WHITE PAPER

The what, how and why of data sharing

Port logistics more efficient, reliable and sustainable

smartport.nl

Knowledge hub SmartPort Rotterdam is a partnership between the Port of Rotterdam Authority, Deltalinqs, the Municipality of Rotterdam, TNO, Deltares, Marin, Erasmus University and Delft University of Technology. We form the bridge between port businesses, knowledge institutions and governments. By inspiring, initiating and forming alliances SmartPort stimulates and finances scientific research for and by the companies in the port of Rotterdam in collaboration with knowledge institutes. SmartPort's goal is the acceleration of innovations to keep the port of Rotterdam future-proof and together form the smartest port in the world in 2030-2050.

www.smartport.nl | LinkedIn: smartportrdam | Twitter: SmartPortRdam | Instagram: smartportrdam

SMARTPORT PARTNERS

Deltares | Deltalinqs | Erasmus | Gemeente Rotterdam | Port of Rotterdam | MARIN | TNO | TU Delft

WITH CONTRIBUTION

TKI DINALOG
Dutch Institute for Advanced Logistics

THE WHAT, HOW AND WHY OF DATA SHARING

Port logistics more efficient, reliable and sustainable

Natasja van der Keijl (CGI) & Hans Moonen (CGI and University of Twente)

CONTENTS

A WHITE PAPER ON DATA SHARING IN THE PORT

Data sharing is a hot topic and one where a lot is currently happening. Parties in and around the port are asking how this affects or will affect their logistics processes. This white paper attempts to answer in accessible terms the what, how and why of data sharing within port logistics. We do this by combining relevant experiences of both authors, knowledge gathered from desk research and insights gained from a series of 15 interviews with direct stakeholders from the logistics field. Finally, the white paper was reviewed and enriched through seven interviews with experts on the topic of data sharing.

-01- Introduction

Digitalisation is a development that no one can ignore anymore. It is no longer a choice, but a necessity to keep up with the competition or to arm yourself against rivals from other industries built around a digital proposition. Port environments and processes are also becoming increasingly digital. Consider, for example, developments such as the Internet of Things, autonomous transport, 3D printing and the increasing focus on cyber security (Horst et al., 2019). However, digitalisation cannot happen without data and data sharing between parties.

Data sharing refers to making data or information from one organisation available and accessible to one or more other parties. Within logistics, data sharing is crucial. After all, logistics is 'the art of smart organisation' as the Van Laarhoven Commission (2008) powerfully put it. Following this definition, logistics is the connecting activity aimed at balancing supply and demand, managing (transport) movements in an environment with scarce capacity (of infrastructure and means of transport), all at low cost and with minimal impact on the living environment. Having and obtaining good information is crucial to properly manage operations. In other words, it is important to obtain data from chain partners (such as clients, suppliers and customers) and other parties in the ecosystem. Take, for instance, the importance of the accurate Estimated Time of Arrival (ETA) of sea-going vessels for many parties in the port and hinterland chain.

Data sharing is not new, but developments are rapid.

This obviously does not make data sharing a new topic. Within logistics chains, (digital) data has been shared between trading partners, with infrastructure managers and authorities for decades. However, developments are moving fast. Exactly what kind of data is shared and why? There are also more and more forms of data sharing. For instance, data sharing bilaterally, through a community system, a blockchain, a commercial platform, or federated data spaces. What exactly is what, how does it work, and why should we share data at all?

This white paper attempts to answer in accessible terms the what, how and why of data sharing within port logistics. It will give readers an insight into the state of data sharing in current practice among industry peers. It also explains the rationale of digitalisation and increasing importance of data and data sharing as an unstoppable development. It then lists three main forms of data sharing, compares them and links them to the challenges of daily practice. Finally, it briefly discusses the innovation tools currently available and provides five concrete tools on how to get started with data sharing.

-02-

Data and data sharing in port logistics today

During the preparation of this white paper, we conducted a round of interviews with some 15 parties involved in port logistics at SmartPort's request. During these interviews, we discussed the importance of data in operations, how data is currently being shared and the challenges they see with regard to this topic.

SMARTPORT'S DATA CHALLENGE

SmartPort is the knowledge hub of the port of Rotterdam. SmartPort has been developing scientific knowledge with and for the port business community since 2015. A large part of the more than 120 studies with more than 200 companies focus on data. Examples include:

- The Physical Internet project (2016-2021) explored how in a (distant) future, logistics processes could become fully self-organising, and what this means for the parties involved.
- SOLport project (2020) explored applying the concept of self-organisation to processes in an inland terminal and in the port.
- Talking Trucks project (2021-2022) aimed to let trucks schedule themselves via smart algorithms at an inland terminal.
- The Digital Twin in Inland Navigation project (2020-2021) investigated the application of a Digital Twin in inland navigation to respond to climate change impacts. Various data sources were combined in the Digital Twins (high/low tide, inland waterway operator data, depth of waterways data).

The projects show a great desire for further automation and digitalisation in the port of Rotterdam. The overall conclusion, however, is that the gap between desire and reality is, unfortunately, still wide. The main challenges during these projects turned out to be data quality and data structuring.

Summing up the interviews, a pretty sobering picture emerged about data and data sharing in port logistics today. To begin with, electronic communication still often takes place in the form of email or interaction via a web interface. So not system-system interaction, but human-system or human-human interaction. With humans being the limiting factor, it is difficult to maintain a real-time information picture. Platform parties Portbase and UTURN also report that the majority of their users still work primarily via the web interface. In other words, they do not use a system link, but interact manually via the web browser. Logistics parties additionally feedback that they still use email as their primary means of communication with many customers and suppliers despite years of collaboration. One example specified by Grolsch was packaging flows of empty beer crates. Regular flows are neatly automated, but recovering packaging requires a lot of manual work, with daily emails, follow-up calls, and maintaining separate records.

Another aspect frequently mentioned is that in (port) logistics chains, a lot of cargo information is still on paper. The information is not available digitally or only to a limited extent, and a lot of work goes into digitising this information.
A third notable point is that many logistics parties report spending a lot of manpower on monitoring the chain. For instance, the arrival of a ship at the seaport. What is the sea-going vessel's estimated time of

arrival (ETA) and estimated time of departure (ETD)? When will specific containers become available at the terminal? Monitoring one system won't suffice in this case. Many parties say they monitor multiple sources, feed information from them (often through web-scraping) into their own local system, combine the data and draw their own conclusions from it. Sometimes because the 'official' information is outdated and not updated often enough.

This point also touches on another frequently mentioned problem, namely the fact that terms are not always clearly and unambiguously defined. For example, what one party considers to be an ETA, is considered something else by another party. The ETA of a sea-going vessel may, for the Pilotage Service provider, be the arrival at the off-shore embarkation location, while for the towage service provider, this is the moment of mooring the tow ropes, and for the terminal, the actual moment of mooring lines at the quay. This is not only a problem between parties, but many organisations seem to suffer from this internally as well – different definitions are used in different processes and systems, causing a lot of confusion.

Major business cases for data sharing centre around making adjustments based on accurate and timely information from the chain. One example mentioned is slow steaming. This allows prevented waiting time to be converted into slower sailing speed, saving fuel. A nice concept, but it will only work if all the information is correct, such as travel time, congestion, and handling time, and practice does not show prioritisation of other parties. A second example concerns the adjustment of conditioned transport, based on, for example, market information and/or the condition of the goods, obtained by real-time sensors measuring the condition of the goods.

It is also mentioned several times that information is often collected only for one specific purpose, while it can be used for many more purposes. A good example is the use of intelligent bollards in the seaport, which measure the pulling force exerted on the bollard. Unfortunately, this data is not yet used for other purposes. The Port of Rotterdam Authority is now considering storing this kind of data in Digital Twins. Digital Twins are digital copies of a fixed or moving (physical) object in the port, in this case a bollard. The data from the Digital Twin can then be accessed by other applications, for example to be used to determine quayside occupancy by monitoring which bollards are collectively under tension and since when. In other words, data reuse.

Finally, some points of criticism regarding data sharing are also discussed. Several parties indicate that transparency, in the form of shared data, can be used against you. In (future) contract negotiations, or in created expectations in the operation that ultimately did not materialise. Others oppose the word 'sharing' in data sharing. This term does not directly imply gain for a party, but seems to hint that effort has to be made to give up something without getting anything directly in return. Concerns are also raised about the pace of implementation/realisation, with the (slow) roll-out of eCMR and iShare being cited several times.

In summary, it can be concluded that market participants have indicated that their needs include:
- Being able to create system links easily and quickly
- Being able to share data with chain parties with whom there is no direct contractual relationship
- A unambiguous conceptual framework in the chain
- Being able to receive data from the (best) source

Parties do see the need for data sharing with chain partners, but struggle with how this can effectively be achieved given the many practical, technical and organisational challenges. Let's first reflect on the why of data sharing.

-03-
Data sharing is a logical consequence of increasing digitalisation of logistics

Fifty years ago, US professor Jay Galbraith formulated the *organizational information processing theory* (Galbraith, 1974; Premkumar et al., 2005). This theory describes two strategies an organisation can adopt to counter uncertainties in their operations: (1) Using buffers, or (2) Committing to structurally better information processing. As far as Galbraith is concerned, these strategies can be applied both within (large) organisations as well as between organisations. An example of the first strategy is focusing on inventories to overcome uncertainty in supply chains. An example of the second strategy is setting up organisation-wide IT systems (e.g. Enterprise Resource Planning (ERP) systems) that give the entire organisation an up-to-date view of orders, inventory, and capacities. What is remarkable, though, is that in contrast to sharing information internally, only limited strides have been made in sharing information between organisations over the past 50 years. Orders are often communicated electronically, but further information exchange between organisations remains limited for the time being. In short, there is still much to be gained – and not only in theory – by focusing on more and better data sharing between organisations in logistics chains. From a theoretical point of view, betting on further digitalisation and data sharing is a logical next step.

Digital is no longer a choice, but a necessity.

This trend can be seen within the logistics sector. Digitalisation, digital transformation and becoming data-driven are themes that have been high on the agenda of most companies within the sector in recent years (CGI, 2022). Digital is no longer a choice, but a necessity to keep up with the competition or to arm yourself against rivals from other industries built around a digital proposition. Digitalisation is a tool that helps to do what you already did as an organisation better (faster, cleaner, cheaper) and/or to do completely new things that you did not do before (Sebastian et al., 2017; Blackburn et al., 2021). For instance, barges become information providers to take dynamic depth measurements on the river. Major challenges facing logistics today, such as major global uncertainties due to disrupted trade flows, geopolitical developments, the impact of (Covid) lockdowns, high fuel prices, not to mention the challenges associated with climate change mitigation measures, create the need for smarter solutions. Data sharing is a logical extension of increasing digitalisation efforts. To put it in terms of the late Prof. dr.ir. Jo van Nunen, former port professor at Erasmus University: "Will you digitise, or will you be digitised?"

"Will you digitise, or will you be digitised?"
- Prof.dr.ir. Jo van Nunen

At the same time, many parties also rightly have reservations about data sharing. After all, increasingly the adage goes: data = power (Snijder, 2017). Data has thus become part of the strategic positioning within a logistics ecosystem, and not something you 'just give away' to others 'for free'. It is not entirely surprising that this causes parties to start thinking about payment models for making their data available.

-04-
What exactly is data?

An almost exponentially growing amount of data is produced within logistics processes. This is partly because hardware is getting smarter (e.g. sensors / the Internet of Things) and the fact that data processing and data communication is getting cheaper. Making data useful, i.e. giving it context and thus translating it into information, is a challenge. Things to consider include the available data, the different data formats to be processed and the various sources to be combined. There are also many different forms of data, which can be divided in several ways. Three of these divisions are detailed in this section. A division by 1) size, 2) type and 3) purpose of use. Figuur 1 provides an illustration of these three possible divisions.

DATA VERSUS INFORMATION

Are data and information synonyms or not? Data comprises unstructured data about business processes, products and customers. Information is created once data is placed in context. In short, information has context, data does not. In practice, the terms are regularly used interchangeably. To align with the concept of data sharing, this white paper uses data as the core term, occasionally possibly indicating information.

First of all, there is an important distinction to be made between the format of the data, namely structured, semi-structured and unstructured data. Structured data is data that conforms to a predefined data model. This makes the data relatively easy to read and interpret for other systems. The data conforms to a tight table format in which the relationship between different rows and columns is fixed. This makes this data quickly searchable and easily reusable. Unstructured data on the other hand, does not follow a predefined data model. Unstructured data may have its own internal structure, but it does not fit neatly into a database. Examples of this type of data include media files (such as photos and videos), social media posts and emails. Text files are also often unstructured in nature. There is also a category that has characteristics of both, semi-structured data. The more unstructured the data, the more difficult it is to analyse and reuse. It is estimated that around 80% of the data within organisations is of the semi-structured or unstructured type, this is where the challenge lies.

It is estimated that around 80% of the data within organisations is of the semi-structured or unstructured type.

Another distinction in data can be made by type of data – see figuur 1. Transactions are formal messages that bindingly initiate a logistical process. Examples include many of the messages that go through Portbase, as well as direct orders between parties. Events are the digital representation of the result of an action in the physical world: for instance, 'expected time of arrival issued' and 'container unloaded'. An event refers to the result of actions, not the action itself. Events often follow previous events. The loading of a container onto a ship in China, the 'container loaded onto ship' event, for instance, can trigger a notification in the Port of Rotterdam. Sensor data is measurement data collected by a sensor. This is often data that is logged periodically, say every five minutes. Within the port of Rotterdam, the previously mentioned Internet of Things sensors measuring the flow rate of the Nieuwe Waterweg, as laid down by the Port Authority together with the Directorate-General for Public Works and Water Management, are an example thereof. Another example are sensors on containers that measure and record various things. Forecasts refers to data resulting from predicting future developments. A well-known example in port logistics is predicting the arrival time of a sea-going vessel, the ETA (Estimated Time of Arrival). This is an important reference point for parties to plan their activities on. Master data (also known as reference data) define the core data of an organisation, or even industry, by capturing structured reference data which provide a single reference point for use in different applications. An example thereof is the reference list of codes as defined in ISO 6346 for shipping containers. Metadata is data that describes the characteristics of certain data. So it is actually data about data. The process of linking metadata to underlying data, and thus making it easier to find, is called attaching metadata. Examples of attaching metadata include recording attributes such as: author, year, and language. Historical data refers to data created in the past which is no longer current. This type of data is often recorded in so-called data warehouses, and can be used for analytical purposes. An example thereof is logging a series of consecutive ETA predictions for the arrival of a specific ship, in order to arrive at better prediction algorithms and issue more accurate ETAs in the future.

A third division can be made according to the underlying purpose for sharing data within the logistics chain: 1) contract, 2) compliance, and 3) coordination (Wagter, 2021) - See Figure 1. An example of a contract is an order. This involves one party entering into an obligation with another party, with financial consequences. Sharing compliance data is often the result of regulatory obligations, for example, information on hazardous substances, emissions or customs documentation. Often, a government is the receiving (and obligating) party in this. The coordination data has everything to do with smart organisation within logistics. In order to organise in a truly smart manner, it is important to have a good overview of what is happening in the logistics chain. Has the sea-going vessel arrived? Has a container already been unloaded? Are customs releases in order? What time will the truck arrive at its destination? What is the latest time at which a train can still be loaded? In short, crucial data that often resides with a party with whom there is no contractual relationship. For instance, a transport carrier that comes to pick up a container at a terminal on behalf of a shipper, which is unloaded from a sea-going vessel (shipping company). The contractual relationships are between the carrier and the shipper, and between the shipping company and the terminal. However, there is a major coordination issue between the carrier and the terminal to agree on the ideal time to pick up the container: the container must be physically present, customs releases must be in order, there must be handling capacity, while both the terminal and the carrier want to keep their costs down. Also making this 'coordination' data accessible to non-contract partners is a challenge since there is no natural relationship.

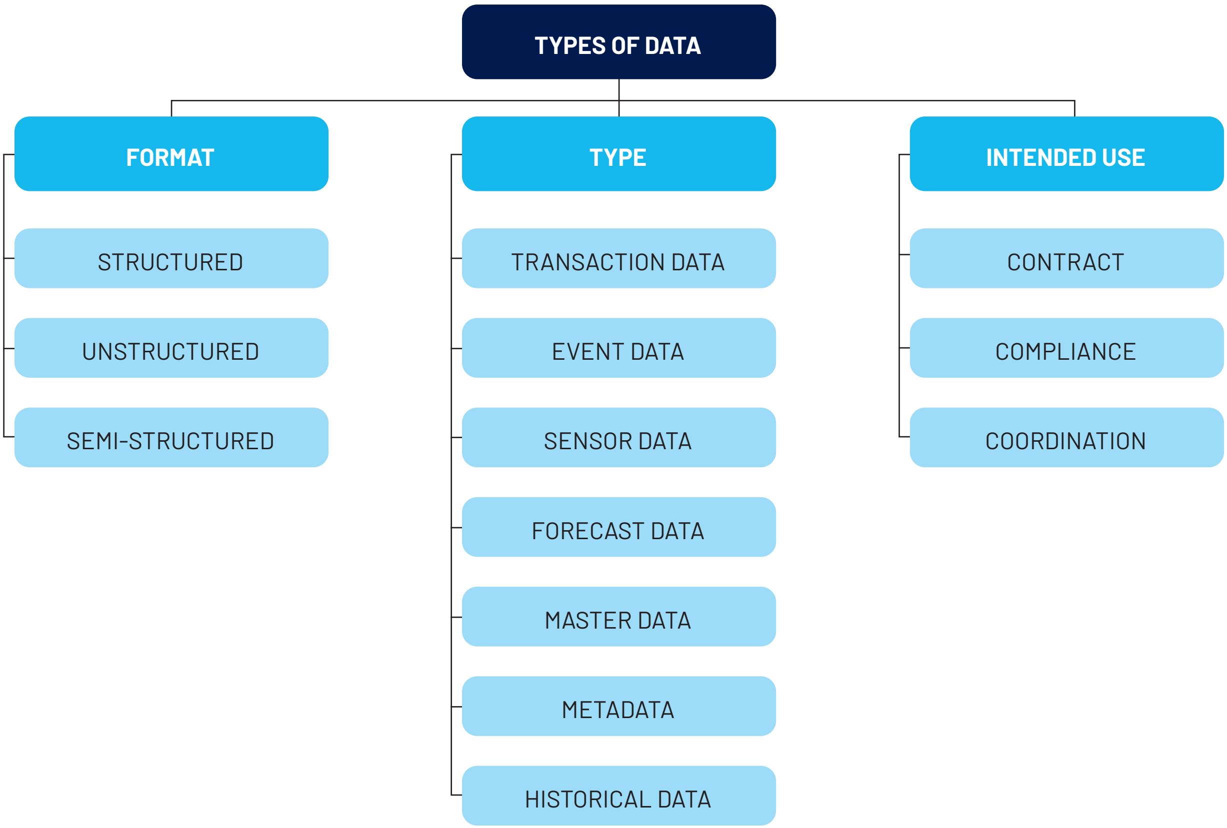


Figure 1 – Visualisation of different types of data, broken down by format, type and intended use

-05-
Different types of data sharing

Data is everywhere, in many different forms, and the amount of data is growing exponentially. In addition, it is clear, from both theory and practice, that it is crucial to be able to use data from elsewhere in the chain within the logistics business process. Time to consider the different types of data sharing, starting with bilateral data sharing. We will then discuss data sharing via a central platform, and finally we will look at federated data sharing – the type of data sharing that is currently emerging.

Data is everywhere and is growing exponentially.

5.1 Bilateral data sharing

With bilateral data sharing, one organisation shares data with one other organisation through a direct (system) link. This is the simplest form of data sharing.

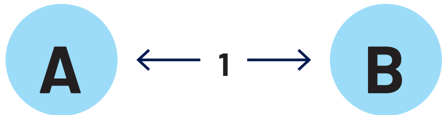


Figure 2 - Example of data sharing using a bilateral link between two parties

For a single link, this is easy to implement, as only two organisations need to agree on the data to be exchanged. A major disadvantage of bilateral is that it is not easy to quickly scale up and exchange data with many parties. Each new link requires a lot of time. Standardisation of messages (e.g. through industry-specific EDI or XML standards) allows faster scaling up in the number of links, but often a standard just does not suffice, and adjustments are necessary.

Linking all parties in a logistics chain through bilateral links will result in an awful lot of links. After all, setting up bilateral links scales up quadratically. To (fully) connect four parties bilaterally requires a total of six links, with five parties there are already ten links (see Figuur 3), and moving to a network of ten parties you already need to set up 45 links. Leaving aside the question of whether it is realistic for all parties in a logistics chain to be linked, this shows that this is a complicated solution in terms of scalability. It also raises a major management issue, as changes in the systems or data fields to be shared directly impact a multitude of links.

If you compare the list of wishes as retrieved in the round of talks with port logistics parties with bilateral data sharing, it is obvious that this is not a solution for all points (also see Tabel 1). For one, the requirement to be able to quickly create system links is not fully met. After all, it requires agreement between both parties on the what and how of data sharing, and this takes time and effort. Bilateral data sharing with a party with whom there is no direct contractual relationship will be difficult. Naturally, there is no logical trigger for both parties to put in the effort. When creating a bilateral link between parties, it will be necessary to discuss specifically what data to exchange, and the conceptual framework involved. After all, with bilateral data sharing, the data is obtained at the source – and therefore, the data will never be outdated.

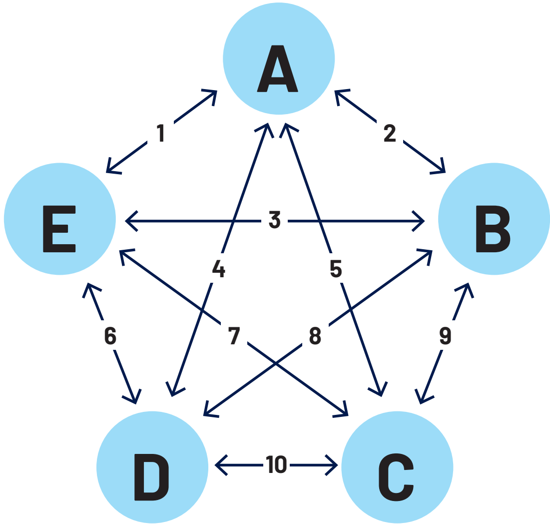


Figure 3 - Example of data sharing through (all possible) bilateral links between five parties

5.2 Data sharing via a central platform

Another form of data sharing is through a central platform, sometimes referred to as a 'hub'. A central entity provides the platform to which individual parties connect. Parties connected on the platform can share data with each other, greatly reducing the number of links for parties to share with each other – see the visualisation in Figure 3. Within logistics, two rather different variants of the central platform can be identified which will be discussed below: the port community systems (PCS) and the (commercial) platforms.

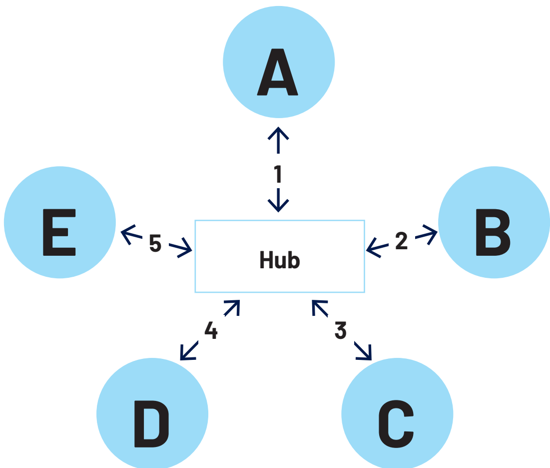


Figure 4 - Example of data sharing by linking five parties to a central data sharing platform

The two port community systems (PCS) for the Dutch main ports, Portbase (for the Dutch seaports) and Cargonaut (for Schiphol) are examples of central platforms. These community systems provide standardised information services for various processes in the mainport. The Hinterland Container

Notification (HCN) service is an example of such an information service. Hinterland carriers use it to pre-notify a visit to a seaport terminal, receive feedback on this, and share the exact loading/unloading list with the terminal shortly before the visit. It is important to mention that ownership of the data lies with the individual market participants at all times, despite central data storage. The PCS is merely a gateway for the data, and without explicit consent, it may not do anything with the data other than what was agreed at the time the service was concluded. However, precisely because of the central data storage, previously delivered data can be reused for other services, provided that this has explicitly been agreed. The main advantage of linking to a PCS, is that with one link, a party has basically established a link with all already linked parties. Connecting to Portbase means digital access to more than 5,000 parties, either manually via a web interface, or through an EDI (Electronic Data Interchange) or API (Application Programming Interface) system link. The strength of an existing community system is that many standard use cases are already realised and usable, precisely because there is such a large community of users. In addition, data security is also professionally set up.

Data sharing through a port community system (PCS) also has drawbacks. The pace of innovation, for instance, is relatively limited, as the PCS provider often has to interact with many parties, regarding questions such as desired functionality, process to be agreed, and data format, before being able to design a new functionality. At the same time, the PCS provider is asked for future functionalities and modifications to existing functionalities. Together, this makes the lead time to realisation of new services relatively long. Furthermore, in principle, only community services that are desired by multiple parties are developed within a PCS. Obviously, it is complicated to realise a unique service, or unique functionality, for just one container terminal. The basic principle that data within a PCS can, in principle, only be reused for purposes other than agreed upon after explicit consent is also seen as a limitation by many parties. Analytics applications such as pattern recognition and benchmarking, for example, are applications that would run excellently on a central platform. These disadvantages were pointed out during several interviews. One of the parties illustrated this by giving an example of an issue they are facing. By storing data centrally on the PCS, it is difficult to gain insight into, and analyse, queries of that data by other parties – after all, the queries take place on the PCS, not on one's own system. Of course, this kind of functionality can also be realised centrally within the PCS, but this will require some lead time.

A second form of a central platform provider is the commercial platform. These are parties that bring together supply and demand for logistics activities (such as transport) the same way a party like Uber does for taxi rides, booking.com for hotel rooms, and thuisbezorgd.nl for meals. Examples of large (high-growth) supply chain platforms include Flexport, Project44, Shippeo, Teleroute, Transfollow, and UTURN. These are international commercial platforms that offer functionality to chain parties and derive their revenue model from it, usually in the form of a percentage of the transaction value. The picture that emerges from the interviews is that many parties indicate that they expect platforms to bring about a lasting change in logistics business models. At the same time, many interviewees – as indicated earlier – articulate not seeing this as a positive development only, due to the shifting power in the chain, ('whose customer is it') and the underlying revenue model. To get a better picture, read Rens Lieman's book 'Uber voor alles' (Uber for everything) (2018), which gives an insight into how platforms in the consumer market add their value at a (high) price for participating companies.

It is worth remembering that in practice, parties connected to a central data sharing platform often have (many) more data links outside the central platform. For instance, bilateral links to major shipping companies, outside Portbase or a commercial platform.
Data sharing through a central platform is not going to fulfil all the wishes from port-logistics parties. In order to quickly create system links: this has partly been fulfilled. If both parties are already connected to the central platform, data sharing can be started quickly. If the parties are not yet on the platform, a

link will first need to be created, which requires lead time. If it concerns a data share service that does not yet exist, it is a different story, since it will require going through the process of designing and realising the data share service. Data sharing without a contractual relationship – the second wish – is in principle possible, provided there are services specifically designed for this purpose. This is the reason there are relatively few such services in practice, simply because most data-sharing services focus on exchanges between contractual partners and mandatory messaging towards, for example, the government. In terms of a common conceptual framework, parties using an existing data sharing service should implement the service's message format and underlying agreements, which, effectively, constitutes enforcement. The last point, the emphatic desire to be able to get data directly from the source, does not apply to a platform. After all, the platform is where information is collected and (potentially) enriched and can be obtained by other parties. As soon as something changes in the source, it needs to be communicated to the platform, which means the platform sometimes works with outdated information.

BLOCKCHAIN – WHAT WAS THAT AGAIN?

Blockchain is a distributed database solution, in which transactions can be stored. These can be all kinds of transactions, such as payments with a digital currency (like Bitcoin), but also the exchange of contracts or property deeds between parties, or the recording of other types of transactions. Technically, a blockchain is a system for storing data in a chain of data blocks, whereby the blocks cannot be changed afterwards. This system, which revolves around a so-called 'ledger', ensures that recorded transactions can always be retrieved unchanged, because in a blockchain, the same information is always stored in multiple places. This makes it virtually impossible to change anything after initial storage, unlike traditional databases. This allows a wide range of applications.

Several blockchain initiatives have been launched within logistics chains. Among others, in the field of global trade flows in order to unambiguously record things like origin, ownership and payments, within chains with – besides a number of large global players – also many small and changing parties. One example thereof is Tradelens, which was a joint initiative of Maersk and IBM. Another example are the various initiatives to unambiguously record the origin of products, such as cocoa and coffee.

A few years ago, Blockchain was quite a hype and it was proclaimed that this technology would replace all data storage (Iansiti & Lakhani, 2017). Meanwhile, the picture is more nuanced as blockchain does have some serious drawbacks, including the large amount of computing time required, which results in relatively long transaction times and high energy consumption. As a result, it is less suitable to apply in environments with high real-time transaction volume.

5.3 Federated data sharing

The previous two sections will have made it clear that both bilateral data sharing and data sharing via a central platform do not fully meet the data sharing wants and needs from the market. The simplicity of system linking is disappointing, it is difficult to structure data sharing with parties with whom no contractual relationship exists, an unambiguous conceptual framework is lacking, which regularly leads to confusion, and data is not always retrieved at the source.

Federated sharing was created from the desire for simplicity in sharing data.

Federated data sharing is another method to share data between parties that overcomes some of the main drawbacks of the methods discussed above. It was created from the desire to share data easily without human intervention, machine-to-machine so to speak. This does not involve setting up a central hub, or arranging a database or blockchain in which data storage takes place, but rather keeping the data as much as possible at the source where it can be retrieved (automatically). In this case, the data owner determines who is granted access to the data. This principle is called data sovereignty: the legally and technically binding information use restrictions. This mechanism can also be applied to data stored (centrally) with a delegated party. By arranging a number of essential matters, a machine-to-machine interface will be able to retrieve the data at the source. This concerns the following building blocks, which are also central to the Dutch Basic Data Infrastructure (BDI), which will be discussed in more detail later, as visualised in Figuur 5:

1. An identification and automatic authentication infrastructure will be used in order to establish that a party is actually who it says it is
2. A mechanism will be used to control authorisation of data, thus neatly determining which data can/cannot be shared
3. An unambiguous conceptual framework will be in place that unambiguously describes the data, thus avoiding any ambiguity about terms
4. A technical access point will be set up at the source for machine-to-machine communication without human intervention

By using these building blocks to share data, parties keep track of their own data, they can analyse what happens to the data, and the data is always up to date, after all, there is no need for synchronisation with a central system. Finally, the speed of innovation is also an important additional argument as parties themselves are and remain in control. If they want to set up new data sharing with chain partners because they collectively see the benefits of it, there will be no need for a central body, which often cause delays due to the pursuit of community-wide services.

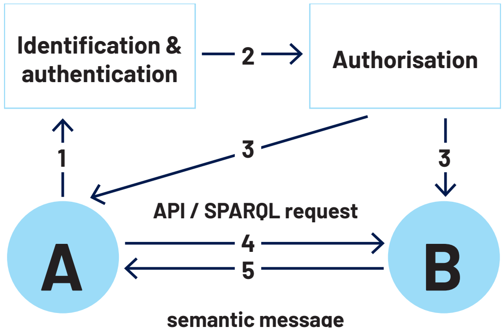


Figure 5 - Example of federated data sharing between two parties

Building block #1: identification and authentication

The first important building block for machine-to-machine messaging is the combination of automated identification and authentication. The first step is to establish the identity (of a person or system) by a trusted organisation: identification. Once this is done, automatic authentication can take place. This process verifies whether the identity actually corresponds to the user. A well-known example in the port of Rotterdam of identification and authentication is the Cargocard. For over 20 years, this personal smart card, combined with biometrics, has ensured a smooth, trouble-free and reliable logistics process.

Building block #2: authorisation

Successful identification and authentication is followed by authorisation to determine which data access rights have been assigned to this user. This can also be conditional access, in which case data becomes accessible to a particular organisation only when certain conditions are met. For example, a logistics provider is not granted access to the full information of a container until it is both unloaded at a sea terminal and customs has released the container.

ISHARE

Since 2017, iShare has been available to the Dutch logistics industry. iShare is a set of agreements for identification, authentication and authorisation that allows logistics parties to share data with each other under mutually established conditions. iShare did not only develop concepts, but also software required to implement the technical components of the set of agreements. Despite a series of pilots, adoption within Dutch logistics is as yet limited. It became apparent from the interview round that familiarity with iShare within port logistics was limited. A small proportion of interviewees said they were familiar with it, but also found it to be abstract and highly conceptual. People also find it difficult to understand how iShare can be applied concretely.

This is likely to change in the coming years. The various national and European innovation programmes around federated data sharing all foresee a role for iShare, precisely as a tool for identification, authentication and authorisation.

Building block #3: use of unambiguous terms within data exchange
For automated data exchange between systems, it is crucial that these systems can unambiguously interpret the data they exchange. Shared data definitions are crucial for this purpose, as the same term can have multiple meanings in different systems or different terms can be used in different systems while referring to the same term. This raises the question, what exactly does a specific data field mean, and how does it relate to other data fields? An order in one system may have a slightly different meaning in another system.

Semantic data models or ontologies are the technical solution for a unified conceptual framework. An ontology is a data structure that contains all relevant entities (usually an 'object') and their interrelationships and rules within a specific domain. An ontology is therefore a semantic network that groups a set of concepts that fully describe a domain. So-called 'triples' are used to make relationships between concepts legible to computers. In triple format, the relationship between data is translated into a structure of subject, object and predicate, see Figuur 6. The predicate is the description of the relationship between subject and object, so in this case, the fact that Maersk is the shipowner of the Maersk Emma (in 2006, the largest container ship in the world).

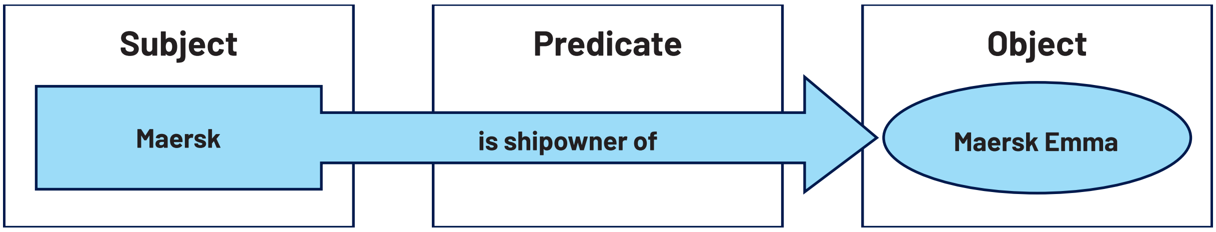


Figure 6 - Example of a Triple view of the relationship between shipping company Maersk and one of its container ships

LINKED (OPEN) DATA

Linked Data are structured data that are linked to other data and therefore more useful in semantic queries. Linked Data is based on the vision of growing the Internet into a global database so that data can be easily and widely (re)used. The concepts are based on a number of design principles aimed at making data available through a so-called Uniform Resource Identifier (URI). A URI is a unique designation of a 'resource', information, data or the like. One of the manifestations of a URI is the URL (uniform resource locator) which is used to access websites.

When Linked Data uses open data, it is referred to as Linked Open Data. Open data is a term used to describe freely available data. A good example is the Dutch government's open data portal, see https://data.overheid.nl, where more than 20,000 government resources are now externally accessible for reuse elsewhere.

One of the interesting applications of ontologies is that they make it possible to build industry (data) reference models, which describe the shared idiom within that specific industry and thus can overcome the problems of inconsistent definitions. An example thereof is the Airline Industry Data Model (AIDM) created by International Air Transport Association (IATA), which is finding wider and wider application within aviation.

An initiative was recently launched from the Top Sector Logistics to turn the (successful) Open Trip Model (OTM) into a fully-fledged semantic version, a logistics ontology. OTM is a shared data format for sharing (real-time) logistics transport data between parties. The model enables communication between different types of systems without complicated links or translations. This provides shippers and carriers with real-time information from municipalities and road authorities, and traffic centres with better insight into (actual) logistics movements.

Likewise, no semantic model for multimodal freight transport was available until recently. A first version of such a model is currently being developed within the European FEDeRATED project, which will be used in the Dutch BDI. UN/CEFACT is working hard on future semantic standards. Building on this, the Digital Container Shipping Association (DCSA) is working with a number of container shipping companies to develop standards for container sea transport, some of which may eventually be applicable to hinterland transport.

Semantic data models proved to be an interesting topic of conversation during the interviews. The data definition challenges appeared to be widely known in the sector, since almost all interviewees came up with examples from their own internal organisation and/or communication with chain partners. Several parties gave the feedback that they are currently running improvement programmes in this area, both internally and within the chain. Finally, almost everyone indicated that they see the development of 'standard' semantic models for specific application areas as a very interesting development.

Building block #4: use of technical gateways for machine-to-machine communication

The fourth building block of federated data sharing is the use of technical gateways (so-called 'endpoints') that enable automated access for systems, aka machine-to-machine communication. In short, the gateway where a computer of another party can retrieve data from the source in an automated way, after the steps of identification, authentication and authorisation have been successfully completed, of course. There are two variants of endpoints, 1) API and 2) SPARQL. An Application Programming Interface (API) endpoint is a gateway to access the functionality of a computer programme in another application, inside or outside the organisation. The other application does not know details of the functionality or specific implementation, but can use that functionality thanks to the API. APIs can be used to request information, but also to share information with another application, or to invoke specific functionality of another application. Many organisations today already make their data accessible to other organisations through one or more (web-based) APIs, using standard Internet protocols. The round of interviews shows that many parties are actively engaged in API development, and are using it more and more widely in their organisations.

Simple Protocol And RDF Query Language (SPARQL) endpoints are less well known. This is a query language used to query semantic data structures (such as linked datasets) through queries. A developer can use SPARQL to ask (very) detailed questions and receive answers (and data). In the interviews, it became clear that the existence of SPARQL is still hardly known among the interviewees. However, the general feedback is that people find this an interesting development as it enables advanced machine-to-machine interactions.

Innovation acceleration

The previously cited Basic Data Infrastructure (BDI) programme has created a blueprint for federated data sharing for Dutch logistics, fully incorporating the previously developed iShare concepts (Bastiaansen et al., 2020) (Boot, 2022). The DIL Growth Fund programme was set up to accelerate the roll-out of the BDI. DIL stands for Digital Infrastructure Logistics, a programme in which a number of living labs are set up. Applications of federated data sharing are developed, tested and improved in practice in the living labs. The existing living labs – with a significant overlap with port-logistics applications – cover the following themes: goods tracking system, arrival timing, confidence chain, transport tracking system, container tracking system, and multimodal (air cargo) chain. The programme team managing BDI and DIL also explicitly reaches out to interested parties to contact them, especially if they want to participate in one of the existing living labs, or have an idea for another living lab (for which space is reserved in the programme).

Besides this national innovation push, there is also a large-scale commitment from Europe to innovation programmes focused on federated data sharing. Europe sees federated data sharing as an important tool for parties to guard themselves against the rise of commercial platforms and leave power with the (existing) parties in the chain (Nagel & Lycklama, 2021). As far as Brussels is concerned, data remains at the source, and (automated) access to that data will be properly organised.

FEDERATED DATA SHARING AND THE FUTURE OF PORTBASE

Portbase is the party behind the Dutch port community system (PCS) for logistics flows to, through and from ports. It achieves this through a series of services focused on well-defined processes. Examples thereof include the mandatory communication with customs, or the message exchange around the visit of a container terminal.

The interviews have shown that Portbase is well aware that it will have to go along with the federated development, as for some services it may make much more sense to set up or develop them in a federated manner than centrally via the traditional PCS. Federated data sharing provides more speed, more confidence, and it allows for truly customised solutions. Not surprisingly, Portbase has been involved in this development since the conception of iShare. It has conducted several pilots, and is gearing up to take a major role within the BDI and DIL programmes.

In line with this, Portbase's latest strategy envisages a split into three types of service:

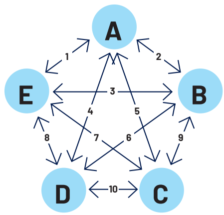
1. Port Solutions → the traditional Portbase PCS services (currently 42 different services)
2. Identity Solutions → implementation of building blocks #1 and #2 of federated data sharing: identification, authentication and authorisation services. Currently, iAMconnected is available for this purpose, this service will become fully iShare compliant in the not too distant future.
3. Platform Solutions → data access via Portbase, for example, of data already in the traditional PCS services, but also road transport information access from, for example, DEFlog is an example of this.

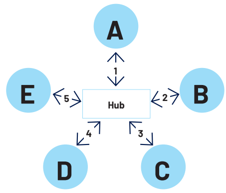
It is not inconceivable that the number of traditional PCS services will decrease in the near future, and Portbase will take a major role within the federated landscape. Reasoning from the use cases behind data sharing, the previously discussed split into contract, compliance and coordination, the following possible picture emerges. Data flows from a compliance perspective, often motivated from a governmental task (such as customs, food safety, and hazardous substances legislation), are obvious to be facilitated centrally via the PCS. The same applies to (some of the) contract-related data flows. Data sharing motivated from a coordination perspective, on the other hand, might be more logical in a federated manner, as it involves many links, and parties all have their own requirements in terms of data to be shared and associated conditions.

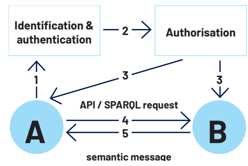
5.4 At a glance

This white paper started from the central question: data sharing, what, how and why? Putting everything together, it is clear that data sharing can answer several challenges around digital collaboration between parties in the logistics chain. Data sharing is going to help make logistics smarter, better and greener. It is important to be aware of the fact that the data to be shared comes in many forms. There are also different types of data sharing, all with their advantages and disadvantages, see the overview based on the above summarised in Tabel 1.

Table 1 - Three types of data sharing compared







	Data sharing	Bilateral	Central platform	Federated
Advantages and disadvantages	Advantages	• Easy to implement for limited number of links • Trust is not an issue	• Easy to connect many parties • Wide range of standard services • Secure data and data communications • Maintenance-friendly for participants	• Data at the source • Scalable • Open to all (based on a specific set of agree-ments) • Low risk of lock-in due to level playing field and open standards
	Disadvantages	• Complicated and time-consuming to scale up • Management issue (due to many links)	• Limited pace of innovation • Focus is on community services (hard to deviate from) • Follow data-sharing rules • (sometimes) Commercial revenue model	• Technology still under development • Few semantic industry standards available.
Port logistics market needs	Quick linking	No; requires agreement on the what and the how	In part; linking to central platform takes time; message exchange start-up between already connected parties easy	Yes; machine-to-machine prepared; linking to new party is simple
	Data sharing without a contrac-tual relationship	Complicated	Complicated	Relatively easy
	Unambiguous conceptual framework	No; must be discussed specifically	Yes; enforced by message format standard used	Yes
	Data directly from the source	Yes	No	Yes

Bilateral data sharing is easy and quick to set up, if it concerns one or a few links, but is difficult to scale up. Data sharing through a central platform reduces the number of links needed, but limits the possibili-ties of rapid innovation. Finally, federated data sharing is the big newcomer, with a number of promising concepts, however, many concepts will have to prove themselves in practice in the coming years. It is, however, clear that innovation is currently accelerating, with many opportunities to participate in innovation projects.

-06-
Getting started

All things considered, it can only be concluded that it makes sense for organisations to take steps with data sharing. But where exactly do you start? By reading these five recommendations on how to get started.

[1] Work on digital awareness of the organisation. The world is becoming increasingly digitised, and as described earlier, the logistics world is moving with it. New technologies keep emerging, with increasingly brighter promises. At the same time, it also seems fleeting sometimes, and it's back to waiting for the next new concept, idea or buzzword. The question, however, is whether parties can continue to afford to wait and see. Digitalisation goes far beyond traditional IT, offering opportunities to change and improve business processes. It is therefore advisable to become more digitally aware as an organisation. What is out there? What are the possibilities? What is ahead? Digital awareness within the organisation helps set the strategic direction, based on the new (digital) opportunities. So work on the digital competences of people within the organisation, and commit to digital leadership. Crucial foundations before starting data sharing.

[2] Think in processes, not in technology. Data sharing requires technology, as this white paper has shown, and as referred to in the first recommendation. However, it is perhaps even more about taking a critical look at processes. What processes are in place in organisations? What is the potential of reusing data from elsewhere in the chain? Is a redesign required? Take a close look at your own processes and their role in the chain and ask questions like: which party has which data available, which party has which data needs, and why is this data exchanged/not exchanged? The examples and cited references discussed in this white paper may serve as inspiration in this regard.

[3] Get your own data management in order. Too many organisations do not have their data management sufficiently in order, which became clear in the round of interviews. For example, what about the quality of the data being recorded within the organisation? Is data collected (partly) manually? What validations are performed on this data? Is someone responsible for data quality within the organisation? Another topic is data definitions: what is meant by a particular term? Is this used consistently within our own organisation? And how does this work from a chain perspective? Do chain partners use the same definitions? Proper data management is crucial when data is shared more intensively, but also to become more data-driven. So take a close look at the ownership of the data, work on data quality, and appoint so-called data stewards within the organisation. Data stewards are those responsible for ensuring data quality, data use and the corresponding security guidelines.

[4] Seize the innovation opportunities out there and start sharing data. As the previous section showed, the innovation engine around federated data sharing is gathering steam. The Netherlands and Europe are betting big on federated concepts, based on the vision that these will help make a fist against large international tech platforms. It is, of course, still innovation, and not everything has crystallised yet. However, the programmes being set up offer an interesting opportunity to actively participate at the forefront of this innovation. Seize that opportunity, start experimenting and innovating, and start extracting value from data sharing. A possible starting point is to start using identities widely within data sharing applications by applying identification, authentication and authorisation.

[5] Don't do it alone. Collaborate with other parties. For instance, directly with chain or technology partners, possibly within an ongoing innovation programme. Also seek collaboration with knowledge institutions. Knowledge institutions can participate in innovation projects, for instance in the form of researchers and interns or graduates. Working with graduate students is a great way to engage with a topic like data sharing in a low-key way.

In short, it's time to take action, and start shaping the digital future.

Appendix A

List of sources consulted

1. Bastiaansen, H.J.M., C.H.M. Nieuwenhuis, G. Zomer, J.P.S. Piest, M. van Sinderen, S. Dalmolen, W.J. Hofman, "The logistics data sharing infrastructure", TKI Dinalog, August 2020
2. Boot, S. (2022) "Basis Data Infrastructuur - Een afsprakenstelsel voor event-gedreven coördinatie in de logistiek", 29 June 2022, fact sheet Ministry of Infrastructure and Water Management
3. Blackburn, S., J. Galvin, L. LaBerge, E. Williams (2021) "Strategy for a digital world" McKinsey Quarterly, October 2021
4. CGI (2022). "Top trends to watch in Transport & Logistics" – see **https://www.cgi.com/en/transportation-logistics/voice-of-our-clients**
5. Galbraith, J. R. (1974). Organization design: An information processing view. Interfaces, 4(3), 28-36.
6. Horst, M., Kansen, M., & Moorman, S. (2019). Trends en hun op invloed zeehavens. **https://www.kimnet.nl/publicaties/rapporten/2019/07/8/trends-en-hun-invloed-op-zeehavens**
7. Iansiti, M., & K. R., Lakhani. (2017) The Truth about Blockchain. Harvard Business Review. January-February 2017, pp 118-127
8. IATA. (n.d.). Airline Industry Data Model. **https://www.iata.org/en/programs/workgroups/passenger-standards-conference/architecture-technology-strategy/industry-data-model/**
9. Lieman, Rens (2018). Uber voor alles. Business Contact. EAN 9789047011101
10. Nagel, L., D. Lycklama (2021), "Design Principles for Data Spaces", OPENDEI, position paper version 1.0, April 2021 **https://design-principles-for-data-spaces.org/**
11. Premkumar, G., Ramamurthy, K., & Saunders, C. S. (2005). "Information processing view of organizations: An exploratory examination of fit in the context of interorganizational relationships." Journal of Management Information Systems, 22(1), 257-294)
12. Sebastian, I.M., M. Mocker, J.W. Ross, K.G. Moloney, C. Beath, N.O. Fonstad (2017), "How Big Old Companies Navigate Digital Transformation" MIS Quarterly Executive; September 2017 (16:3), pp 197-213
13. Snijders, D. (2017). Data is macht – Over Big Data en de toekomst. Stichting Toekomstbeeld der Techniek, The Hague (ISBN 978-94-91397-15-8)
14. Wagter, H. (2021), "Introductie federatief informatie delen", August 2021, Connekt / Top Sector Logistics
15. Wikipedia (NL / international) for various definitions in the list of abbreviations and terms

Appendix B
List of abbreviations and terms

Set of agreements	A set of agreements includes joint agreements between parties for data sharing. Having these agreements in place creates trust, hence the name 'trust framework'.
AIDM	Airline Industry Data Model. Data reference model for aviation prepared by (see) IATA.
API	Application Programming Interface. A technical gateway to access the functionality of a computer programme in another application, inside or outside the organisation.
Authentication	The process of verifying whether the identity of a user actually corresponds to this user.
Authorisation	The process by which a user (natural person, or machine) is granted rights to access certain data.
BDI	Basic Data Infrastructure. A set of agreements for event-driven coordination in logistics. Within the (see) DIL programme, the principles of BDI will be rolled out and tested in the coming years.
Bilateral data sharing	Form of data sharing where one organisation shares data with one other organisation through a direct (bilateral) link.
Blockchain	Blockchain is a distributed database solution, in which transactions can be stored immutably. Data is stored in a chain of blocks, whereby it is guaranteed that the blocks cannot be changed afterwards. This is ensured by the fact that the blockchain 'ledger' is kept at multiple locations.
Data	Data comprises structured or unstructured information / facts about business processes, products and customers, available in digital form. Data needs to be put in context to be used as information.
Data sharing	Making an organisation's data available and accessible to one or more other parties (see also the forms Bilateral data sharing, Central data sharing platform, Federated data sharing).
Data sovereignty	The legally and technically binding data use restrictions.
Digital Twin	A Digital Twin is a virtual representation of a physical object. By turning a physical object into a digital counterpart and keeping it up to date, it is easy for applications to use up-to-date information from the physical world in software applications.
DIL	The programme Digital Infrastructure Logistics from the Dutch government with the main objective of setting up a series of federated data sharing living labs based on the principles as they have been and are being developed within the (see) BDI programme. The DIL programme has an intended duration of 2022-2024, with an option to extend for two years (in that case, running until 2026).
eCMR	The digital bill of lading, which can be used instead of the traditional paper CMR bill of lading.
EDI / EDIFACT	Electronic Data Interchange. EDI is an (older) technique for electronic message exchange. It is used for repetitive transactions since the documents to be exchanged must comply with standard templates, usually developed by standardisation organisations such as UN/EDIFACT.
ERP	Enterprise Resource Planning. ERP (software) systems provide support for core processes within an organisation.
ETA	Estimated Time of Arrival

ETD	Estimated Time of Departure
FEDeRATED	FEDeRATED is a European Commission-funded applied research programme working on federated data sharing concepts within logistics applications, among others, during 2019-2023.
IATA	International Air Transport Association
Identification	The process by which the identity of a person or (in the case of federated data sharing) a computer system is established and linked to a known identity.
Internet of Things	The collective network of devices ('things') that interact with other devices or systems and exchange data with them via Internet connections.
Information:	Contextualised data that gives it value for decision-making or direction, for example, within an organisation.
iShare	A set of agreements for identification, authentication and authorisation that allows logistics parties to share data with each other.
Federated data sharing	Federated data sharing refers to automated data sharing between parties, according to an established set of agreements. Data sovereignty is paramount to this. It is based on an infrastructure for automatic identification, authentication and authorisation. Data is typically exchanged through API and/or SPARQL endpoints and takes the form of semantic model or ontology form.
Platform	A platform is a central entity to which individual parties connect. This white paper distinguishes between community systems, such as (see PCS) Portbase and Cargonaut, and commercial platforms that bring together supply and demand for logistics activities (such as transport).
Ontology	See semantic model.
OTM	Open Trip Model. OTM is a shared data format for sharing (real-time) logistics transport data between parties.
PCS	Port Community System (also see Portbase)
Portbase	Portbase is the Dutch Port Community System that connects parties in the logistics chains of the Dutch ports. Portbase facilitates data sharing between companies and information exchange with public authorities to work faster, more efficiently and at lower cost. Portbase is neutral, belongs to and is intended for the port community and is a non-profit organisation.
Query	A command to a database to trigger an action.
Semantic model	A data structure containing all relevant entities and their interrelationships and rules within a specific domain. It describes the shared vocabulary (of concepts and interrelationships) to exchange information between different stakeholders. Often also referred to as ontology.
SPARQL	Simple Protocol And RDF Query Language. A query language used to query semantic data structures (such as linked datasets) through queries.
XML	eXtended Markup Language. A syntactic standard that allows structured data to be represented in the form of plain text. To this end, it uses so-called elements and attributes to structure data. This structure makes the files both machine-readable and human-readable. XML is used to store data (as in the OpenDocument format) and to send data.

Appendix C
List of persons consulted

Name	Organisation	Interview	Expert	Review
Willemien Akerboom-van der Win	ECT	X		
Huibert Alblas	CGI / Top Sector Logistics		X	
Donald Baan	Portbase	X		
Martijn Bekke	CGI			X
Roy van den Berg	SmartPort			X
Sjoerd Boot	Ministry of Infrastructure and the Environment		X	
Ries Bode	digiGO / iShare		X	X
Michel van Dijk	Van Berkel	X		
Simon Dalmolen	TNO		X	X
Richard Donselaar	Pilotage services	X		
Niels Duijvestijn	UTURN	X		
Paul Ham	ECT	X		
Almar van Herk	Kotug	X		
Wout v/d Heuvel	TLN / SUTC	X		
Judith Hofstede	Grolsch	X		
André Ijsselstijn	Kühne+Nagel	X		
Michel Jansen	Total Produce	X		
Dirk Koppenol	SmartPort			X
Rob Kraaijenbrink	Vopak	X		
Anique Kuijpers	SmartPort			X
Alwin Kulsdom	Grolsch	X		
Laurens Lapré	CGI		X	
Dennis Lenting	Pilotage services	X		
Iwan Maessen	Portbase / DILOG			X
Bart van Riessen	Poort8	X		
Tim van Soest	CGI / Top Sector Logistics			X
Dalibor Stojakovic	NPRC	X		
Oscar van Veen	Port of Rotterdam Authority	X		
Ilona Verveer	UTURN	X		

Name	Organisation	Interview	Expert	Review
Herman Wagter	Top Sector Logistics		X	
Sifra Westendorp	Anthony Veder	X		
Rob Zuidwijk	Erasmus University Rotterdam		X	X

We would like to thank Rob Zuidwijk, Roy van den Berg, Anique Kuijpers, Dirk Koppenol, Ries Bode, Simon Dalmolen, Iwan Maessen, Tim van Soest and Martijn Bekke for their critical reviews of earlier versions of this white paper

Colophon
©SmartPort
December 2022

Design: IJzersterk.nu
Photography: Shutterstock

All information included is the property of SmartPort. Reproduction of content, in whole or in part, is permitted provided that reference is made to the source.

Disclaimer
SmartPort has taken the greatest possible care in compiling this document. Nevertheless, SmartPort accepts no liability for any inaccuracies in the information, nor for any damage, nuisance or inconvenience or other consequences arising from or related to the use of this information.

---

connecting knowledge

DO YOU HAVE ANY QUESTIONS?

SmartPort
info@smartport.nl
tel +31 (0)10 402 03 43