistics.

Keywords: Seamless interoperability · Data pipeline · Federated platforms · Service · Protocol · Physical internet

1 Introduction

Customs authorities require additionally data to their current declaration for risk analysis improvement and introduced the concept of data pipeline for seamless data sharing as a solution [1]. Such a data pipeline consists of a large number of stakeholders like shippers, consignees, forwarders, and carriers, exchanging value according a transaction hierarchy, called logistic chain, in an organizational network (reference). The actual implementation of a data pipeline is by interconnecting legacy systems of the stakeholders and/or support by commercial – and community solutions [2]. It is not to be expected that one global system will implement the data pipeline, but interoperability between existing systems and solutions needs to be constructed [2]. As of currently, many interoperability implementations in trade and logistics are based on the message paradigm, but also other mechanisms are explored to address for instance real time data sharing for dynamic planning or resilience [3], Service Oriented Architecture (SOA) supported by Enterprise Services Busses (ESBs) [4] or Linked Data [5]. For real time data sharing, the current generation of platforms supports an Application Programming

Interface (API) registry [6], a particular SOA implementation based on the REST protocol. APIs are still technical specifications that require interpretation to derive semantics.

Seamless data sharing between systems and components of different stakeholders requires universal connectivity [7]. In this respect, scoping of specifications is also important like taking a bilateral or multilateral interoperability [8] or a modeling approach covering organizational chains [9]. Interconnecting internal business processes resulted in reference models either specifying both processes and data [10] or only data with a messaging choreography [11]. Implementation of these reference models still lead to closed systems, since, organizations make bilateral or community agreements based on these models [12]. Several sources [13, 14] stress the importance of unambiguous semantics as part of interoperability, but do not address the implementation of this semantics in legacy systems or other solutions. It is yet unclear how process aspects need to get addressed in interoperability. Interoperability layering [14] considers pragmatics without presenting a way to model pragmatics like taking the bilateral or multilateral, chain approach.

A complicating factor is that Small and Medium sized Enterprises (SMEs) cover some 80 % of the logistics market [15] performing some 20 % of the business. These SMEs have either simple or no IT solutions or systems, but interface manually with systems of their customers, potentially supported by web interfaces. Thus SMEs have to deal with different interfaces to become interoperable with their customers instead of having simple applications running on smart devices with cloud solutions of one or more communities and or providers, since these SMEs operate international and require interfacing with many systems and solutions.

This paper proposes a set of platform services that enables an enterprise to connect once to an infrastructure of federated platforms and compose a data pipeline. Standardization of this set of services allows development of applications on smart devices, where these applications can interconnect to any given platform thus creating a sustainable business model for app developers in logistics. Each solution provider in this infrastructure can have its particular implementation of the services, thus satisfying their customer requirements and have sufficient market share. Firstly, requirements to platform services leading to design choices are formulated and secondly the services and the protocol for platform federation are introduced. The research presented by this paper is based on an action design research approach [16] across several EU funded and Dutch projects addressing interoperability in logistics. Each project has constructed artifacts that do however not meet requirements formulated in [7, 17].

2 Design Choices

A (federation of) platform(s) has to meet particular user requirements. Since it is fairly complex to assess user requirements for all global data pipelines, those stemming from various European Union (EU) funded projects and literature will be transformed into design choices. An example of a design choice is for instance the support of the messaging paradigm, common to most interoperability implementations between organizations. By making

these design choices explicit, discussion on their applicability to meet user requirements is supported. Design choices are on distinction of ‘service’, ‘protocol’, and ‘interface’, bilateral versus multilateral business process modeling, semantics, and data governances supported by privacy-enhanced technologies. This section presents choices based on practice inspired research [16] and briefly reflects the state of the art in research. It does not pretend to be complete, but identifies some basic research questions. The answers to these questions have to be supported by a federation of platforms; the next section shows the mechanisms to do so. Semantics is core to all choices made.

2.1 Participating in a Federated Infrastructure

Currently, organizations bilateral or multilateral develop interoperability agreements, encompassing both functional and non-functional aspects, like message implementation guidelines and process alignment, based on open standards or with proprietary formats leading to closed solutions [12]. Each time a business relation with another enterprise needs to be established, investments in agreements has to be done. Seamless interoperability [7, 17] addresses this problem, but does not provide solutions. [9, 18–19] introduce business process modeling either for bilateral or multilateral interoperability as a solution, but [20] argues to model only behavior between any two peer entities. There are different ways to specify behavior; [21] provides transaction templates for bilateral interoperability to construct chains. A generic specification of behavior will not be applicable to all resources, since they all have different goals and capabilities [22]. A generic specification can however serve as a reference framework for specifying these particular goals and capabilities.

This paper proposes to apply the concept of ‘resource’ offering both real time data and providing or requesting logistic services as specified by an ‘Information Profile’ of such a resource. In this respect, several issues need to be addressed, namely how to express the external behavior of a resource in terms of interactions and semantics. The concept of transaction templates to express external behavior for business transactions can be applied [21]; other mechanisms like events might be required to share any logistic state changes like arrival of a vessel and delivery of a container at its destination. Semantics of one’s profile can be expressed as an ontology, based on a networked ontology of logistics concepts and services (see for instance ontology.tno.nl for a logistics ontology). Semantics of data and behavior need to have a technical binding to a paradigm like messaging and SOA [4] supported by a syntax like XML or JSON. In case any two communicating organizations have different technical bindings, binding negotiation and a data transformation function have to be implemented, either by a data provider, a consumer, or as service of the federation of platforms (see next section). In case Information Profiles of any two communicating organizations are based on an identical - or matched semantic models, an on-the-fly technical binding by a platform can be constructed, as long as the federation of all platforms support that technical binding.

The concept ‘resource’ with its ‘Information Profile’ needs further research and examples stemming from real use cases. These examples are currently developed in EU funded projects like EU FP7 CORE.

2.2 Data Governance and Privacy Enhanced Technologies

Organizations are hesitant in sharing information due to for instance commercial or liability reasons, e.g. the amount of free capacity of a barge might reduce prices, the location of a truck might increase vulnerability for cargo theft or providing real-time and predicted depth of a waterway might increase liability. There are currently a number of barriers that block the adoption of data sharing amongst resources, e.g. data ownership, privacy, commercial sensitivity, liability, and culture [23]. In this respect, data and events are classified as:

- Open data. Data is publicly available to everyone. Open data is normally considered to be available without any costs, but in some occasions like the Cadastre data, one needs to pay.
- Community data. Data is shared within a community according agreed rules. Like with open data, one might distinguish free - and paid data.
- Partner data. Data is shared with a specific partner.
- Internal data. Data is only shared within an organization, according internal data policies (Fig. 1).

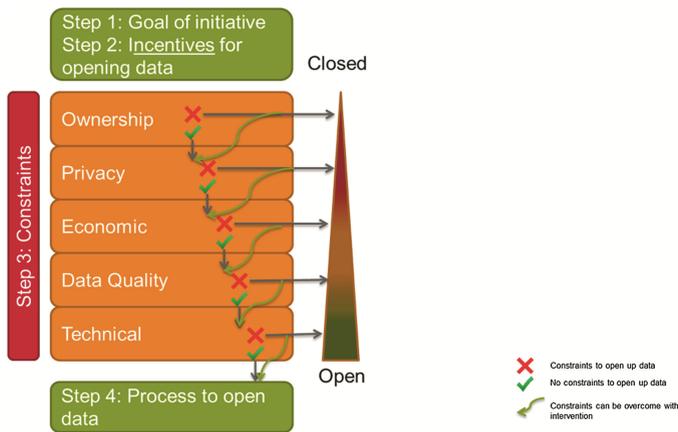


Fig. 1. Decision support instrument for data sharing [23]

Data classification has a lifetime, e.g. it may change over time and/or may be applicable for one or more calls or interactions. For instance, available capacity on a trip may be shared only once in a community at the start of the trip or can be updated during the trip. Communities can also be flexible, e.g. organizations can join and leave a community over time.

The previous figure shows the decision model developed by [23]. It addresses various aspects like data ownership, privacy and commercial sensitivity, and economic aspects,

resulting in a data policy supported by interventions. Many of these interventions are supported by privacy-enhanced technologies [24] like identification and authentication, access control, filtering, and homomorphic encryption [25]. One particular technology might support data sharing for one or more decisions, but we have not yet found these for logistics data sharing. For instance, Role Based Access Control expresses access control for internal data, but a more fine grained access control mechanism like Attribute Based Access Control might be required for partner – and community data. Templates for particular roles can be specified to ease the specification of data policies, e.g. a template for a role like a forwarder. These templates are a form of Role Based Access Control that can thus be refined by organizations meeting their particular requirements as they operate in one or more roles. These templates may implement formal restrictions from a liability and financial perspective, e.g. carriers should not receive any data on the content of a container, whilst they are otherwise liable for any damage or loss to the content.

Role – and Attribute Based Access Control can be expressed as an ontology of a set of rules, based on the earlier mentioned networked ontology for logistics. To support global logistics and supply, protocols are required for a federation of identities [26]. Further research is required with respect to the relation between privacy-enhanced technologies as interventions for data governance and implementing these technologies in real world cases.

2.3 Service, Protocol, and Interface

The Internet design principles of ‘service’, ‘protocol’, and ‘interface’ [20] are applied for specifying a federation of platforms supporting data sharing between organizations. A federated platform is said to offer a ‘service’ to back office systems of supply and logistic stakeholders, e.g. the ability to exchange messages, validate the message structure and content, and validate the message sequence, whereas a ‘protocol’ between any two platforms provides the ability to actually share data with for instance messages. The protocol is the set of agreements for sharing data between any two platforms, independent of a local implementation of the service by each of those platforms to their users, which is called ‘interface’. The service is the conceptual representation of this protocol to a one or more back office systems and/or end-user. The same service can be implemented by various interfaces, e.g. a file sharing mechanism or an API can serve as an implementation of a service. In fact, ‘interface’ is the technical binding of a service offered to back office systems of an organization. The technical binding of an interface can differ from that of an agreed binding of the protocol, which requires transformation by a local implementation of the protocol and service by a platform (Fig. 2).

Introducing these concepts allows conceptually specification of a service and a protocol with different technical bindings, both for the service by its interface and a protocol for interoperability between two local implementations. Complexity reduction of federated platforms is achieved if all participating platforms support the same semantics and technical bindings of the protocol.

Each service – and protocol primitive has a particular structure with control information and a payload, where a semantic model specifies the semantics of the payload.

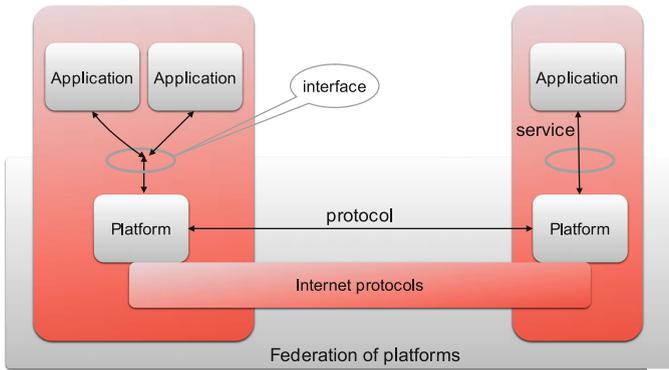


Fig. 2. The concepts service, protocol and interface

Control information of a service primitive is applied by a local implementation for processing that primitive and constructing a protocol primitive with a negotiated - or on-the-fly technical binding (see before). It allows for instance the transformation between a SOA interface of a service to a message based protocol and vice versa. Communication is already implemented by various protocols and is therefore considered out of scope. Thus, the service and protocol for logistics will be elaborated, whereas a local interface is discussed in the next section considering back office integration.

3 A Federation of Platforms

The previous section has introduced a number of research questions and concepts for federated platforms. This section further elaborates the services of a (federation of) platform(s) and its protocol. The services reflect the concept of Information Profile and the support of data governance, all based on networked ontologies.

3.1 Federated Platform Services

A (federation of) platform(s) provides generic services to resources, where these generic services are configured by semantics for a particular application area like supply and logistics. The services can be categorized in two main groups that can be further decomposed (Fig. 3): operational services consider actual sharing of logistics data utilizing various technical bindings and mechanisms; administrative services consider registration and other types of support services for operating a (federation of) platform(s). The decomposed services can be defined as:

- Registration Services. These services support an organization in its registration and connection to a platform (profile specification services) and specify its data policies (data policy specification services).

- Real time data sharing services. These support data policy negotiation, both for events and data, search for a particular (composite) service matching a goal (matching services), and sharing the state of supply and logistics chains via events (visibility services). Visibility services can be on particular objects like trucks with their location, speed, etc. in an area (geo-fencing), timeframe (time-fencing) or a combination of both [27]. Visibility services can only provide the state of an object or also evaluate the state against requirements, which is supported by (complex) event processing [28]. Matching services can be on search for a (structured set of) profile(s) meeting a customer goal, where a structured set composes a logistics chain or all retrieved profiles exactly meet a customer goal. These matching services can be applied in various ways, e.g. to support synchmodal booking [28].
- Transaction support service: validating the sequencing of interaction according an agreed transaction protocol specified by a choreography. In this particular case, the initiation and processing of transactions is in applications registered at a platform.
- Data sharing service: reliable and secure exchange of data and events according a particular technical binding, potentially supported by data transformation. Reliability services consider data resubmission in case the receiving platform or application responds and are not always required. The same is applicable for secure data sharing.
- Supporting services. These services support the operation of a federation of platforms providing particular services to platform users. These services not necessarily have a supporting protocol. The following supporting services are foreseen:
 - Semantic services providing the networked ontologies (see Sect. 2).
 - Publish/subscribe services providing the ability to subscribe to particular events, where these events provide state information. Publish/subscribe may require policy negotiation services.
 - Non-repudiation services that provide proof of actual data shared between any two users of a federation of platforms. Non-repudiation services are supported by an audit trail registering all actions in terms of data sharing between any two actors, e.g. sending or receiving particular data with a timestamp, a log containing the actual data that has been shared, and monitoring services providing both access to the audit trail and the log.
 - Accounting and billing services supporting paid data according agreed pricing structures. These services utilize the non-repudiation services.
 - Certification services providing identification and authentication of platform users.

Services are interrelated. Visibility services can for instance use publish/subscribe services and data exchange services for events, transaction support services can utilize data exchange services by messaging and reliability services to assure a reply is received in time. Secure and reliable data exchange is for instance specified by Electronic Business XML with ebMS [19] and implemented by an eFreight access point [29]. A formal service specification considers the control information and payload [20], which yet needs to be performed.

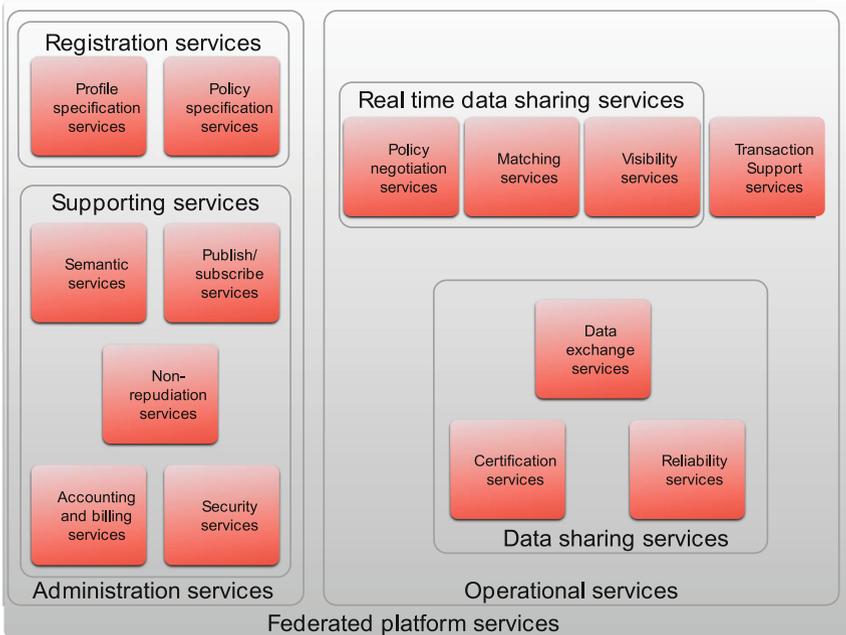


Fig. 3. Federated platform services

3.2 Federation of Platforms

Like the services, the protocol can also be decomposed. The protocol should support each of the services, but some protocols can support more than one service by a different payload of the protocol primitives [20]. The protocol is decomposed as follows:

- Data policy negotiation protocol: negotiate the data that can be shared amongst two organizations.
- Matching protocol: sharing goals and (a structured set of) service(s) to support matching services.
- Visibility protocol: how to access particular data on the supply chain status, e.g. it basically consists of events either received upon subscription or by a query.
- Transaction protocol: the agreed choreography of interactions for transaction support services.
- Data sharing protocol: sharing data between two communicating systems, where the payload is provided by the aforementioned protocols. The data sharing protocol is decomposed in:
 - Binding negotiation protocol: to establish the technical binding for sharing data.
 - Data exchange protocol: the actual sharing of data (and events) according an agreed or selected (on-the-fly) technical binding.

- Reliability protocol: resubmission and identification of resubmitted messages, potentially resulting in a receipt acknowledgement, and timers to detect timely replies to support reliability services.
- Security protocol: selecting a secure protocol with agreed certificates, e.g. https.
- Authentication protocol: a protocol for federation between certification authorities to support authentication of an identity [26].

One might consider to introduce a protocol for supporting registration services, e.g. to share complete Information Profiles and data policies. Such a protocol would provide complete transparency. In our proposed approach, the data negotiation -, matching -, and binding negotiation protocol support this type of transparency, but does not provide a generic data policy of any registered user.

3.3 Implementation and Deployment

There are several ways to implement a federation of platforms, e.g. if any two enterprises are connected to the same platform like a Port Community System, data sharing services between them might be implemented via a database and registration is based on administrative services with a proprietary format. Only in case two enterprises connected to different platforms, the protocols will be required. Currently, most of these protocols are message based (see before).

There are two dimensions to the deployment of the Connectivity Infrastructure, namely an business dimension and ICT dimension. Both will be discussed here. The ICT dimension addresses the development of a local implementation, with potentially different components providing different services and supporting particular parts of the protocol. The following options are feasible:

- Open source: the services are provided by an open source software solution that every resource can implement. Like indicated before, eFreight Access Points provide particular functionality, so do iCargo Access Points in supporting virtualization of logistics actuator objects like containers and trucks [3].
- COTS (Commercial Of The Shelf): the implementation of the protocol with a local interface to its back office systems based on COTS. The software offering the service and supporting the protocol is licensed to a resource or its owner that implements the functionality. The COTS provider is responsible for correct (and complete) implementation of the protocol.
- Proprietary: the IT department of an enterprise develops the implementation of the protocol or its implementation is outsourced to an external software developer. The implementation can be based on open source solutions and/or components like available integration brokers. The solution is owned by the enterprise (with the exception of COTS components used for implementation) and the enterprise is responsible for correctness of the protocol implementation.

From a business perspective, the service can be implemented in many ways, for instance:

- **Resource Deployment:** each resource implements and operates the protocol by itself. The services are internal to the resource, but utilizing an ICT solution implementing the services reduces development costs. The iCargo Access Points try to provide this functionality [3].
- **Community Systems:** two or more organizations might decide to implement the services themselves with different local interfaces to back office systems and/or end-user functionality. These organizations thus own the community system. Port Community Systems are examples of these types of systems.
- **Cloud Platform:** the service and the protocol are provided by a platform of a commercial provider. The latter ones can construct their particular solutions on top of the services. Note that this type of solution is identical to telco or other providers offering services with COTS and/or open source solutions for communication protocols.

The services can also be used for open innovation, implying the development of apps for SMEs (see also www.logicon-project.eu). To support open innovation, the services have to be more tailored to a specific target group of end-users, e.g. barge operators are expected to have other services than truck drivers.

A number of services is currently proprietary to a particular solution, e.g. registration of resources at a community system or cloud platform. These deployment solutions do not yet support the protocol, which makes it difficult to find for instance resources in an infrastructure. Furthermore, these deployment solutions have their specific data policies implemented by message implementation guides, which restricts data sharing across these platforms and require all types of transformations between them.

4 Conclusion and Further Research

The paper presents a set of services for federation of platform solutions and services supported by a protocol for logistics and supply. By standardizing supply and logistics services and their underlying protocol, the so-called data pipeline for interoperability in trade facilitation will be enabled. The services and protocol enable each individual object or actor to act as a information resource with particular capabilities for data sharing. By examples, we have mapped functionality to existing components like developed for eFreight and iCargo. Dedicated solutions and services can be developed addressing particular interoperability aspects like virtualization of actuators representing physical objects like trucks and containers, thus also contributing to the concept of the Physical Internet [7]. Standardisation of services of federated platforms also contributes to development of sustainable business models for deployment of apps on smart devices for SMEs. By separating service and protocol, each resource implementing a protocol stack known to the infrastructure will be able to participate, without additional costs and effort for development of bilateral or community guidelines. Each ICT solution – and service provider will also be able to tailor its services to optimally integrate resources in the infrastructure and provide added value like complex event processing and transformations to these resources.

The solutions provided by this paper need to be developed further, including construction of low cost connection to the federation of platforms. We have indicated that lots of existing systems and cloud solutions have a role in implementing the services. Federation of platforms requires additional research into the business models of these platforms, like a sustainable business model of community systems. Not yet, all identified services are fully supported by software solutions.

Acknowledgements. The content of this paper is based on research and development in EU FP7 SEC Cassandra, EU FP7 SEC CORE, EU FP7 INFSO iCargo, EU FP7 INFSO Logicon, several other projects for data sharing in logistics and supply chains, and discussions in various fora like ALICE on the subject of seamless interoperability.

References

1. Klievink, B., Van Stijn, E., Hesketh, D., Aldewereld, H., Overbeek, S., Heijmann, F., Tan, Y.-H.: Enhancing visibility in international supply chains: the data pipeline concept. *Int. J. Electron. Gov. Res.* **8**(4), 14–33 (2012)
2. Hofman, W., Bastiaansen, H.: A global IT infrastructure improving container security by data completion. In: ECITL, Zaragoza, Spain (2013)
3. Dobler, M., Schumacher, J.: A pan-european ecosystem for intelligent cargo and its impact on emission reduction. In: *Transport Research Arena 2014*, Paris (2014)
4. Erl, T.: *Service Oriented Architecture – Concepts, Technology and Design*. Prentice-Hall, Upper Saddle River (2005)
5. Heath, T., Bizer, C.: *Linked data - evolving the web into a global data space*. In: *Synthesis lectures on the Semantic Web: Theory and Technology*, Morgan & Claypool Publishers (2011)
6. Hofman, W., Rajagopal, M.: A technical framework for data sharing. *J. Theoreticla Appl. Electron. Commer.* **9**(3), 44–57 (2014)
7. Montreuil, B., Meller, R.D., Ballot, E.: Physical internet foundations. In: Borangiu, T., Thomas, A., Trentesaux, D. (eds.) *Service Orientation in Holonic and Multi Agent Manufacturing Robots*, pp. 151–166. Springer, Heidelberg (2013)
8. Schonberger, A., Wilms, C., Wirtz, G.: A requirements analysis of business-to-business integration. In: *Fakultat Wirtschaftsinformatik und angewandte Informatik Otto-Friedrich-Universitat, Bamberg* (2009)
9. Huemer, C., Liegl, P., Motal, T., Schuster, R., Zapletal, M.: The development process of the UN/CEFACT modeling methodology. In: *10th International Conference on Electronic Commerce (IEC)*, Innsbruck, Austria (2008)
10. Pedersen, J.T.: One common framework for information and communication systems in transport and logistics: facilitating interoperability. In: Golinska, P., Hajdul, M. (eds.) *Sustainable Transport. EcoProduction. Environmental Issues in Logistics and Manufacturing*, pp. 165–196. Springer, Heidelberg (2012)
11. World Customs Organization. *WCO Data model - cross border transactions on the fast track*. World Customs Organization (2010)
12. Hofman, W.: Applying semantic web technology to interoperability in freight logistics. In: *e-Freight 2011*, Munich (2011)
13. European Commission. *European Interoperability Framework for European public services*. European Commission, Brussels (2010)

14. Wang, W., Tolk, A., Wang, W.: The levels of conceptual interoperability model: applying systems engineering principles to M&S. In: Spring Simulation Multiconference (2009)
15. Transport and Logistics Netherlands (TLN). Transport in figures, TLN (2015)
16. Sein, M.K., Henfridsson, O., Purao, S., Rossi, M., Lindgren, R.: Action design research. *MIS Q.* **35**(1), 37–56 (2011)
17. Chituc, C.-M., Azevedo, A., Toscano, C.: A framework proposal for seamless interoperability in a collaborative networked environment. *Comput. Ind.* **60**, 317–338 (2009)
18. Van den Heuvel, W.-J., Papazoglou, M.P.: Towards business transaction management in smart networks. *IEEE Comput. Soc.* (2010)
19. Kotok, A., Webber, D.: ebXML - the new global standard for doing business over the internet, New Riders (2002)
20. Tanenbaum, A.S.: *Computer Networks*, 3rd edn. Prentice Hall, Upper Saddle River (1996)
21. Dietz, J.: *Enterprise Ontology, Theory and Methodology*. Springer, Berlin (2006)
22. Spohrer, J.K.S.: Service science, management, engineering, and design (SSMED) - an emerging discipline - outline and references. In: *International Journal on Information Systems in the Service Sector*, May 2009
23. Eckartz, S., Hofman, W., Van Veenstra, A.F.: A decision model for data sharing. In: *eGov2014*, Dublin, Ireland (2014)
24. Legendijk, I.L., Erkin, Z., Barni, M.: Encrypted signal processing for privacy protection. *IEEE Signal Process. Mag.* **82** (2013)
25. Gentry, C.: A fully homomorphic encryption scheme. Stanford University (2009)
26. Pruksasri, P., van den Berg, J., Hofman, W., Deskapan, S.: Multi-level access control in the data pipeline of the international supply chain system. In: *iCETS*, Macau (2013)
27. Merrienboer, S., Veenstra, A.W., van den Haak, W., Tavasszy, L.: Using floating truck data to optimise port logistics. In: *10th Intelligent Transport Systems European Congress (ITS2014)*, Helsinki (2014)
28. Hofman, W.: Control tower architecture for multi- and synchromodal logistics with real time data. In: *ILS2014*, Breda (2014)
29. Angholt, J., Wackerberg, M., Olovsson, T., Jonnson, E.: A First Security Analysis of a Secure Intermodal Goods Transport System. Chalmers University, Gothenburg (2013)