Title:	Document Version:
D2.1 Industrial Data Platforms and seaport community requirements and challenges	1.0

Project Number:	Project Acronym:	Project Title:
H2020-871493	DataPorts	A Data Platform for the Cognitive Ports of the Future

Contractual Delivery Date:	Actual Delivery Date:	Deliverable Type*-Security*:
M15 (March 2021)	M15 (March 2021)	R-PU

*Type: P: Prototype; R: Report; D: Demonstrator; O: Other; ORDP: Open Research Data Pilot; E: Ethics.

**Security Class: PU: Public; PP: Restricted to other programme participants (including the Commission); RE: Restricted to a group defined by the consortium (including the Commission); CO: Confidential, only for members of the consortium (including the Commission).

Responsible:	Organisation:	Contributing WP:
Achilleas Marinakis	ICCS	WP2

Authors (organisation):	
Achilleas Marinakis (ICCS)	Héctor Iturria (PRO)
Anthousa Karkoglou (ICCS)	Ignacio Márquez (PRO)
Anastasios Nikolakopoulos (ICCS)	Pablo Giménez (VPF)
Nikos Tzagkarakis (ICCS)	Miguel Llop (VPF)
Hanane Becha (TRX)	Francisco Valverde (ITI)
Grigorios Dimitriadis (ThPA)	Paolo Calciati (ITI)
Elsa Koukouloudi (ThPA)	Santiago Cáceres (ITI)
Dimitris Tsitsamis (ThPA)	Javier Navarro (EVR)
Shayan Ahmadian (UKL)	Fernando Sanz de Galdeano (EVR)
Konstantinos Votis (CERTH)	Andreu Belsa (UPV)
Sofia Terzi (CERTH)	Inna Skarbovsky (IBM)
Vasileios Siopidis (CERTH)	Andreas Metzger (UDE)
Evi Manganopoulou (CERTH)	

Abstract:

This deliverable is a report describing the State-of-the-Art in related technologies and the benchmarking of existing data platforms related to the use case providers, together with an evaluation of requirements and challenges of the port community in general and the proposed pilots.

Keywords:

State-of-the-Art, benchmarking, requirements

Revision History

Revision	Date	Description	Author (Organisation)
V0.1	14.02.2020	First version of the document	Achilleas Marinakis (ICCS)
V0.2	03.07.2020	Added contribution from UKL on SotA analysis section	Shayan Ahmadian (UKL)
V0.3	31.08.2020	Added contribution from ICCS on SotA analysis section	Achilleas Marinakis (ICCS) and Anthousa Karkoglou (ICCS)
V0.4	06.10.2020	Added contribution from PRO on SotA analysis section	Ignacio Márquez (PRO) and Héctor Iturria Sánchez (PRO)
V0.5	13.11.2020	Added requirements elicitation section	Achilleas Marinakis (ICCS)
V0.6	11.01.2021	Added infrastructure benchmarking section and updated document with ThPA comments	Anthousa Karkoglou (ICCS), Ignacio Márquez (PRO), Héctor Iturria Sánchez (PRO), Shayan Ahmadian (UKL), Grigorios Dimitriadis (ThPA) and Pablo Gimenez (VPF)
V0.7	02.02.2021	Added contribution from TRX on infrastructure benchmarking section	Hanane Becha (TRX)
V0.8	10.02.2021	Updated document with VPF comments	Achilleas Marinakis (ICCS), Anthousa Karkoglou (ICCS), Shayan Ahmadian (UKL) and Héctor Iturria Sánchez (PRO)
V0.9	16.03.2021	Updated requirements elicitation section	Achilleas Marinakis (ICCS) and Anthousa Karkoglou (ICCS)
1.0	30.03.2021	Final version	Achilleas Marinakis (ICCS) and Anthousa Karkoglou (ICCS)

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement № 871493.

More information available at https://DataPorts-project.eu

Copyright Statement

INDEX

LIST OF FIGURES

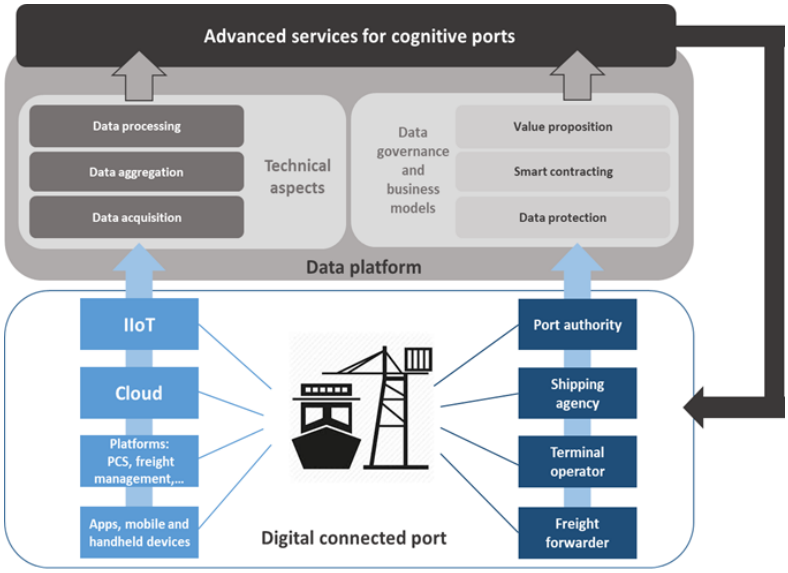
LIST OF TABLES

1 INTRODUCTION

1.1 DATAPORTS PROJECT OVERVIEW

DataPorts is a project funded by the European Commission as part of the H2020 Big Data Value PPP programme, and coordinated by the Technological Institute of Informatics (ITI). DataPorts rely on the participation of 13 partners from five different nationalities. The project involves the design and implementation of a data platform, its deployment in two relevant European seaports connecting to their existing digital infrastructures and addressing specific local constraints. Furthermore, a global use case involving these two ports and other actors and targeting inter-port objectives, and all the actions to foster the adoption of the platform at European level.

Hundreds of different European seaports collaborate with each other, exchanging different digital data from several data sources. However, to achieve efficient collaboration and benefit from AI-based technology, a new integrating environment is needed. To this end, DataPorts project is designing and implementing an Industrial Data Platform.



The DataPorts Platform aim is to connect to the different digital infrastructures currently existing in digital seaports, enabling the interconnection of a wide variety of systems into a tightly integrated ecosystem. In addition, to set the policies for a trusted and reliable data sharing and trading based on data owners' rules and offering a clear value proposition. Finally, to leverage on the data collected to provide advanced Data Analytic services based on which the different actors in the port value chain could develop novel AI and cognitive applications.

DataPorts will allow establish a future Data Space unique for all maritime ports of Europe and contribute to the EC global objective of creating a Common European Data Space.

1.2 DELIVERABLE PURPOSE AND SCOPE

Specifically, the DoA states the following regarding this Deliverable:

This deliverable will be a report describing the State-of-the-Art in related technologies and the benchmarking of existing data platforms related to the use case providers, together with an evaluation of requirements and challenges of the port community in general and the proposed pilots in particular.

The purpose of this document is to introduce the State-of-the-Art analysis, the infrastructure benchmarking, and the requirements elicitation, which are going to define the course of the project throughout its lifespan. The objectives of the deliverable are firstly to study the existing platforms that are found in the literature as well as the platforms that are already provided by the end users of the piloting partners in the maritime domain in relation to all the other domains that DataPorts is going to address. Secondly, to elicit and evaluate the requirements from port community and pilots for each aspect of the DataPorts project, especially those that are going to be utilized as incentives when designing the platform's architecture.

1.3 DOCUMENT STRUCTURE

This deliverable is broken down in the following sections:

- **Section 1** includes the introduction of the document which entails the description of the scope, purpose and structure of the present document. In addition, document dependencies and context are presented to the reader.
- **Section 2** includes the State-of-the-Art analysis that is found in contemporary scientific literature. This section also includes a State-of-the-Art analysis of the commercial products and technologies that are already in use by ports. Furthermore, Section 2 includes a presentation of related H2020 EU projects that would be potentially contacted for collaboration with DataPorts.
- **Section 3** includes the comparative analysis and benchmarking of the infrastructure that is already in use by the pilot ports of Thessaloniki and Valencia, and also the required technologies in other ports where Posidonia Operations is deployed, a software solution that will be integrated in the pilots as part of the global use case.
- **Section 4** presents the methodology according to which the requirements of DataPorts are aggregated and organized. Moreover, the statistics that stem from these requirements are presented and analysed here.
- **Section 5** describes the conclusions that are drawn from the State-of-the-Art analysis, the infrastructure benchmarking, and the requirement elicitation process that are presented in this deliverable.

Annexes

- **Annex 1 (Section 7):** Requirements at WP Level

2 STATE OF THE ART ANALYSIS

As part of the European landscape, DataPorts should identify and evaluate the State-of-the-Art solutions, and methodologies that currently address the industrial needs of a port's ecosystem with respect to Big Data solutions. This analysis is performed in a three-fold manner, where its first aspect concerns the survey of scientific literature regarding the related State-of-the-Art technologies in the areas of Smart Ports, Big Data, and Blockchain. The second aspect aims at the documentation of current commercial products and technologies that are available for the implementation of port processes and functionalities. Lastly, the final aspect of the analysis is considered the enlisting of European projects and activities related to data-driven innovation, including European projects that target ports and maritime applications. In this chapter are presented the most significant European projects alongside their information including their objective to identify the current research projects that could lay the foundation for a future collaboration with DataPorts. In this way, the technological and innovation map within which the project's research and development takes place is identified.

The State-of-the-Art analysis, as described above, provides a broader understanding of the functionalities and attributes that the DataPorts Platform should exhibit to be adopted widely by the industry. Moreover, it highlights fruitful areas of development as far as the DataPorts Platform is concerned, while indicating potential limitations and hurdles that might occur. Overall, this multi-level analysis describes extensively the technological and scientific background that DataPorts is going to be built upon.

2.1 SCIENTIFIC LITERATURE

In literature there is a plethora of sources and scientific publications that describe technological advances and their implementation within a port's ecosystem. The following analysis, after the establishment of the seaports' significance, is focused firstly on the selected Big Data model that depicts and describes the maritime and freight data reserves. Next are presented the different aspects of maritime data and the applications that stem from their leverage with primary applications in the energy optimization and chartering operation areas. Lastly is underlined the necessity for secure storage and exchange mechanisms of the maritime and port data and is presented the Blockchain technology as a powerful tool in this direction.

Seaports constitute a cornerstone of the global economy, as they operate as the connecting tissue between all other modes of transport. Furthermore, they play a detrimental role as the pivot of economy concerning the wider region that they are in [1]. Port activity in terms of international shipping trade has exploded in the recent decades, resulting in the dynamic evolution of the harbours and the increasing degree of complexity of their management [2]. In support of the port management process under transparency, data-driven intelligence is introduced. Since ports do not remain static, their evolution can be evaluated against size, connectivity, as well as operational capacity. As such, the decision-making process regarding port investment and policy establishment needs to take into consideration port evolution over time and space [1]. The latter can be achieved through the aggregation of miscellaneous data that can be harvested from the port itself, vessels, as well as the stakeholders that are involved in the operations of the port.

The data analysis mentioned above addresses different aspects of the port operations' lifecycle that require to be inspected and managed. The maritime domain encloses data sources of a high magnitude that include data about vessels, oceans, wave stations, observations from environmental conditions, fishing and maritime biodiversity, routes, and trajectories, and incidental or voluntary oil spill events [3]. The challenges imposed by the maritime data need to be faced for knowledge to be extracted through a sustainable analysis process. In terms of Big Data, the data can be described in terms of the 5V's model. In more detail, the five V's that consist of a definition model for Big Data are described below [3] [4]:

- **Volume** represents the magnitude of data that is generated and collected.
- **Velocity** denotes the pace at which the data flows.
- **Variety** indicates heterogeneity in data types, formats, structuredness, and data generation scale.

- **Veracity** refers to noise and quality issues in the data.
- **Value** denotes the value that can be obtained from processing and mining Big Data.

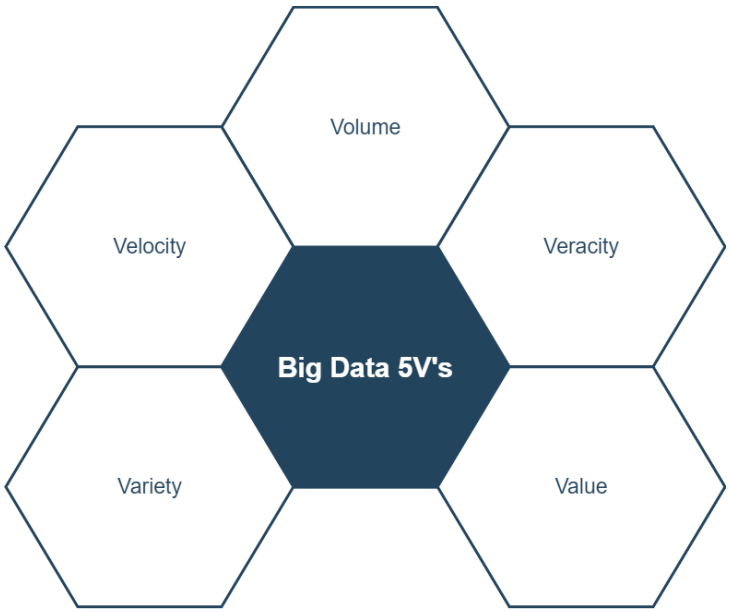


Figure 1 – Big Data 5V's

According to the metrics mentioned above, the maritime data are characterized as depicted below:

- **Volume:** The volume of data that is generated from the existing maritime data sources corresponds to an enormous scale. Terabytes of data on a daily basis are collected from public websites such as NOAA, NASA, and NODC that allow access to a plethora of oceanographic and maritime biodiversity datasets. Moreover, enormous databases containing information about live and historical ship conditions and trajectories are accessible in different maritime company websites. It is recorded that data to perform live maps that originates from satellites and marine databases can be estimated to correspond to more than 18 million vessels and ports, and 800 million vessel positions recorded monthly [3].
  Concerning Smart Ports and the interrelated Internet of Things (IoT) technology, the rapidly growing presence of smart logistics mechanisms and smart sensing systems in ports realize the automation of the port's functionality and the generation of large amounts of data. The range of the generated information is wide while involving a plethora of communication, sensor systems and control technologies to facilitate data collection throughout the supply chain for decision-making in real time and information sharing with the various stakeholders of the port [5].
  Sensors, bar codes, ERP, transport management systems and database technologies amongst others generate bulks of data on the scale of Zettabytes (ZB) or even Yottabytes (YB) [6].
  There is a variety of storage approaches as far as Big Data generated from the supply chain industry is concerned, including rational database management systems (RDBMS) of auxiliary systems such as servers and database management systems, Direct Attached Storage (DAS), as well as Network Storage (NS). The latter can be encountered either as Network Attached Storage or Storage Area Network and are both directly attached to the network, thus enabling a unified network interface platform for real-time data sharing and accessibility. DAS consists of different types hard-disks/hard-drives, which are directly attached to the DBMS [7].
- **Velocity:** Maritime regulations in force establish that ships are obliged to collect and store large amounts of data that describe information such as location, speed, trajectories, communication audios and engine, rudder, and thruster status. In this case, data recording should be continuous,

while it should also be retained in the vessel's devices for decision making and analytics. Moreover, there is a vast providing of real-time data about meteorological and oceanographic conditions that are collected via the Geostationary Operational Environmental Satellites (GOES) and commercial satellites. It is recorded that the NOAA National Centre for Environmental Information (NCEI) only houses over 20 petabytes of atmospheric, coastal, oceanic, palaeoclimatological, and geophysical data, and archives over 26 terabytes of oceanic, atmospheric, and geophysical data each month. Moreover, maritime companies make available live maps reporting trajectories and location of more than half million vessels all over the world [6].

As far as Supply Chain management is concerned, the velocity of data relies on the speed that data is collected, the efficiency of data storage, the added value that the knowledge extraction provides, algorithms and decision-making models [8]. The agility of data is proven to increase the efficiency of the processes in the supply chain. Assuming a multi-level supply chain, as the number of intermediaries increase, the latency of the data reaching the last actor in the channel is getting bigger. Therefore, multistage supply chains, such as the ones that are present within the port's ecosystem, exhibit room for advancement [9].

- **Variety:** Vessel data is collected from a series of different devices and in different formats e.g., in-situ Automatic Identification Systems (AISs) receive data about locations from navigational satellite systems, collect information from other ships or ports through radio communication systems, while a ship's engine performance is monitored by several sensors. In similar fashion, oceanographic data is collected through time-series measurements at a fixed location via instrumentation arrays, or with other types of equipment, e.g., surface drifters or gliders can freely move and gather oceanographic data. Moreover, sensors affixed to orbiting satellites can continuously gather data about the ocean [3].

  The collection of maritime data is performed through a variety of ways and stored in many different formats and data structures [3]. The harvested data might be structured, unstructured or semi structured, while its format could be just as variant, e.g., XML or video and SMS. This polymorphism includes data from diverse sources such as retailers, distributors, suppliers, inventory, sales, consumers etc. [6]. The data aggregation in the supply management chain includes the upstream and downstream sources. The latter includes logistics, distribution, or retailer side, while the upstream source includes supplier's side [10].
- **Veracity:** Disparities between the instruments of data aggregation, as well as different degrees of accuracy and uncertainty that can be observed in maritime data sources significantly affect the veracity of maritime data. For instance, poorly trained personnel might lack the skills to correctly calibrate and utilize the measurement instruments on vessels, while oceanographic buoy measurements may be impacted by buoy size, shape, ballast, and mooring. Thus, inconsistent data that contain missing or inconsistent observations and ambiguities may come across maritime data sources [3].

  Furthermore, data veracity plays a detrimental role in the maintenance of perishable goods, which is by itself a key challenge for logistic companies in the food manufacturing sector [8]. For instance, smart containers carrying food products are equipped with sensors that are constantly collecting and monitoring data, such as temperature, humidity, door status etc.

  In terms of achieving quality of maritime data, quality assessment methods like deduplication, disambiguation, and domain specific data cleaning process are highly significant [3].
- **Value:** Big Data in the maritime industry can be applied to a number of diverse application areas in the industry. Under these maritime data-driven applications can be enlisted operations of schedule management, fleet allocation, chartering, condition monitoring and maintenance, hull and propeller cleaning, safe operation, as well as energy saving operations.

  Diving deeper into some key application areas, the chartering operations include the optimization of

vessel selection for cargo shipment, which is highly dependent on data provided in the form of AIS information, position reports, estimated time of arrival, vessel technical specifications etc. In this way, available alternatives and freight forecasts are of high reliability. Concerning voyage operations, terminal operators, voyage managers or port agents need the information of the Estimated Time of Arrival (ETA) and cargo information [11] . In this way, any deviations from expected performance can be forecasted and managed in real-time with the aid of aggregated data or data streams. The most significant function of charterers is the selection of the most suitable ship for cargo at the most economical price. Detrimental role to this task plays the information provided by brokers and ship owners, which is nevertheless limited. However, Big Data analytics can provide charters with readily available, accurate and actionable information in terms of improving the decision-making process. All the available information can be integrated into an exchange portal that provides all the feasible alternatives alongside freight forecast [11].

Furthermore, vessel operations regarding speed and maintenance are concerned with several decisions that affect the correlation between optimum vessel speed and fuel consumption and the vessel maintenance. The former can be assisted by big maritime data analytics to define the optimal balance between optimal speed and fuel consumption, while taking into consideration bunker costs, freight rates and schedules. Ships exhibit optimum speeds and various tests are conducted at the time of vessel delivery to determine the optimum speed for fuel consumption. Nevertheless, operating a vessel at its optimum speed is not easily achievable since its calculation varies due to a variety of factors, such as engine wear and maintenance.

On the other hand, the decision process regarding vessel maintenance, including hull cleaning and propeller polishing, is based upon intuition or scheduling tactics and not on actual vessel performance. In this direction, fuel consumption data can be leveraged for a cost-benefit analysis to be performed in terms of vessel maintenance. Thus, data analytics are of great significance in the cost-benefit analysis of vessel maintenance for the timing and the benefits of performing the actual maintenance to be accurately defined [11].

Focusing on the energy optimization of the port's infrastructure as one of the enhancements that a Big Data approach can achieve, it has been documented that a port's operational efficiency depends to a high degree on the efficiency with which the available resources are managed. In more detail, there seems to be a positive correlation between the optimization of the operations mentioned above (i.e., reduction of the transport times and transport time of containers in the yard) and energy efficiency [12]. Specifically reducing port stay times results in the reduction of ship sailing speed at sea [13], which in turn adds energy savings up to 25.4% [14]. Moreover, virtual arrival, as in reduction of the approach speed due to port congestion, is also examined as an energy related aspect of ports [15]. Speed optimization and virtual arrival are also extended to the energy study of ship routes [16].

In the yard side, the energy planning addresses mainly the transport and stacking of containers [12]. The yard allocation problem and yard handling equipment planning are solved within an energy-aware frame and is transformed to a variant of the vehicle routing problem [17]. Recent energy-aware studies have focused on automated container terminals, which as a predictive control model is provided for balancing the throughput and energy consumption of resources [12]. Energy-aware planning focuses also on gate operations and inter-terminal transport in the landside [12]. Small shifts of truck arrival times at the port gates contribute significantly to the reduction of energy consumptions and truck emissions [18]. Peak workloads of trucks can be reduced with an intelligent inter-terminal transportation schedule [19].

As far as containerized reefer trade is concerned, the continuous refrigeration of each container to keep the products cool, has been gaining terrain in the market segment of shipping in the recent years [20]. The determination of the number of plugs for reefers, the location of reefer area in terms of minimizing travel distances, the formulation of a power plan for each reefer cargo, the optimization of the electrical distribution systems design and the exact measurement of the energy consumption for reefer containers have been selected as important research perspectives for the energy efficiency [21]. As a conclusion of the above, energy management systems consist of energy

demand and energy supply planning, as well as smart energy management system linking demand and supply. Efficient implementation of energy management systems includes the ability of ports to measure and estimate the energy consumption of its facilities properly [12]. All the above should be included and leveraged by an energy management strategy that could be founded on the aggregated data and the resultant knowledge that emerges from the corresponding analysis.

Even though legislation and privacy norms are becoming increasingly strict, IT solutions for logging and auditing access to sensitive data, modelling, managing, and enforcing privacy and consent policies, as well as providing the ability to anonymize sensitive data are lacking. The goal is to make privacy and compliance part of the IT infrastructure and to ensure close coupling of all data with relevant consent and policies [22]. In the context of Internet of Things (IoT), Blockchain contributes significantly to constructing a truly decentralized, trustless, and secure environment. A Blockchain is a distributed ledger capable of maintaining an immutable log of transactions taking place in a network [23].

The validation of the transactions in terms of user security and ledger consistency is performed with asymmetric cryptography and distributed consensus algorithms. Asymmetric cryptography is utilized in the authentication process of the transactions. Each user owns a pair of private and public key. The private key is kept in confidentiality to sign the transactions. Next, the digital signed transactions are broadcasted across the network and after their validation are added to the Blockchain if consensus is reached. The typical digital signature algorithm used in Blockchains is the Elliptic Curve Digital Signature Algorithm (ECDSA). As far as the consensus is concerned, in Blockchain there is no central node that verifies the consistency of the ledgers in different nodes [24]. Therefore, different algorithms determine the consensus that the Blockchain should reach for a new block to be added to the Blockchain, while avoiding malicious attempts to attack the network.

According to the utilization of Blockchains in different application cases, there is a classification into multiple types that showcase distinct attributes [23]. Each of these Blockchain types are described below:

- Public Blockchains are truly decentralized, and its members can participate in publishing new blocks and accessing Blockchain contents. Public Blockchains are termed “permissionless” because they allow anyone to maintain a copy of the Blockchain and participate in the process of the validation of new blocks. This kind of Blockchain is designed to accommodate a high number of anonymous nodes, creating the need for protection against potential malicious behaviour. Publishing new blocks in the network involves either computationally expensive puzzle solving or staking one’s cryptocurrency. The incentive for peers to publish new blocks in the Blockchain has a processing fee attached to it. In this way it becomes too costly to tamper a block’s content. Most popular public Blockchain implementations are cryptocurrency networks, such as Bitcoin, Ethereum etc.
- Private Blockchains are permissioned and every node joining the network is a known member of a single organization. They are destined for single enterprise solutions and are utilized as a synchronized distributed database to keep track of data exchanges occurring between different entities. Furthermore, private Blockchains do not require currency or tokens to operate, while there are no processing fees included in their transactions, and the organization has the option to roll back their Blockchain status to any state from the previous ones.
- Consortium Blockchains, also named federated Blockchains, are like private Blockchains in the sense that they are permissioned networks. Consortium networks span across multiple organizations and assist in the transparency among involved parties. This kind of Blockchain is utilized as an auditable and reliable distributed database that is kept synchronized, that keeps track of data exchanges between the participating consortium members. They provide auditability and lower latency in transaction processing, while it is not entirely decentralized or censorship resistant.

As the summary of the information depicted above, a table with the distinctive features of each Blockchain type follows below [23].

	Public Blockchain	Private Blockchain	Consortium Blockchain
Participation in Consensus	All nodes	Single organization	Selected nodes in multiple organizations
Access	Public read/write	Can be restricted	Can be restricted
Identity	Pseudo-anonymous	Approved participants	Approved participants
Immutability	Yes	Partial	Partial
Transaction Processing Speed	Slow	Fast	Fast
Permissionless	Yes	No	No

Table 1 – Comparison of different Blockchain types

Within the Blockchain infrastructure, smart contracts are programmable applications stored in the Blockchain that manage transactions under specific terms and conditions. They can be viewed as the digital equivalent of traditional economic contracts between various engaging entities that are enforced not by centralized authorizing entities, since a Blockchain network does not require authorizing intermediaries to ensure that the conditions in a smart contract are fulfilled. Invoking functions in smart contracts entails an execution fee since an invocation is considered a transaction that is logged in the Blockchain. Execution fees play the role of mitigating flooding attacks and incentivizing peers publishing new blocks on the network. However, smart contract functionality varies within a Blockchain network. A number of instances of the smart contract role are enlisted below:

- Smart contracts enable multi-signature transactions that are carried out only when a majority or a required percentage of peers agrees upon.
- Automation of transactions that are triggered by specific events, such as fixed time intervals, completion of other transactions, receival of a message at the smart contract's address.
- Facilitation of utilities for other smart contracts. For instance, in Ethereum, inheritance can be written into smart contracts where one contract can invoke functions written in another contract.
- Manifestation of storage space for application-specific information such as membership records, lists or Boolean states etc.

Concluding on smart contracts, they are stored within the Blockchain and visible to all participants in the Blockchain. Security issues can occur if a peer exploits any bugs or loopholes in a deployed contract, therefore becoming critical to practice stringency in the design process.

Concerning the applications of Blockchain in the IoT sector, the major technological aspects that have been revolutionized concerning the port ecosystem are presented below.

First and foremost, the smart energy sector has attracted significant attention from the IoT community over the past recent years. Most applications that the Blockchain technology can be applied to are the preservation of privacy of the Users alongside their personal information, as well as the protection of the system from malicious transactions, such as Users attempting to trade unreasonable amount of energy [25]. Blockchain smart contracts in specific increase speed, scale, and security of exchanges of distributed energy

resources [26]. Furthermore, the transportation sector has been advancing significantly in terms of IoT technologies over the past few years. Blockchain is applied on smart vehicles for tamper-proof data exchange, monitoring of vehicle-related data, as well as, the design of intelligent transportation system architecture, which includes application, contract, incentive, consensus, data, physical, and network layers. In addition, Blockchain is also leveraged to implement general share data without third-party centralized management [25]. Lastly, supply chain systems have been enhanced to include cloud-based and on-demand manufacturing functionality. In this way, a safe distribution framework to share knowledge and services across enterprises is achieved through Blockchain technology [27].

A more extensive literature survey concerning the Blockchain technology is included in the D2.3 – Blockchain design specification deliverable [28].

2.2 COMMERCIAL PRODUCTS AND TECHNOLOGIES

2.2.1 General Data Integration Platforms

The insights unveiled by collecting data and combining information from multiple sources (data integration) have revealed that data-driven organizations are 23 times more likely to acquire new customers and 19 times more likely to be profitable than their competitors (according to McKinsey & Company[1]). The good news is that there are plenty data integration tools in the market. The bad news is that the "one-size-fits-all" approach does not work, and a closer look must be taken regarding the company's situation, IT infrastructure, types of data sources and regulations/compliances (e.g., government laws and standards).

Some of these Data Integration Tools are[2]: IBM InfoSphere[3], Microsoft SQL Server Integration Services[4], Informatica, Panoply[5], Oracle Data Integration platform[6] and Talend[7], to name a few.

The sourced data is stored into data warehouses and then it can be analysed via complicated SQL queries, AI tools and machine learning algorithms to produce valuable Business Intelligence.

Note that these platforms are only dealing with internal data which will not be the case when having multiple actors (e.g., multiple organizations, entities, ports). A new set of challenges will arise, different data sources, data governance, smart contracts and so on.

2.2.2 Ports Data Sharing Platforms

When narrowing down general Data Sharing Platforms (DSP) to the environment of a port we find a relatively small amount of work complete compared with other industries such as automobile or aviation. In this latest industry data sharing is allowing transparency, improving general efficiency, optimizing processes, and ultimately minimizing costs. However, the advantages have gone well beyond the limits of each individual airport and now every stakeholder can benefit from data sharing between airports. The similarities between airports and ports are indisputable: passengers, shipping goods, borders, aircrafts/vessels, authorities and so on. Therefore, an ever-increasing number of ports, institutions and software companies are investing on

[1] https://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/five-facts-how-customer-analytics-boosts-corporate-performance

[2] https://www.xplenty.com/blog/17-best-data-integration-platforms/

[3] https://www.ibm.com/analytics/information-server

[4] https://docs.microsoft.com/es-es/sql/integration-services/sql-server-integration-services?redirectedfrom=MSDN&view=sql-server-ver15

[5] https://panoply.io/

[6] https://www.oracle.com/middleware/technologies/data-integrator.html

[7] https://www.talend.com/

different approaches for data sharing platforms within the port ecosystem. As shown in Figure 2, these DSP solutions can interconnect any actor within the port to collaborate and share data. These functionalities can be extended beyond the port to communicate with upstream and downstream ports.

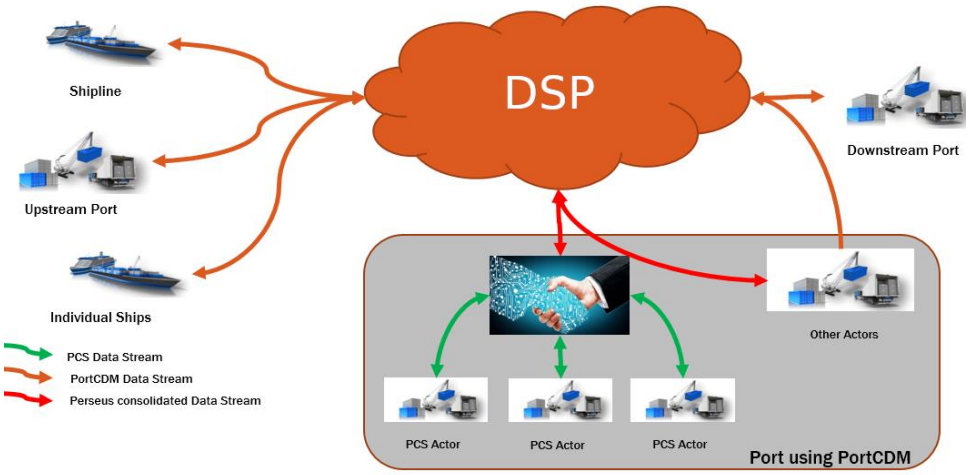


Figure 2 – General DSP workflow in port environments (source: Marine Fields)

Between the various platforms already in use there are, however, some common capabilities although they are implemented differently without common rules:

- Disparate supply chain data is synthesized into a standardised single source of truth.
- Easy integration into existing IT systems through API-driven architecture.
- Modern user interfaces for displaying processed data.
- Data protection through access control and cryptography.
- Data governance that ensures high quality through the complete life cycle of the data.

Next, we will be presenting some of these platforms, their features, and its assessment with the aid of the "four step digitization of ports" from the Port of Rotterdam Authority from lowest to highest efficiency (Figure 3):

- 0 - No automation. No digitization.
- 1- Single automation: where each individual organization focuses on automating their own processes internally.
- 2- Integrated system in the port: where all organizations involved in the port activities actively share information in a single common environment.
- 3- Integrated system from port to hinterland: where this common environment extends to actors in the logistic chain to and from hinterland such as ship/rail/road operators and inland terminals and ports.
- 4- Connected ports and supply chains: the future scenario where ports and supply chain members will share information worldwide taking efficiency to the highest level.

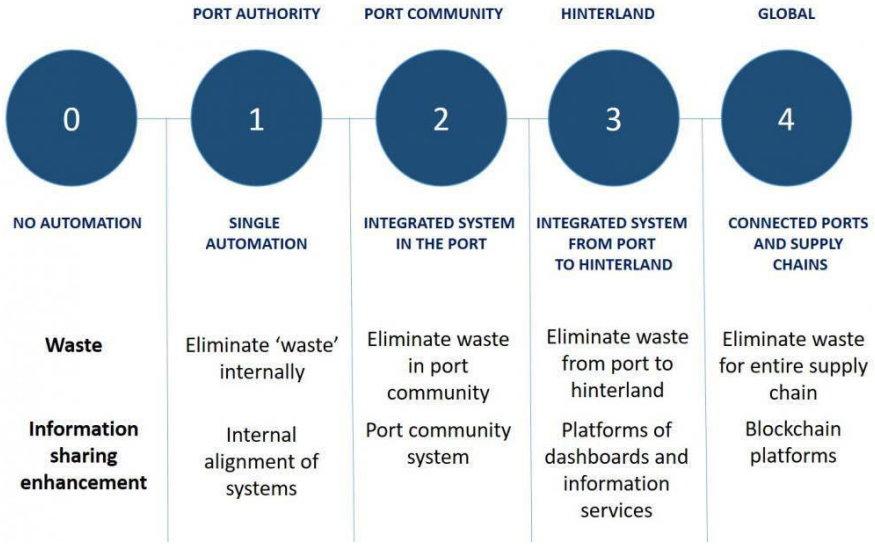


Figure 3 – Digitisation of ports in four steps (source: Port of Rotterdam Authority)

The following solutions are located primarily in the second step "Port Community" and lack intercommunication outside the port although their implementations include a common ontology along with authorization and security capabilities:

Port Optimizer (GE Transportation)[8]. It is a cloud-based software solution implemented in the Port of Los Angeles gathering data from Shipping Lines, Marine Terminals and third parties. Scalable to support over 10 million TEU facilities, built-in Big Data functionalities, rapid production deployment, no disruption upgrades and minimal setup.

NextPort (Port of Antwerp)[9]. It is a data utility platform collecting data from various stages in the supply chain of the Port of Antwerp and delivering consistent data by means of an extensive pool with over 15 APIs to choose from. NextPort is an open project welcoming new partners from all over the world.

Posidonia (Prodevelop)[10]. It is a suite of proactive solutions for the management and optimization of port activities also with an extensive number of modules comprising PMS, PCS, CDM, vessel operations management and business data intelligence among others. It is implemented in all major Spanish Ports to a greater or lesser extent.

Perseus (Marine Fields)[11]. It is a system to interconnect any maritime actor allowing data sharing and collaboration, independent of the systems used by the individual participant. Also, it allows Users to share BDI and connectivity of external actors with a limited individual adaption.

PortForward (Port of Rotterdam)[12]: A set of software tools for the optimization of port operations via collaboration and data sharing between all agents in the port. It also displays new insights and control of the logistic chain for shippers and freight forwarders. PortForward includes real-time information for a PCS platform (Portinsider), port call platform (PortXchange), worldwide route planner for containers via Rotterdam (Navigate), container cargo tracking (Boxinsider), rail network monitoring (OnTrack), vessels arrivals predictions (Timetoport), optimum port assets management (Port Asset Tooling) and a flexible port

[8] https://www.wabtec.com/business-units/ge-transportation-a-wabtec-company/port-optimizer

[9] https://www.nxtport.com

[10] https://www.prodevelop.es/en/ports/posidonia

[11] https://www.marine-fields.com/product

[12] https://www.portofrotterdam.com/en/port-forward/products

management system (Portmaster).

ValenciaportPCS[13]: is an open data sharing platform for port and logistic processes allowing secure information exchange. Over twenty transactional and informative services are available to more than 600 public and private agents in the ports of Valencia, Sagunto and Gandia. It works with Inttra and GT Nexus merging major ocean carriers to offer Booking, Shipping Instructions and Track & Trace services.

SmartPort Logistics (Hamburg Port Authority)[14]: is a platform created by the Port Authority to optimise the transportation of goods as an entire system by means of making its own data publicly available and allowing all stakeholders to integrate their data into the platform. Due to the distinctive constraints posed by nearby urban areas, the platform has a primary focus on infrastructures assets and flows such as roads, bridges, rail network, parking spaces, traffic flows and GPS location.

Other platforms already provide a wider scope outside the Port Community and Hinterland reaching the fourth step:

Tradelens[15]: is a Blockchain powered, cloud-based platform by Maersk and IBM providing control and management of shipping data to every stakeholder in the supply chain through a refined permissions-based system. The platform can be accessed via open APIs that enables all the participants to integrate their systems adopting a data model and access control scheme based on UN/CEFACT. It is currently processing two million shipping events and 25,000 trade documents a day with over 170 active stakeholders worldwide.

TRAXENS-HUB[15]**:** is a cloud-based platform by TRAXENS providing smart container visibility data mainly to the shipping lines enabling them to better manage their fleet. It also offers to every stakeholder in the supply chain that is interested in real-time actionable insight the ability to track shipments and make decisions that increase supply chain efficiency, namely the freight forwarders and the beneficial cargo owners. The platform can be accessed via open APIs that enable all the participants to integrate their systems adopting a data model and access control scheme based on their needs. It is currently processing more than 100,000 equipped containers, over than 60,000 booking trips per year, over then 40 types of data, and four hundred million of smart container events and with many active stakeholders worldwide including the two major shipping lines MSC and CMA CGM.

CargoSmart[16]: another Blockchain powered, cloud-based platform where Shippers, Consignees, Logistic Service Providers and Ocean Carriers have access to secure, global shipment data allowing them to better manage their shipments improving visibility, execution, collaboration, and compliance. It is available on application, integration and mobile platforms and even performs machine learning to further boost planning and management. Over 40 ocean carriers are connected to CargoSmart monitoring 12,000 vessels' live movements and 800 container ports worldwide.

Inttra[17]: It is a set of solutions shipper-oriented offering a wide variety of services. Ocean Trade Platform offers them access to the carrier industry allowing for booking, digital shipping orders, real-time container status events and alerts. Ocean Schedules offers shippers all information regarding the carrier itinerary (transhipment details, cut-off date, ETA, ETD, vessel type, voyage number). Inttra also offers electronic submission of Verified Gross Mass (VGM) and a dashboard with reports and analytics. It is cloud-based with web interface or integration via web API.

[13] https://www.valenciaportpcs.com/en/

[14] https://www.hamburg-port-authority.de/fileadmin/user_upload/150618_spl_app_web.pdf

[15] https://www.tradelens.com/

[16] https://www.cargosmart.com/en/default.htm

[17] https://www.inttra.com/shipper_solutions/

Calista[18]: a Blockchain-based global supply chain platform that offers cargo owners and logistic Service Providers wide data-related services through a single electronic gateway within a network of business and governments. It is cloud-based with a web interface and alerts on some port events.

PI System (Osisoft)[19]: It is a suite of software products running on Windows OS that supplies a continuous flow of relevant data by collecting, analysing, and displaying it. Data collection is carried out by the third-party or proprietary PI Interfaces from a diversity of sources regardless the system, language, format or device, unifying, storing, and centralizing all incoming data in the PI Servers. Then, real-time and historical information can be searched and analysed either ad hoc or automatically by third-party systems (e.g., ERP). Finally, operational and business information can be accessed by many different PI Clients, including a PI Web API, to display visualizations to the end user on any device.

The above platforms' common technologies and features are shown below and have been extracted from the available information on their websites although in many cases references to the technologies have been found without further details. We can clearly see that DataPorts is the only platform IDS compliant meaning trustful, transparent, and secure exchange of data between participants. Also, data owners can stay in control of their data at any time.

Platform	IDS compliant	AI and Analytics	Blockchain	Cloud Based
Port Optimizer	No	Yes	No	Yes
NextPort	No	No	No	No
Posidonia	No	No	No	No
Perseus	No	Yes	Yes	No
ValenciaportPCS	No	No	No	No
SmartPort Logistics	No	Yes	No	Yes
Tradelens	No	Yes	Yes	Yes
TRAXENS-HUB	No	Yes	No	Yes
Cargo Smart	No	Yes	Yes	Yes
Inttra	No	Yes	No	Yes
Calista	No	No	Yes	Yes
PI System	No	Yes	No	Yes
Port Forward	No	Yes	No	Yes
DataPorts	Yes	Yes	Yes	No (TBD)

Table 2 – DSP platforms' common technologies and features

[18] https://calista.globaletrade.services/

[19] https://www.osisoft.com/pi-system/

Besides all the data platforms mentioned above that allow companies in the global supply chain to enhance its operations it is worth mentioning another complementary data provider: **Traxens**[20]. It provides high-value, real-time, door-to-door, IoT insight data along the supply chain life cycle. Based on its cutting-edge smart container solution, Traxens generates, collects, qualifies, and certifies data such as geolocation, geofencing, shock detection, door opening, temperature and humidity.

2.2.3 Technologies

2.2.3.1 Message Brokers

Message brokers allow senders (data producers) and receivers (data consumer) to communicate large volumes of data asynchronously through a publish-subscribe paradigm, meaning they are loosely coupled and do not need to be online at the same time. Messages are queued in and out and this becomes useful for unreliable networks and Big Data applications. Three of these differences are: backlog behaviour, clustering, and protection of data. Some examples are:

RabbitMQ: It is focused on delivery consistency and suited for complex routing to consumers. The approach is "smart broker/dumb consumer" meaning the broker ensures that messages are delivered and dequeued once it gets acknowledgement from all consumers having read a particular message. Speed up to 50 thousand messages per second. It also relies on many plugins for different scenarios and has an API for monitoring and management.

MQTT Route: Is a complete IoT Application suite with an inbuilt MQTT Broker secured with TLS/SSL. Features device Authentication, schedulable built-in Rule Engine, error notification support, customizable User Interface, API to manage Edge Devices and speeds up to 30 thousand messages per second. It can be deployed on the cloud or on premise.

GRPC: It is a high-performance, open-source universal remote procedure call (RPC) framework by Google. It is based on Protocol Buffers which are language and platform neutral for serializing structured data. It uses HTTP/2 to highly support performant and scalable APIs resulting in reduced response times, network usage and battery consumption in mobile devices. This two-way streaming allows for server-push notifications. It is also more secure as it can use TLS/SSL.

Apache Kafka: It is focused on high volume messages (up to 1 million/sec) and streams, speed, and scalability. It also keeps the history of messages for a set amount of time. The approach is "dumb broker/smart consumer" which leaves the consumer responsible for tracking messages in each log (consumer state) relieving Kafka of such operations. If required, external libraries for transaction support may be implemented for delivery consistency. The main drawbacks for Kafka are the need for external services to run (e.g., Zookeeper which is an Apache service for distributed system) and external monitoring tools as Kafka uses CLI management.

We also have brokers within the architectures of FIWARE and IDSA that will be examined on 2.2.3.4 and 2.2.3.5.

MBS	Vendor	URL	License
RabbitMQ	Pivotal Software	https://www.rabbitmq.com	Mozilla Public License
MQTT Route	Bevywise	https://www.bevywise.com	Commercial
gRPC	gRPC	https://grpc.io	Open Source (CC-BY-4.0)

[20] https://www.traxens.com/

MBS	Vendor	URL	License
Apache Kafka	Apache Software foundation	https://kafka.apache.org	Apache License 2.0

Table 3 – Most popular Message Broker Systems in 2020 (www.slant.co)

2.2.3.2 Distributed Coordination System

To build a system with high availability, a distribution across nodes needs to be implemented which leads to the use of a mediator acting as a hub for communication. This will avoid internal network failures while communicating and gathering information, but at the same time this hub becomes a point of single failure. It can become highly available, however, by replicating it in so-called Distributed Coordination Systems enabling nodes to communicate in runtime sharing configuration, variables, locks, and distributed data stores[21]. Some examples are:

Consul: is a service providing configuration, segmentation, and service functionality. Each of these features can be used individually or they can all be used together to build a full-service mesh. It requires a data plane and supports both a proxy and native integration model. Consul ships with a simple built-in proxy for quick start and supports third party proxy integrations for further configuration.

etcd: is an open-source, strongly consistent, distributed key-value store providing a reliable way to store data that needs to be retrieved by distributed systems or cluster of machines. It can tolerate machine failure even in the leader node and applications of any complexity/size can read and write from/into etcd.

Zookeeper: Centralized, open-source service for maintaining configuration information, naming, providing distributed synchronization and providing group services aiming to reduce the management complexity when deploying a distributed application.

Kubernetes: is a portable, extensible, open-source platform for managing containerized workloads and services, that facilitates both declarative configuration and automation. It has a large, rapidly growing ecosystem. Kubernetes services, support, and tools are widely available.

DCS	Vendor	URL	License
Consul	HashiCorp Inc.	https://www.consul.io	Mozilla Public License
Etcd	CoreOS	https://coreos.com/etcd	Open Source
Zookeeper	Apache Software foundation	https://zookeeper.apache.org	Open Source
Kubernetes	Google	https://kubernetes.io	Kubernetes

Table 4 – Most popular Distributed Coordination System in 2020 (www.slant.co)

2.2.3.3 Storage

From all the diverse categories in existence for database/storage the most remarkable distinction to consider when selecting one is the SQL/NoSQL model[22].

SQL or Relational databases are a mature technology from the early 80's, widespread and well stablished with plentiful choices in the market. They are mostly general-purpose but have been expanded over the years

[21] https://medium.com/@Imesha94/distributed-coordination-5eb8eabb2ff

[22] https://www.dataversity.net/review-pros-cons-different-databases-relational-versus-non-relational/

with non-relational concepts such as not atomic attributes, user-defined data types, inheritance and hierarchies allowing adaptability for specific usage scenarios.

Relational databases store data as rows and columns in tables. Each row in a table has its own unique key. Rows in a table can be linked to rows in other tables by adding a column for the unique key in the linked row. Such columns are known as foreign keys. When using relational databases, we want to be able to select or modify one and only one row in a table. Relational databases allow this by implementing a unique primary key for each row in a table. When a new row is written to the table, a new unique value for the primary key is generated. This key is used by the system primarily for accessing the table. These primary keys are also used within a database to define the relationships among tables[23].

Relational databases have a vast amount of documentation, resources and use support available. They are considered the "truly databases" for their maturity and consistency since they are ACID compliant. This means that even in case of transaction or failure the state stays consistent, but this consistency comes at a higher complexity and performance although it is possible to trade off ACID rules for performance improvement in some implementations.

The main downside of SQL databases is the ability to work well only with well-structured data, while unstructured or semi-structured poses big challenges due to schema and type constraints. These constraints are also related to a problem with mapping to object and classes representing the data. As the mapping often is not one-to-one, this presents difficulties in data interpretation as well as implementation of persistence mechanisms for classes defined in programming languages. Because of all these constraints, complex datasets are difficult to handle.

Another downside is a complex horizontal scalability. This means adding more machines to a setup of resources which usually consists of partitioning the data so that each node only contains a part of the whole data. This can be done without downtime or major performance impacts, while vertical scaling always comes with downtime as the hardware (CPU and RAM) must be replaced. Also, it is limited by the expandability of the machine that is to be upgraded (max supported RAM, max CPU cores/slots). Some examples are:

Oracle: It is a multi-model database supporting relational, columnar, XML, JSON, Spatial, Graph and unstructured data. It is the world's leading database for online transaction processing, data warehousing and mixed database workloads.

MySQL: It is an open-source database engine available for all major platforms. It is easy to use, secure, reliable, scalable and one of the fastest database languages.

MS SQL Server: It is a secure, scalable RDBMS that supports business intelligence and analytics applications in corporate IT environments. It delivers advanced analytics by using R and Python directly within the RDBMS.

DBMS	Vendor	URL	License
Oracle	Oracle	http://www.oracle.com/database	Commercial and restricted free version
MySQL	Oracle	http://www.mysql.com/	GPL - open source and commercial
Microsoft SQL Server	Microsoft	http://www.microsoft.com/en-us/sql-server	Commercial and restricted free version

Table 5 – Most popular RDBMS in 2020 (DB-Engines Ranking)

[23] https://www.ibm.com/cloud/learn/relational-databases

Non-Relational databases are schema agnostic thus allowing for efficient unstructured/semi-structured data storage. They are generally more horizontally scalable and fraud-tolerant than Relational databases and can be easily distributed across different nodes.

The main downsides in relation to Relational databases are weaker consistency (instead of ACID), denormalized data and narrower spread among the community.

The "NoSQL" database model comprises a set of various approaches that have been designed to cover different application domains reaching distinctive levels of maturity for each approach[24]. These implementations are key-value stores, wide column stores, document stores, graph databases, search engines and time series databases.

Key-value stores work by storing data in dictionaries or hash tables that contain a collection of objects/records which are stored and retrieved using a key that uniquely identifies one record. These collections require no predefined structure and may have different fields for each record. The key-value implementation is good for non-complex data and speed. Some examples are:

Redis: It is an open-source, non-relational, in-memory data structure store, cache and message broker featuring transactions, publish/subscribe, Lua scripting, keys with a limited time-to-live, LRU eviction of keys and automatic failover. It is mostly tested in Linux and OS.

Memcached: It is a NoSQL, open-source, high-performance, distributed memory caching system intended to speed up dynamic web applications by reducing the database load. It is a key-value dictionary stored in the memory, resulting from database calls, API calls or page rendering.

Etcd: has already been introduced in 2.2.3.2.

DBMS	Vendor	URL	License
Redis	Salvatore Sanfilippo	http://www.redis.io/	Open Source and Commercial
Memcached	Danga Interactive	http://www.memcached.org/	BSD License
etcd	etcd	http://www.etcd.io/	Open Source

Table 6 – Most popular Key-Value stores in 2020 (DB-Engines Ranking)

Wide column stores keep data in records with an ability to hold very large numbers of dynamic columns. These column names and record keys are not fixed and since a record has no limitation in the number of columns, wide column stores can be seen as two-dimensional key-value stores. They are well suited for scalability in massive distributed systems. Some implementations use CQL, which is a variant of SQL, for data definition and manipulation. Some examples are:

Cassandra: It is an open-source, non-relational DBMS designed to handle large amounts of data and offer continuous availability, low latency, linear scale performance and easy data distribution across multiple data centres and cloud services.

HBase: It is an open-source, NoSQL, distributed, horizontally scalable, persistent and consistent DBMS built on top of the Hadoop file system ideal for real-time analysis due to its low latency reads, writes and updates.

MS Azure Cosmo DB: Is a horizontally scalable, non-relational, globally distributed, highly available, fully managed, low latency, multi-model (including key-value, document, graph and column family-based), multi-

[24] https://www.thoughtworks.com/es/insights/blog/nosql-databases-overview

API database for managing large amounts of data running in Azure.

DBMS	Vendor	URL	License
Cassandra	Apache Software Foundation	https://cassandra.apache.org/	Open Source
HBase	Apache Software Foundation	http://hbase.apache.org/	Open Source
Microsoft Azure Cosmo DB	Microsoft	http://www.azure.microsoft.com/services/cosmos-db	Commercial

Table 7 – Most popular Wide Column stores in 2020 (DB-Engines Ranking)

Document stores are schema-free where data is stored as JSON documents. They are similar to key-value stores in the way that they use document names as keys and contents as values. These databases are well suited for managing semi-structured data in distributed systems. Some examples are:

MongoDB: It is the most popular NoSQL, open-source, large volume, distributed, scalable, JSON-like, high throughput and efficiently indexed DBMS allowing for complex ad hoc queries.

Amazon DynamoDB: It is a NoSQL, serverless database provided by Amazon Web Services. It is multi-model (key-value and document based), scalable, highly available and distributed.

MS Azure Cosmo DB: It has just been introduced in the previous section.

DBMS	Vendor	URL	License
Mongo DB	Mongo DB, Inc	http://www.mongodb.com/	Open Source
Amazon Dynamo DB	Amazon	http://www.aws.amazon.com/dynamodb	Commercial
Microsoft Azure Cosmo DB	Microsoft	http://www.azure.microsoft.com/services/cosmos-db	Commercial

Table 8 – Most popular Document stores in 2020 (DB-Engines Ranking)

Graph databases represent data as a network of nodes/objects. These databases are used mostly when an in-depth analysis of relationships between data points is required. Some examples are:

Neo4J: It is an open-source, highly scalable and available, real-time based, and schema-free Graph database developed in Java. It provides a powerful declarative query language (Cypher), supports full ACID rules and indexes by using Lucene.

ArangoDB: ArangoDB is a scalable open-source multi-model (graph, document and search) database. All supported data models & access patterns can be combined in queries allowing for maximal flexibility. ArangoDB runs either on premise or in the cloud.

OrinteDB: It is an open-source Multi-Model, NoSQL database technology designed to work with the Graph, Document, Key-Value, Geospatial and Reactive models while managing queries with SQL syntax.

DBMS	Vendor	URL	License
Neo4J	Neo4J, Inc	http://www.neo4j.com/	Open source

ArangoDB	ArangoDB, Inc	http://www.arangodb.com/	Open Source
OrientDB	OrientDB LTB	https://orientdb.org/	Open Source

Table 9 – Most popular Graph databases in 2020 (DB-Engines Ranking)

Search engines are NoSQL database management systems with general focus on making data easily available via text-based search. Other available features are distributed search, support for complex search expressions, stemming, ranking and grouping results and geospatial search. They are suitable for high scalability. Some examples are:

Elasticsearch (see also 2.2.3.6): It is a popular, real-time, scalable, distributed search and analytic engine with high availability built on top of Apache Lucene used for full-text search, structured search, analytics or all three in combination. It is a schema-free, document-oriented data store exposing RESTful API that allows you to integrate, manage and query index data in a variety of ways.

Splunk: It is a software which processes and brings out insight from machine data and other forms of Big Data. This machine data is generated by IoT devices, running webserver, logs from mobile apps, etc. It can read this unstructured, semi-structured or non-structured data and then allow to search, tag, create reports and dashboards on it.

Solr: It is a popular, open-source enterprise search platform written in Java from Apache Lucene project. It features full text search, faceted search, highlighting, near-real-time indexing, dynamic clustering, rich document handling and geospatial search, highly scalability and reliability.

DBMS	Vendor	URL	License
Elasticsearch	Elastic	https://www.elastic.co/elasticsearch/	Open Source
Splunk	Splunk Inc.	http://www.splunk.com/	Commercial
Solr	Apache Software Foundation	https://lucene.apache.org/solr/	Open Source

Table 10 – Most popular Search Engines in 2020 (DB-Engines Ranking)

Time series database management systems are specifically designed to collect, store and query time series with high volumes. They allow for creation, enumeration, update, deletion, and organization of various time series as well as basic calculations. This handling of time series can be done with other categories of DBMS but at a lower performance. Some examples are:

InfluxDB: It is an open-source, fast, high-availability and retrieval time series database written in Go. Data management, injection and integration are simple and has excellent documentation. Access to the database is carried out with a SQL-like language.

Kdb+: It is a high-performance relational database platform with time series built in supporting real-time analysis for billions of records. Q is the language for working with kdb+. Both kdb+ and “q” are written in K programming language.

Prometheus: Written in Go it is an open-source, metrics-based monitoring system through a simple yet powerful data model and query language (PromQL). Labels and names for each time series allow for aggregation alerts management. Performance is up to millions of records per second on a single server.

DBMS	Vendor	URL	License

InfluxDB	InfluxData	http://www.influxdata.com/products/influxdb-overview	Open Source
Kdb+	Kx Systems	https://code.kx.com/q/	Commercial
Prometheus	Cloud Native Computing Foundation	http://www.prometheus.io/	Open Source

Table 11 – Most popular Time Series databases in 2020 (DB-Engines Ranking)

2.2.3.4 FIWARE

FIWARE is an open source, royalty-free framework aiming to accelerate the development of smart solutions through the definition of a universal set of standards for Context Information (FIWARE Core or Context Broker) and the provision of a set of general-purpose platform functions (Generic Enablers or GE) built around it.

The basic FIWARE architecture normally consists of five blocks where there can be components from different layers which can connect to each other[25]. The Context Broker Generic Enabler is the core and only mandatory component. It receives information in NGSI v2 format, holds the current context/state of the overall system and can pass information onto other layers. The other four components can be FIWARE components or 3rd-party platform components to design a custom hybrid platform.

The Interface to IoT, Robots and external systems offers various GE to gather context information or trigger actuations in response to context updates. The Context Processing, Analysis and Visualization component allows several GE to implement the "smart behaviour" by means of processing, analysing and visualizing context information. Under the Context Data/API Management, Publication and Monetization Component we can find a series of GE dealing with secured access (OAuth2, XACML), publication and monetization of context data resources. Finally, the Deployment Tools available in the FIWARE platform aim to ease the deployment and configuration of components either within or outside the platform.

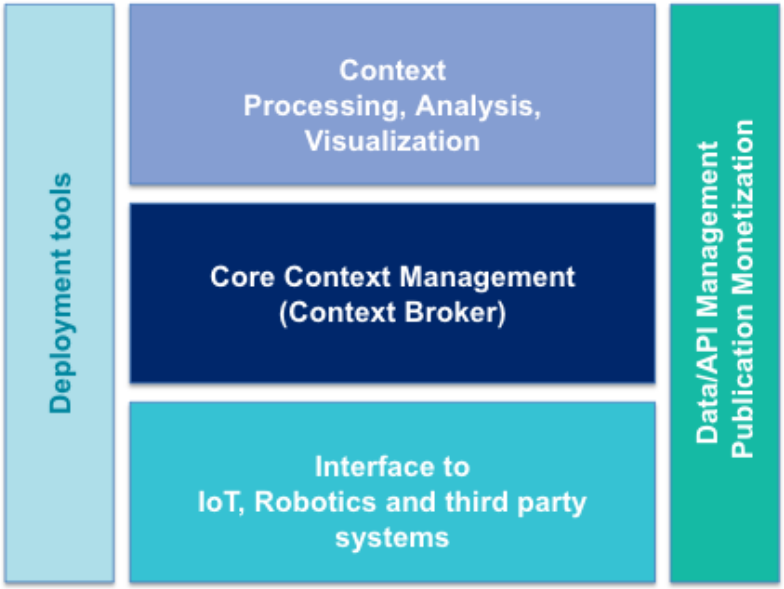


Figure 4 – Basic FIWARE Architecture

[25] https://www.FIWARE.org/developers/catalogue/

A small/medium enterprise that can make their application "FIWARE ready", i.e., able to manage context information through the Context Broker, means that it can be plugged into any system. Similarly, if there is a processing layer which communicates through NGSI, it can pass information down to devices which are "FIWARE ready" regardless of the actual platform resulting in a "vendor-locking" free system.

Some important FIWARE Generic Enablers are[26]:

- Context Broker: enables to perform updates, queries or subscribe to changes on context information through the NGSI v2 API. There are new brokers being incubated supporting NGSI-LD API: Orion-LD, Scorpio and Stellio.
- Cygnus, QuantumLeap, Draco and STH Comet: Used for managing the persistence of historic context into databases.
- IDAS: capable to deal with the current wide variety of IoT protocols through different Agents and a customizable Agent to deal with potential sources of context data.
- OpenMTC: brings an open-source implementation of the OneMTM standard.
- FIROS, Fast DDS and Micro XRCE-DDS: middlewares for Robot-based Systems.
- Domibus: allows for Users to exchange electronic data and documents.
- Kurento and OpenVidu: capable of getting data out of stream information like video.
- WireCloud: capable to read data from the NGSI format and display it in whichever graphically method is chosen.
- Knowage: is a Business Intelligence suite offering a wide range of analytical tools.
- Perseo: advance event processing enabling the user to fire events which send HTTP requests, emails, SMS etc.
- FogFlow: orchestrates dynamic processing flows over cloud and edges based on various context.
- Cosmos: intended to bring tools enabling Big Data analysis of both batch and stream context information.
- Ckan: is a data management platform that makes data accessible by enabling the publication, search, and consumption of datasets.
- Biz Framework: brings backend support to Context API/Data monetization based upon TM Forum APIs.
- Wilma, Keyrock and AuthZForce: secure access control to FIWARE components.

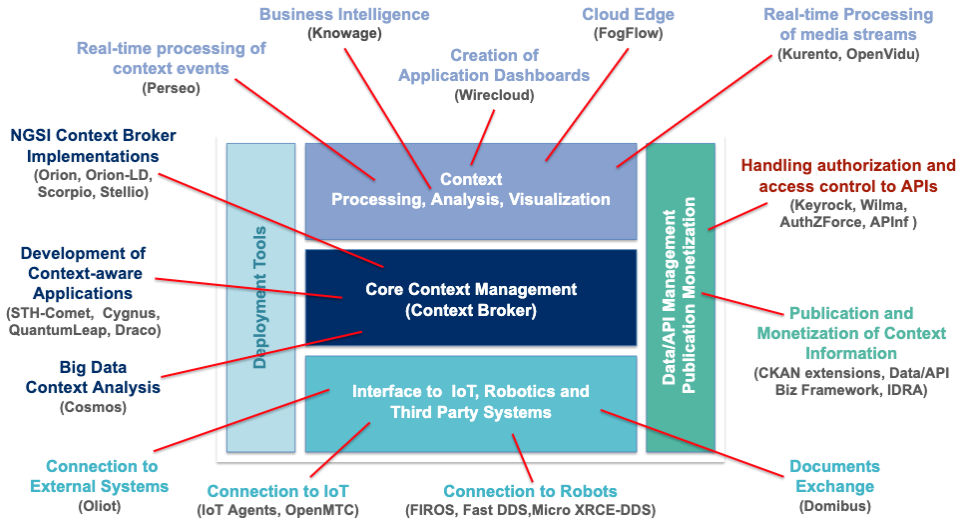


Figure 5 – FIWARE Generic Enablers

[26] https://www.FIWARE.org/developers/catalogue/

2.2.3.5 International Data Spaces Association

IDSA aims to bring a standardized, trustworthy, secure, sovereign approach to both data trading and sharing between participants. It connects the lower-level architectures for communication and basic data services with more abstract architectures for smart data services establishing a secure supply chain from data sources to Data Users.

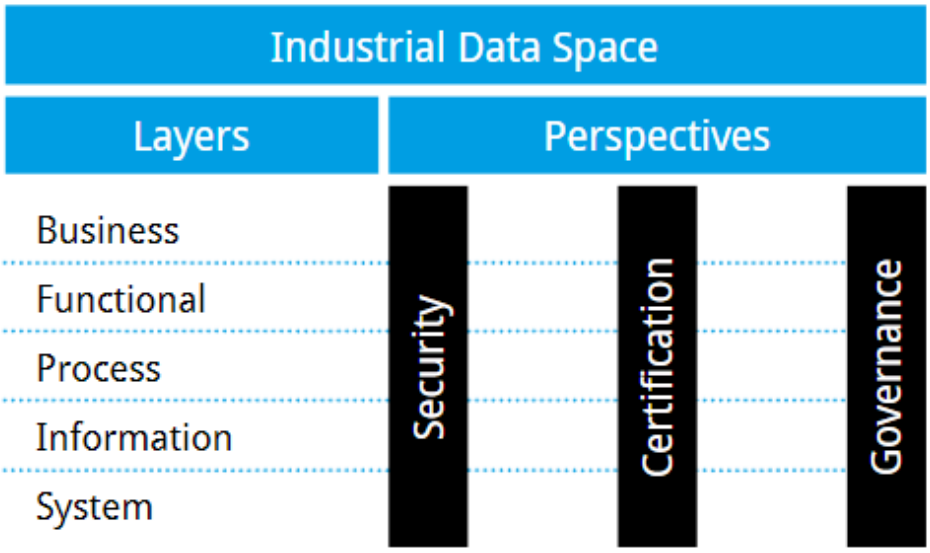


Figure 6 – IDSA – General Structure of Reference Model

Its Reference Architecture Model[27] is divided in five layers:

- **Business layer**. Specifies and categorizes the different roles which the participants of IDS can play and specifies the main activities and interactions connected with each of these roles.
- **Functional layer**. Defines the functional requirements of the IDS such as trust, security, sovereignty, interoperability, description, processing, and financial value.
- **Process layer**. Specifies the interaction among the different components of the IDS, the main interactions being the provision, exchange, and publishing of data.
- **Information layer**. Defines a conceptual model which is the central agreement between all participants and components facilitating interoperability and compatibility.
- **System layer**. Determines a particular data and service architecture from the roles in the Business layer to meet requirements in the Functional layer resulting in three major technical components: Connector, Broker and App Store. The DataPorts Platform will adapt to the component's interaction schema (Figure 7).

To enable data exchange between participants while retaining control over it. This will be achieved through software components (connectors) that link a device or cloud environment to a data space. Each connector can define its own rules that must be met by other connectors in order to access its data. The architecture of IDS consists of a distributed network of connectors, data brokers and clearing houses interacting without a centralized data storage.

The connectors are the gatekeepers between data and internal infrastructure and the data space. Each actor willing to consume or provide data in this data economy will need to build its own. The broker only exchanges (machine readable) metadata of the data source and data usage policies. The actual data flows directly between connectors (not through a broker) once the identity and rules from the provided connector have

[27] https://www.internationaldataspaces.org/wp-content/uploads/2019/03/IDS-Reference-Architecture-Model-3.0.pdf

been met. Further, the connectors may load and execute data apps, which encapsulate data processing or transformation services.

Thus, a participant in a data space can technically control to what extent, for how long and with what accuracy another connector can access its data.

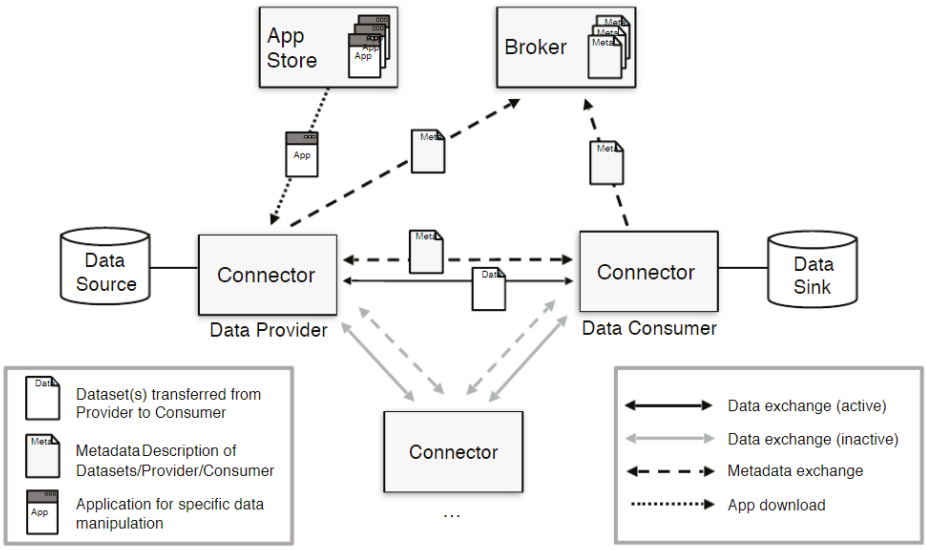


Figure 7 – IDSA – Interaction of Technical Components

2.2.3.6 Elastic Stack

The Elastic Stack[28] is designed to provide a solution for centralized log management and analytics and ensures that systems generating an ever-growing amount of data, whether event logs or metrics, can make fast, reliable and meaningful use of it at all times.

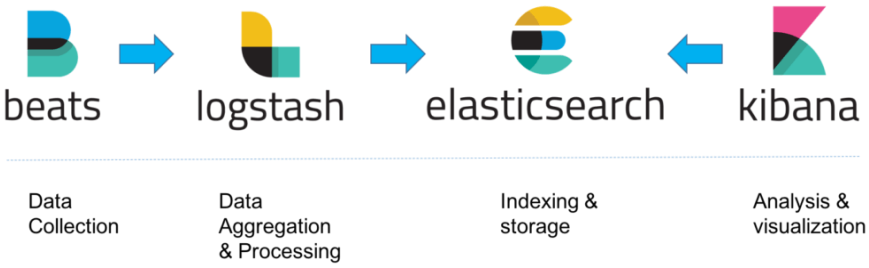


Figure 8 – Elastic Stack Flow

As shown in the above figure, the Elastic Stack consists of 4 independent software products each one having a particular role (Beats, Logstash, Elasticsearch and Kibana):

Beats are a collection of open-source, lightweight, single-purpose data shippers written in Go that act as agents installed on the different servers in your environment collecting either metrics or logs. They can ship data directly into Elasticsearch and some even support processing although the normal flow will ship logs into **Logstash** for collection, processing, and dispatching. These three stages are defined with plugins:

- Input plugin: responsible for aggregating logs and events using more than 50 Input plugins for

[28] https://logz.io/learn/complete-guide-elk-stack/#intro

different platforms, databases, and applications.
- Filter plugin: enables for logs to be enriched, manipulated, and processed making sure the log messages are parsed and structured correctly for efficient use down the pipeline.
- Output plugin: allows to push your data to various locations, services, and technologies

The number of combinations of inputs and outputs along with the use of codecs makes Logstash a versatile event transformer that outputs the processed logs into the central component in the Elastic Stack: **Elasticsearch** (see also 2.2.3.3). It plays the role of data indexing and storage with a strong focus on search and analysis. It is categorized as a NoSQL database although it is possible to use SQL statements to interact with the data from the latest versions. It is horizontally scalable. The structure of the data has a particular terminology where "indexes" are the equivalent to tables in Relational databases and "documents" relate to table rows. It also exposes an extensive REST API to integrate, manage and query the indexed data in countless different ways. These functionalities can be extended even further by adding different plugins such as alerts, analysis, security, restoration and many more.

The last piece in the Elastic Stack is **Kibana**, an open-source, browser-based user interface that can interact with the data stored in Elasticsearch indices: searches, filtering, autocompletion, visualizations, dashboard, analysis, and monitoring.

2.2.3.7 Linked Data (Epimorphics)

The concept of Linked Data[29] transfers the principles of the Web into data sharing: each data "item" will have its individual identity or URI the same way web resources are identified by URLs. These data URIs will be linked to others thus creating a web of linked information the same way the Web links documents and other resources to each other.

This approach provides several benefits, the data is:

- placed in context: each item has a web address through which it can be referenced.
- linked enabling information to be combined across silos.
- accessible over the web so that downstream applications can run from the live data.

The linked data approach builds upon several standards:

- URIs are used to identify anything of interest in the data. However, the recommendation is to use HTTP URLs so that standard web client software can fetch from them.
- The Resource Description Framework (RDF) is the recommended approach to representing the data. It represents data as statements so called "triple" of resources with a subject, a property, and a value. These statements are schema free so different authorities can publish different statements about the same subjects using their own set of properties although vocabulary standards are still required.
- The vocabulary defining what terms are available may range from a simple list of terms to a complete ontology depending on the needs of a particular application.
- The last step is to query the data. The URI may provide certain information but in a normal situation there is the need of a query language suited to this graph-like web of data called SPARQL that can access data aggregation.

Epimorphics offers a Reference Data Manager and a Data Sharing platform. The former helps in the management of the standardized terms by providing tools and services such as controlled, authoritative lists of identifiers as URIs thus supporting data governance, data collaboration and data use. The latter provides a cost-effective, flexible approach to publishing data on the web and exploiting the benefits of linked data. A resilient, scalable, and cloud-based solution for publishing data.

[29] https://www.epimorphics.com/

2.3 RELATED EU PROJECTS

One of the main aspects in the context of the State-of-the-Art analysis is identifying and evaluating related EU projects. This considerably increases the potential collaborations between various EU projects, improving the results delivered by each project.

We performed a survey to identify the related EU projects. Within this survey in the first step, iteratively all project partners were asked to insert any relevant project to an excel sheet. The usefulness and the relevance of the projects were discussed in several online meetings in the context of WP2 online calls.

The result was an excel sheet with a list of projects and several aspects defined for each project. The excel sheet can be found in the dedicated project repository[30].

Afterwards these projects were evaluated one by one by obtaining more details about them using the dedicated websites or available deliverables.

The result includes nine projects which are listed in the following sections. For each project, a set of aspects are specified and at the end in the summary sections, a short vision on collaboration possibilities is provided.

2.3.1 PortForward

Towards a green and sustainable ecosystem for the EU Port of the Future.

Programme

H2020-EU.3.4. - SOCIETAL CHALLENGES - Smart, Green And Integrated Transport

Status

Started at 01.07.2018 (42 Months)

Link

https://www.portforward-project.eu/

Objective

The Port of the Future will be able to enhance sustainable development and to manage the resources to be invested and their employment for a competitive advantage. Therefore, the port of the future must be oriented to port community and have an operative strategic capability to work, in line with European purposes, on the following:

- Smart, through ICT solutions, because it is important to improve exchange of information flows between the port and the port community.
- Interconnected with the use of a combination of different modes of transport and the integration of different technologies, because it is important to achieve better monitoring and controlling of the freight flows.
- Green through the adoption of green technologies because it is important to reduce the environmental Impact of port operations saving the resources.

All in all, sustainable development is the present and future for ports that want to lead the industry supported by three cornerstones: Operational Excellence, Insightful Collaboration with partners through the supply chain, and top-notch Safety, Health and Environmental practices.

PortForward proposes a holistic approach that will lead to a smarter, greener, and more sustainable port ecosystem and which will include the following features:

[30] https://dataports.iti.upv.es/nc/index.php/s/y8KX3YdM3iYeRLg, (accessed: 14.12.2020)

The introduction of an Internet of Things (IoT) concept for port assets (infrastructure, vehicles, cargo, people):

The socio-economic analysis of the port interface with its surrounding area and the port-city, as well as the rest of the logistics value chain.

Summary

Like the DataPorts, PortForward is defined in the domain of EU ports. A potential collaboration on the design of platforms as well as the descriptions and realization of the pilots and case studies is foreseeable.

Corresponding/Participating Partners

FHG (Fraunhofer Institute), not the same group as in DataPorts.

2.3.2 Smartship

A data analytics, decision support and circular economy-based multi-layer optimization platform towards a holistic energy efficiency, fuel consumption and emissions management of vessels.

Programme

H2020-EU.1.3.3. - Stimulating innovation by means of cross-fertilization of knowledge

Status

Started at 01.04.2019 (48 Months)

Link

https://smartship2020.eu/

Objective

SmartShip aims to offer a multi-layer optimization in the fields of fuel consumption, energy efficiency and emissions control management, in full respect to the implementation of the requirements of maritime sector regulations and considering applications of circular economy concepts in the maritime as well.

Knowledge exchange between the partners that are already involved in the maritime sector, the ICT technology industry partners and the academia partners is one of the major SmartShip's objectives and will be materialized through corresponding secondments during the whole project's timespan.

SmartShip will capitalize on available COTS technologies and will deliver an ICT and IoT-enabled holistic cloud-based maritime performance and monitoring system, for the entire lifecycle of a ship, aimed to optimize energy efficiency, emissions reduction and fuel consumption, whist introducing circular economy concepts in the maritime field.

Summary

In fact, SmartShip similar to the DataPorts is defined in the maritime domain. However, SmartShip mainly focuses on developing a platform towards a holistic energy efficiency. SmartShip aims to develop a cloud-based maritime performance and monitoring system for the entire lifecycle of a ship, therefore, the main objective of the SmartShip is not to transfer data with other source in way that is planned in the context of DataPorts Platform. In case of the definition of the pilots and case studies, a collaboration is foreseeable.

Corresponding/Participating Partners

No common participants with DataPorts.

2.3.3 BigDataStack

High-performing, data-centric stack for Big Data applications and operations.

Programme

H2020-EU.2.1.1. - INDUSTRIAL LEADERSHIP - Leadership in enabling and industrial technologies - Information and Communication Technologies (ICT)

Status

Started at 01.01.2018 (36 Months)

Link

https://bigdatastack.eu/

Objective

BigDataStack aims at providing a complete infrastructure management system, which will base the management and deployment decisions on data from current and past application and infrastructure deployments. This complete infrastructure management system is delivered as a full "stack" that facilitates the needs of operation data and application.

- Data-driven infrastructure management system which ensures computing, storage and networking resources management will be fully efficient and optimized.
- Data as a service promotes automation and quality and ensures that the provided data are meaningful.
- Application dimensioning workbench facilitates data-focused application analysis.
- Process modelling and optimization framework allowing the flexible, functionality-base process modelling.
- Data toolkit enabling open and extensibility.

Summary

Like DataPorts, BigDataStack benefits from a data driven system. Apart from the differences in domain and main functionalities, DataPorts project may leverage from the specification of the BigDataStack management system to further design and develop different functionalities of its platform.

Corresponding/Participating Partners

No common participants with DataPorts.

2.3.4 Pixel

Port IoT for Environmental Leverage.

Programme

H2020-EU.3.4. – SOCIETAL CHALLENGES – Smart, Green and Integrated Transport

Status

Started at 01.05.2018 (36 Months)

Link

https://pixel-ports.eu/

Objective

PIXEL is the first smart, flexible and scalable solution for reducing environmental impacts while enabling the optimization of operations in port ecosystems through IoT.

PIXEL enables a two-way collaboration of ports, multimodal transport agents and cities for optimal use of internal and external resources, sustainable economic growth, and environmental impact mitigation,

towards the Ports of the Future. Built on top of the State-of-the-Art interoperability technologies, PIXEL centralizes data from the different information silos where internal and external stakeholders store their operational information. PIXEL leverages an IoT based communication infrastructure to voluntarily exchange data among ports and stakeholders to achieve an efficient use of resources in ports.

Summary

Among the identified H2020 relevant projects, PIXEL is one of the projects that DataPorts might considerably benefit from. DataPorts will learn from PIXEL technical and business approaches of the application of IoT to seaports. Specifically, in the context of developing two main components of the DataPorts Platform: Data Access component and Semantic Interoperability component, several technical details that are already used in the PIXEL project may be adapted and used for the purpose of developing the DataPorts Platform.

Corresponding/Participating Partners

- UPV (Universitat Politecnica de Valencia)
- PRO (Prodevelop)
- ThPA (Thessaloniki Port Authority SA)

2.3.5 TransformingTransport

Transforming Transport.

Programme

H2020-EU.2.1.1. - INDUSTRIAL LEADERSHIP - Leadership in enabling and industrial technologies - Information and Communication Technologies (ICT)

Status

Started 1 January 2017 - End 31 July 2019

Link

https://transformingtransport.eu/

Objective

The Transforming Transport project will demonstrate, in a realistic, measurable, and replicable way the transformations that Big Data will bring to the mobility and logistics market. To this end, Transforming Transport, validates the technical and economic viability of Big Data to reshape transport processes and services to significantly increase operational efficiency, deliver improved customer experience, and foster new business models. Transforming Transport will address seven pilot domains of major importance for the mobility and logistics sector in Europe:

- Smart Highways.
- Sustainable Vehicle Fleets.
- Proactive Rail Infrastructures.
- Ports as Intelligent Logistics Hubs.
- Efficient Air Transport.
- Multi-modal Urban Mobility.
- Dynamic Supply Chains.

The Transforming Transport consortium combines knowledge and solutions of major European ICT and Big Data Technology Providers together with the competence and experience of key European industry players in the mobility and logistics domain.

Summary

Transforming Transport project like DataPorts project is defined in the mobility domain. Insights and tools from the project in general, and the pilots in seaports, will be considered for the design and implementation of the platform. The main aim of Transforming Transport project is to refine transport processes and services to increase efficiency mainly based on Big Data analytics. The partners participating in the WP3 of the DataPorts project (especially in the context of Big Data analytics) may benefit from the concepts of this project.

Corresponding/Participating Partners

- ITI (Instituto Tecnológico de Informática)
- UPV (Universitat Politecnica de Valencia)
- UDE (Universitaet Duisburg-Essen)
- VPF (Fundacion de la Communidad Valenciana para la Investigacion, Promocion y Estudios Comerciales de Valenciaport)

2.3.6 INTER-IoT

Interoperability of Heterogeneous IoT Platforms.

Programme

H2020-EU.2.1.1. - INDUSTRIAL LEADERSHIP - Leadership in enabling and industrial technologies - Information and Communication Technologies (ICT)

Status

Started 1st January 2016, finished 31st December 2018.

Link

https://inter-iot.eu/

Objective

The overall goal of the INTER-IoT project is to provide an interoperable and open IoT framework, with associated engineering tools and methodology, for seamless integration of heterogeneous IoT platforms, regardless of the application domains.

The objectives are as follows:

- Design and Implementation of an Open Cross-Layer Framework for Interoperability of IoT Platforms.
- Definition of Techniques and Tools for interoperability at the different IoT Platform Layers.
- Definition of a CASE-driven Engineering Methodology Driving the Application of the IoT Platform Interoperability Framework.
- Design and Implementation of an Integrated Interoperable Open Platform for Transport and Logistics in Port Environments (INTER-LogP).
- Design and Implementation of an Integrated Interoperable Open Platform for Mobile Health Monitoring (INTER-Health).
- Successful completion of field trials: The INTER-IoT developed pilots (Objectives 4 and 5) will be further evaluated in the two proposed application domains in: Nichelino (Turin) (IT) for m-health and Valencia (ES) for port transportation. Additionally, a cross-domain use case and associated field trial will be performed to proof extendibility and interoperability of platforms from different application domains. Moreover, the project will analyse the provided solutions from the perspective of relevant stakeholders by considering their specific benefits, requirements, and constrains, and involving stakeholders from other application domains to evaluate the extendibility of the results.

Summary

Concerning one of the two proposed application domains of the INTER-IoT project, namely port transportation (in Valencia), in the context of the DataPorts Platform (specifically the definition of use cases), a collaboration is foreseeable. Knowledge and techniques from INTER-IoT may be used and applied to DataPorts for connection to existing heterogeneous port infrastructures and collection of data.

Corresponding/Participating Partners

- UPV (Universitat Politecnica de Valencia)
- PRO (Prodevelop)
- VPF (Fundacion de la Comunidad Valenciana para la Investigacion, Promcion y Estudios Commerciales de Valenciaport)

2.3.7 SYNERGY

Big Energy Data Platform and AI Analytics Marketplace for new viable solutions.

Programme

H2020-EU.2.1.1. - INDUSTRIAL LEADERSHIP - Leadership in enabling and industrial technologies - Information and Communication Technologies (ICT)

Status

1 January 2020 - 30 June 2023

Link

https://www.synergyh2020.eu/

Objective

SYNERGY introduces a novel reference Big Data architecture and platform that leverages data, primary or secondarily related to the electricity domain, coming from diverse sources (APIs, historical data, statistics, sensors/ IoT, weather, energy markets and various other open data sources) to help electricity stakeholders to simultaneously enhance their data reach, improve their internal intelligence on electricity-related optimization functions, while getting involved in novel data (intelligence) sharing/trading models, in order to shift individual decision-making at a collective intelligence level.

To this end, SYNERGY will develop a highly effective a Big Energy Data Platform and AI Analytics Marketplace, accompanied by Big Data-enabled applications for the totality of electricity value chain stakeholders (altogether integrated in the SYNERGY Big Data-driven EaaS Framework).

Summary

SYNERGY project mainly delivers a Big Data platform and AI analytics marketplace regarding the data received from various data sources.

In the context of the AI analytics, a collaboration is foreseeable. However, since the SYNERGY project has started in January, currently not much information is available.

Corresponding/Participating Partners

ICCS (Institute of Communication and Computer Systems), not the same group.

2.3.8 Sauron

Scalable multidimension Al sitUation awaReness sOlution for protectiNg European ports.

Programme

- H2020-EU.3.7.4. - Improve cyber security
- H2020-EU.3.7.2. - Protect and improve the resilience of critical infrastructures, supply chains and transport modes

Status

1 May 2017 - 30 September 2020

Link

https://www.sauronproject.eu/

Objective

SAURON project addresses prevention, detection, response, and mitigation of the combination of physical and cyber threats to the critical infrastructure of Europe and put the focus on protection of EU Ports under Transport Infrastructure and means of transportation type of CI.

SAURON proposes to ensure an adequate level of protection and resilience against physical, cyber and a combined threat for the EU ports and limiting, as far as possible, the detrimental effects for the society and citizens of a declared attack.

The vision of SAURON is to provide a multidimensional yet installation-specific Situational Awareness (SA) platform to help port operators anticipate and withstand potential cyber, physical, or combined threats to their freight and cargo business and to the safety of their employees, visitors, passengers and citizens in the vicinity.

Summary

In the context of the DataPorts Platform (especially within the WP4), security and privacy aspects of data exchanged between different organizations are of importance. SAURON aims to deliver an adequate level of protection and resilience against cyber physical threats for the EU ports.

The partners participating in WP4 may benefit from the concepts introduced in the context of the SAURON project.

Corresponding/Participating Partners

VPF (Fundacion de la Comunidad Valenciana para la Investigacion, Promocion y Estudios Comerciales de Valenciaport).

2.3.9 Cyber-MAR

Cyber preparedness actions for a holistic approach and awareness raising in the MARitime logistics supply chain.

Programme

H2020-EU.3.7.4. - Improve cyber security

Status

1 September 2019, 31 August 2022

Link

https://www.cyber-mar.eu/

Objective

Cyber-MAR is an effort to fully unlock the value of the use of cyber range in the maritime logistics value chain via the development of an innovative simulation environment adapting in the peculiarities of the maritime

sector but being at the same time easily applicable in other transport subsectors. A combination of innovative technologies are the technology enablers of the proposed Cyber-MAR platform which is not only a knowledge-based platform but more importantly a decision support tool to cybersecurity measures, by deploying novel risk analysis and econometric models. CSIRTs/CERTs data collected will be analysed and feed the knowledge-based platform with new-targeted scenarios and exercises. Through Cyber-MAR, the maritime logistics value chain actors will increase their cyber-awareness level; they will validate their business continuity management minimizing business disruption potential. Cyber-MAR will act as a cost-efficient training solution covering the maritime logistics value chain.

Summary

Cyber-MAR aims to increase the cyber-awareness level for the maritime logistics value chain actors. The partners participating in the WP4 and WP7 of the DataPorts project may benefit from the concepts introduced within the Cyber-MAR project, and therefore a collaboration is foreseeable.

Corresponding/Participating Partners

VPF (Fundacion de la Comunidad Valenciana para la Investigacion, Promocion y Estudios Comerciales de Valenciaport)

3 INFRASTRUCTURE BENCHMARKING

3.1 ANALYSIS OF CURRENT INFRASTRUCTURE AND PROCESSES

For the management and control of all the activities that take place in a port, it is required a large infrastructure built from multiple components that carry out their own processes and generate or consume information provided by any of the elements integrated in it. The DataPorts project aims to provide a mechanism to share this data among all the port ecosystem, so current processes should be identified to take advantage of the new platform.

In this section are described the current processes that somehow are related with the pilot development and evaluation.

3.1.1 TRAXENS

ID	TRX
Short description	Smart Container Service
Process description	• Smart containers produce and communicate physical data remotely. The platform collects, qualifies, enriches, and certifies data to provide organizations with real-time insight into their end-to-end global supply chains. • Geolocation Data: Automatically detect transport mode, calculate mileage, and compute CO_2 emissions with global, high precision geolocation. • Geofencing and predictive ETA: Creation and geofencing of areas of interest to accurately estimate arrival dates and times. • Shock detection: Detection of shocks on containers and severity assessment through our shock algorithm. • Door opening: Alerts to unexpected container openings during the cargo journey. • Temperature and humidity fluctuation: Monitoring of the cargo remotely and measurement of environmental data, such as temperature and humidity, inside the asset. TRAXENS-Hub platform collects the data for the tracking devices and transforms data and distributes the right information to the right stakeholder at the right time. This is described in the following three steps: 1. Data Production Traxens devices are permanently attached to assets—dry containers, reefer containers, wagons, or trailers. The devices capture and transmit real-time data from anywhere in the world. 2. Data Aggregation Asset data is sent to the Traxens IoT platform, which aggregates high volumes of data traffic from devices and many other external sources. 3. Data Usage Traxens analyses, cleanses, and enriches data, converting it into valuable insight for time-sensitive, data-driven decisions.
Department(s) involved	IoT solution provider

Type	Operational
Involved technologies	• Linux Server ○ Description: Operating System for the PCS ○ Provider: Debian • Cassandra & postgreSQL ○ Description: Database server for the PCS ○ Provider: Datastacks • Apache Tomcat & Apache HTTP server ○ Description: application server for hosting the application and web services ○ Provider: open source by Apache Foundation • API REST ○ Description: The API services are based on REST

Table 12 – TRAXENS Smart Container Service

3.1.2 Valencia Port

This section describes the main processes related to the DataPorts Platform carried out in the Port of Valencia.

3.1.2.1 VPF_PCS

ID	vpf_pcs
Short description	Port Community System processes
Process description	The information flow around Valenciaport is very complex and involves a huge number of agents. Each container movement requires multiple communications between the Port Community members, creating a complex information network. ValenciaportPCS is an open and neutral electronic platform that allows a safe and smart information exchange between public and private agents in order to improve our competitive position as a Port Community. ValenciaportPCS offers more than 20 transactional and informative services to more than 600 companies and public entities. ValenciaportPCS optimizes, manages, and automatizes port and logistic processes in an efficient way by a single data transfer as well as connecting transport and logistic networks. ValenciaportPCS services supply advanced management procedures to its Users, which are: a better transaction efficiency, resources optimization, process automation, costs saving, error reduction, time saving, and better client support. ValenciaportPCS is integrated with the world's major carriers through the INTTRA and GT Nexus technological platforms. It pools together the services of these two platforms, thereby offering a single gateway to the world's most important ocean carriers.

	The high-level services or processes provided are: • Port Calls Management: allows port call and mooring authorisation requests to be processed directly with the Port Authority of Valencia (PAV) and the Harbourmaster's office as well as for the corresponding authorisations for these requests to be received directly. • Dangerous Goods Management: allows shipping agents to present authorisation requests directly to the Port Authority of Valencia (PAV) and the Harbourmaster's office for the loading, discharging and transit of dangerous goods in the ports managed by the PAV. The service also serves to present the respective notifications required for dangerous goods remaining in vessels during port calls at the ports managed by the PAV. • Loading and Discharge Orders: permits shipping agents to send the vessel loading and discharge lists to the container and car terminals and obtain the respective confirmations of the loading and discharge of the containers and cars figuring on these lists from the terminals. • Road Transport Management: allows the agents involved in the transportation of goods by land to generate and manage the transport orders, cargo acceptance and delivery orders required to transport these goods within the port facilities managed by Valenciaport, and to notify the delivery and acceptance of containers in the container terminals and/or depots. • Rail Transport Management: allows the agents involved in the transportation of goods by rail to generate and manage the transport orders, cargo acceptance and delivery orders required to transport these goods within the port facilities managed by Valenciaport, and to notify the delivery and acceptance of containers in the container terminals and/or depots. • Cargo Tracking: allows Users of ValenciaportPCS to obtain information to track and trace their shipments, such as the current status of their cargo, transhipments carried out and/or documents processed. The platform also enables Users to integrate this information into their systems to present it to their customers. • Goods Declaration: allows shipping agents to present import and export cargo manifests directly to the Port Authority of Valencia (PAV) and the Spanish State Tax Agency as well as to amend such documents when needed pursuant to established procedures. • VGM: With the VGM Service of ValenciaportPCS, Shippers, Freight Forwarders, Shipping Agents, Shipping Companies and Terminals have the fastest and most efficient way to notify, receive and consult the Verified Gross Weight. The VGM Service of ValenciaportPCS is made up of a web application and the EDI / XML messaging exchange of the VERMAS message so that shippers and their logistics Service Pfinroviders can notify the VGM to shipping agents and shipping companies and terminals. Shipping Agents and Shipping Companies registered in ValenciaportPCS can notify the VGM to the terminals through the Web application or by electronic messaging directly from their management system. • Customs Information: customs clearance of containers and the consignments declared in the cargo manifests to be cleared by the customs office.

	• Equipment Status: allows ValenciaportPCS Users to know the status of their containers and vehicles both at load and unload operations. This service provides ValenciaportPCS Users automated paperless customs control of containers and vehicles for export and transhipment. • Departures and Arrivals / Schedule: offering a single source of information on the departures and arrivals of vessels of the world's major ocean carriers at the ports managed by Valenciaport. • Bookings: allows its clients to send bookings through electronic messaging and/or using the ValenciaportPCS application. Users of the bookings service save time whilst improving the accuracy of data. • Shipping Instructions: allows Users of ValenciaportPCS to automatically carry out the documentary processes to compile cargo manifests and create bills of lading (B/Ls) with the world's major ocean carriers. More information at: https://www.valenciaportpcs.com/en/our-services/
Department(s) involved	Port Authority of Valencia – PCS and Commercial Department Technological provider
Type	Operational

Involved technologies	• Windows Server
	  ○ Description: Operating System for the PCS
	  ○ Provider: Microsoft
	• SQL Server
	  ○ Description: Database server for the PCS
	  ○ Provider: Microsoft
	• Internet Information Services (IIS)
	  ○ Description: Web server for hosting the application and web services
	  ○ Provider: Microsoft
	• API REST / SOAP
	  ○ Description: The PCS API services are based on REST and SOAP

Table 13 – VPF_PCS process in the Port of Valencia

3.1.2.2 VPF_CONPESO

ID	vpf_conpeso
Short description	Verified Gross Mass
Process description	In 2014, the International Maritime Organization (IMO) amended (MSC.1Circ.1475) the Convention on the Safety of Life at Sea (SOLAS) to require, as a condition of loading a full container on a ship for export, that the container has a verified the gross weight. This requirement became mandatory worldwide in 1st July 2016. The shipper became the responsible

	for obtaining the verified gross weight of a full container and communicating it to the shipping company. The regulation prescribes two methods by which the loader can obtain the verified gross weight of a full container: • Method 1: once the packaging and sealing of a container has been completed, the loader can weigh, or arrange for a third party to weigh the full container. • Method 2: the shipper or, by shipper's order, a third party may weigh all packages and cargo items, including the weight of pallets, stowage wood and other packaging and fastening materials packed in the container, and add the tare weight of the container that appears on the door of the container to the sum of the individual weights of the contents of the container. (subject to national regulation, currently in draft) With respect to both method 1 and 2, the weighing equipment used must meet the precision standards and requirements of the State in which the equipment is being used. Weight estimation is not allowed. The full text of the regulation is available on the Internet through the following link[31]. There are also some guides published by the IMO regarding the verified gross weight of a container transporting goods[32], which can be consulted on the Internet through the following link[33]. To attend this requirement, Fundación Valenciaport created in 2016 conPESO. conPESO is a marketplace platform that facilitates the compliance of SOLAS regulations on weighing containers for the port logistics community. The platform offers Users an effective solution to allow containers to arrive at the port with verified gross weight, reducing last minute incidents or delays at container terminals or the appearance of congestion situations. Furthermore, conPESO offers a fast and automated method so that the verified gross weight reaches the shipping line and the terminal, and allows the port to be more competitive[34] through a connection with ValenciaportPCS platform through SOAP web services and sending the VERMAS message in XML format. The processes included are: - System administration processes for the system administrator - User registration processes for the user or company/system administrators - VGM container request process for the VGM requestor companies

[31] http://www.worldshipping.org/industry-issues/safety/SOLAS_CHAPTER_VI_Regulation_2_Paragraphs_4-6.pdf

[32] MSC.1/Circ.1475 Annex. Guidelines regarding the verified gross mass of a container carrying cargo. International Maritime Organisation.

[33] https://www.worldshipping.org/industry-issues/safety/MSC_1-Circ_1475_-_Guidelines_Regarding_The_Verified_Gross_Mass_Of_A_Container_Carrying_Cargo_-Secretariat-.pdf

[34] https://conpeso.com

	- Credit payment processes for the VGM requestor companies - VGM invoicing processes for the system administrator, VGM requestor and weight scale operators - Weight registration processes for the weight scale operator - VGM communication process to ValenciaportPCS
Department(s) involved	Fundación Valenciaport - Digital Transformation Department. Technology provider
Type	Operational
Involved technologies	PHP application Apache Web Server • MariaDB database • Linux managed Operating System. For the DataPorts Platform sandbox, this application is being evolved to use the following technologies: • Angular 9 Single Web Page Application • API REST Server in Nodejs • MongoDB • Ubuntu Server

Table 14 – VPF_CONPESO process in the Port of Valencia

3.1.2.3 VPF_PI

ID	vpf_pi
Short description	Port industrial platform
Process description	Industries and big infrastructures where several systems, devices and sensors are deployed need a powerful tool which allows them to store big amounts of data in a fast way. Furthermore, it should have the features to structure the information and analyse the information. One of the most common tools in the industrial environment is PI software from OSI, which is also being used in the port of Valencia. The PI System is a suite of software products that are used for data collection, historicizing, finding, analysing, delivering, and visualizing. It is marketed as an enterprise infrastructure for management of real-time data and events. Data can be automatically collected from many sources (control systems, lab equipment, calculations, manual entry or custom software). Most information is gathered using one of the many OSIsoft and third-party PI Interfaces. Users can then access this information using a common set of tools (such as Microsoft Excel, web browser, or external tools) and look for correlations. In the port of Valencia there are currently some sensors and systems connected to the industrial platform for monitoring environmental data, the gate access system, electric smart meters and AIS system.

Department(s) involved	Port Authority of Valencia - IT Systems and OT (Operational Technology) Departments
Type	Operational
Involved technologies	The OSIsoft PI products include: • PI Data Archive: The PI Data Archive collects, stores, and organizes data from data sources, providing an information infrastructure. The PI Server also includes tools for analytics, alerts, and auditing. The PI Server may be connected to almost any existing automation, lab, or information system. Operators, engineers, managers, and other plant personnel can use client applications to connect to the PI Server to view data stored in the PI Server or in external data archive systems. • PI Asset Framework (AF): PI AF allows the definition of consistent representations of organizational assets and/or equipment and uses these representations in analyses that yield critical and actionable information. It allows the creation of digital twins of sensors and objects. • PI Web API: The PI Web API is a RESTful interface to the PI system. It gives client applications read and write access to their PI Asset Framework and PI Data Archive over HTTPS. • PI Vision: A simple web-based tool for quick ad hoc displays. PI Vision is a thin client tool and can be used anywhere an Internet connection is available. It allows to create dashboards easily including all the data from the sensors. • PI Connectors and PI Interfaces: PI Connectors and PI Interfaces are designed to gather data from a data source, convert to a PI readable format, and send to the PI Data Archive to be stored. Unlike Interfaces, Connectors are able to collect and send metadata to the PI Asset Framework in addition to real-time data. There are more than 400 connectors and interfaces, including OPC, Modbus, RDBMS, and Universal File Loader (UFL). • PI Integrators: PI Integrators is a line of web-based applications that prepare and send time-series data to third-party analytics platforms. The PI Integrator cleanses, augments, shapes, and transmits data from PI into common business intelligence and data warehouse tools, such as Power BI, Tableau, Hadoop, Microsoft Azure, and Amazon Web Services. The infrastructure needed is: • Windows Server o Description: Operating System for PI o Provider: Microsoft • PI Data Archive o Description: Database for PI data o Provider: OSIsoft • SQL Server o Description: Database for configuration data o Provider: Microsoft

	• PI Web API o Description: PI API services are based on REST

Table 15 – VPF_PI process in the Port of Valencia

3.1.2.4 VPF_CUSTOMS

ID	vpf_customs
Short description	Customs
Process description	The Tax Agency is entrusted with the effective application of the state and customs tax system, as well as those resources of other national public administrations or the European Union whose management is entrusted to it by law or by agreement. Therefore, all imports and exports made through the port must be monitored and approved through the appropriate documentation. The goods owner or their representative has to declare all import, export or transhipment good when arriving to the European Union. Then they receive the customs release or indications for some inspections. Tax Agency is providing open data statistical information about foreign trade including anonymised container trade. https://www.agenciatributaria.es/AEAT.internet/en_gb/datosabiertos/catalogo/hacienda/datos_estadisticos_Comercio_Exterior.shtml
Department(s) involved	Importer/exporter, Freight forwarders Customs
Type	Operational
Involved technologies	The tax agency has an electronic office[35] to submit all the information related to: o Summary tax returns for unloading o Notification of movement of goods in temporary storage (G5) o Cargo manifests o Entry summary declaration (ENS) o Exit summary declaration (EXS)

Table 16 – VPF_CUSTOMS process in the Port of Valencia

3.1.2.5 VPF_TRADELENS

ID	vpf_tradeLens
Short description	TradeLens

[35] https://www.agenciatributaria.gob.es/AEAT.sede/en_gb/Inicio/Procedimientos_y_Servicios/Aduanas/Presentacion_de_declaraciones_sumarias_y_manifiestos_de_carga/Presentacion_de_declaraciones_sumarias_y_manifiestos_de_carga.shtml

Process description	TradeLens is a proprietary global trade digital platform jointly developed by IBM and Maersk. The platform was officially launched in August 2018, and currently has the participation of an important group of shipping companies, which represent 60% of the operating fleet. Data sharing is the essence of the TradeLens platform. As a cargo flows from origin to destination through the supply chain, TradeLens participants involved in the transportation and logistics of that cargo provide data to the platform such as logistics milestones, movement information, and business documents. The TradeLens platform, through an authorization and permission model, enables or restricts access to this data to other participants. The model promotes secure and rapid access to supply chain information and ensures that sensitive business information is not available to competitors or other unauthorized parties. TradeLens provides visibility into movements of international container shipments. Includes application programming interfaces (APIs) and user interfaces for: • Publishing and subscribing to event data describing the physical progress of cargo through the supply chain • Verifying compliance with regulatory objectives • Viewing events and objectives of previous points • Managing Users and access permissions TradeLens defines several roles depending the type of company. The Port Authority has the PCS role and it is providing data on gate entry and exit and container loading/discharging in the port. As a data consumer, Port Authority has access to most all the events regarding all the containers that go through the port though an API or a web dashboard. This data is only for internal use and cannot be shared.
Department(s) involved	IBM
Type	Operational
Involved technologies	TradeLens is accessible through an open API and a Web UI interface, which links the ecosystem through open standards. The digital platform is powered by Blockchain technology in the cloud, using the Hyperledger Fabric solution (private and permissioned Blockchain network), which allows information to be shared and collaborate safely. TradeLens employs a cloud approach. The APIs offer the functionality of TradeLens. These are documented on the web https://docs.tradelens.com/ and are aimed at the general public[36]. The platform services layer uses containerized and horizontally scalable microservices, linked together through a publish/subscribe service.

[36] https://platform.tradelens.com/documentation/swagger/

	Multiple layers of persistence are used, including object storage, document databases, relational databases, and Blockchain, depending on the type of information being pursued and how it is accessed. Managed services, including PaaS and DBaaS components, are also used to take advantage of cloud economies of scale and drive efficiencies.

Table 17 – VPF_TRADELENS process in the Port of Valencia

3.1.3 Thessaloniki Port

In this section there are described the processes carried out in the Port of Thessaloniki.

3.1.3.1 THPA_01_VESSEL_PREANNOUNCEMENT

ID	THPA_01_vessel_preannouncement
Short description	Vessel preannouncement
Process description	This process presents the details about a vessel preannouncement (CALINF EDIFACT message), sent to ThPA through a 2-step process. There are two different ways to conclude step 1, but step 2 is common for both. Step 1: a. A stakeholder (e.g., shipping agent) fills the details of the call on a web form. The form is served by a ThPA server, running Apache Tomcat, over a secure (SSL) connection. Only registered Users of ThPA's TOS can use this form, so the form requires successful login. After the data is filled and the form is submitted, a JSP script transforms the entered data into XML format, compliant with a corresponding DTD, and saves it in plaintext format in a local directory on the server. b. An SSL client, used by the stakeholder (shipping agent), connects to a service called *XML Server*, running as an NT service, on the same ThPA server. The service listens on a custom TCP port. After a secure connection is established successfully (a process that includes authentication of the stakeholder's account), the stakeholder's SSL client uploads the vessel call info in XML format. *XML Server* receives the XML document and saves it in plaintext in the same local directory as 1a. Step 2: A second service (also running as an NT service), called *XML Document Server*, continuously polls the aforementioned directory. When it discovers the newly-saved file, it loads the file, parses it and validates it against a DTD. If all is correct, it saves all the XML document information in the TOS database (hosted on a **different** server). If there is a parsing or a DTD validation error, it fails, moving the XML file to a different local directory. Either way, the stakeholder is notified by email – since only registered Users can use this service, the company uploading the information has provided an email address.

	Finally, as the XML information has been parsed and validated, it is stored in the TOS database, but flagged as '*Pending*'. A ThPA TOS user can load the message in the *Document Deposit* application (a module of the TOS), and flag it as '*Accepted/Final*'. After that, all the information about the call is treated by the TOS as operational data. The application, like all of the modules of the TOS, runs locally on a Windows workstation and communicates with the main TOS database via an ODBC connection.
Department(s) involved	Container Terminal
Type	Operational

Involved technologies	• Windows 2003 Server
	  ○ Description: The server on which the NT services *XML Server* and *XML Document Server* run
	  ○ Provider: Microsoft
	  ○ Versions: 2003 Server Std Edition Service Pack 2
	  ○ Infrastructure: Operations
	  ○ Used by the services/applications: NT Services (*XML Server* and XML *Document Server*)
	• NT Services (*XML Server* and *XML Document Server*)
	  ○ Description: Custom software for accepting the XML messages
	  ○ Provider: TREDIT S.A.
	  ○ License: Private
	  ○ Versions: 1.6 (build 0.101)
	  ○ Infrastructure: Operations
	  ○ Used by the services/applications: TOS
	• Windows 2012R2 Server
	  ○ Description: The main TOS Server, hosting the TOS Database
	  ○ Provider: Microsoft
	  ○ Versions: Windows 2012R2 Server
	  ○ Infrastructure: Operations
	  ○ Used by the services/applications: TOS, NT Services (*XML Server* and XML *Document Server*)
	• SQL Server Database
	  ○ Description: Database server, hosting the central TOS Database
	  ○ Provider: Microsoft
	  ○ Version: SQL Server 2012 (11.0.7493)
	  ○ Infrastructure: Operations
	  ○ Used by the services/applications:
	  ▪ NT Services (XML Server and XML Document Server)
	  ▪ TOS
	• Apache Tomcat

	○ Description: Web server for hosting the web form ○ Provider: Apache Software Foundation ○ License: Apache License 2.0 ○ Version: 5.5 ○ Infrastructure: Operations ○ Used by the services/applications: ▪ Web Application ▪ Web Services (services / database operations)

Table 18 – THPA_01_VESSEL_PREANNOUNCEMENT process in the Port of Thessaloniki

3.1.3.2 THPA_02_TRUCK_ENTRY/EXIT_CONTROL_FREE_ZONE

ID	THPA_02_truck entry/exit control (at Free Zone gates)
Short description	Truck entry / exit control (at Free Zone gates)
Process description	This process describes the entry/exit control of a truck at the Free Zone gates. A truck arriving at the Port to deliver or pick-up goods, must fulfil certain criteria. At Gates 10 and 16 (Free Zone gates) RFID sensors are installed. A truck arriving must have an RFID tag (sticker) on its windshield. The RFID sensor scans the tag. The Java application *EpiGateAccess* reads the tag through the RFID sensor's API, and queries an application called *accesscards*, that contains (among others) information about all ThPA-registered trucks – registration number, related stakeholder, etc. The query is sent in the form of a standard web call to a PHP script (HTTP GET request). The PHP script queries the *accesscards* application database and sends the response in JSON format. Response can be that the RFID is unknown, that the truck is not permitted to enter (blocked), or that all is OK. In the last case, the JSON reply contains the trucks details (registration number, company, etc.), for visual confirmation by the Gate personnel. Either way, the PHP script logs the event in the MySQL database. If the truck is to enter the Container Terminal (and not the Free Zone), then extra steps follow: The truck goes through the GOS (Gate Operating System) Gate, where the container number and the truck's license plates are scanned with OCR. The information from the sensors and cameras is processed by the GOS server and stored in an SQL Server database that interacts with the TOS. Finally, at the Container Terminal Gate, the related paperwork (permit to pick-up or deliver containers) is scanned via a (USB) barcode reader, that transmits the barcode information to the *Gate Control* application (module of FRETIS, the TOS ecosystem), that runs at a Container Terminal workstation). The *Gate Control* application queries via ODBC the central TOS database to confirm the validity of the permit. More details for the gates at container terminal are provided in the process described in 3.1.3.3 (at Container Terminal). When a truck exits the Container Terminal Gate, the related paperwork (permit to pick-up containers) is scanned via a (USB) barcode reader, that transmits the barcode information to the *Gate Control* application (part of FRETIS, the TOS ecosystem), that runs at a Container Terminal workstation).

	The *Gate Control* application queries via ODBC the central TOS database to confirm the validity of the permit. When the truck exits Gate 10 or Gate 16, the same process as described for the entry is followed (RFID tag read by RFID sensor, *EpiGateAccess* reads the tag via the sensor's API, queries *accesscards* web application via HTTP GET request, gets reply in JSON, event logged in MySQL database).
Department(s) involved	Container Terminal / Free Zone
Type	Operational

Involved technologies	• Windows 10 Pro
	  ○ Description: The "server" on which the Java application runs (as a service)
	  ○ Provider: Microsoft
	  ○ Versions: Windows 10 Pro (1909)
	  ○ Infrastructure: Operations
	  ○ Used by the services/applications: EpiGateAccess Java application
	• EpiGateAccess
	  ○ Description: Custom Java application for reading RFID sensors' readings through sensors' API
	  ○ Provider: Epi Systems
	  ○ License: Private
	  ○ Versions: 2.0
	  ○ Infrastructure: Operations
	  ○ Used by the services/applications: *accesscards* Web Application
	• Ubuntu Server
	  ○ Description: Operating System for the *accesscards* Web Application
	  ○ Provider: Canonical Ltd / Ubuntu Foundation
	  ○ License: GPL
	  ○ Versions: 12.04.5 LTS
	  ○ Infrastructure: Operations
	  ○ Used by the services/applications: *accesscards* Web Application
	• MySQL Database
	  ○ Description: The DB of the *accesscards* application with trucks' details
	  ○ Provider: MySQL AB
	  ○ License: GPL v2
	  ○ Versions: 5.5.54
	  ○ Infrastructure: Operations
	  ○ Used by the services/applications:
	  ▪ *accesscards* Web Application
	  ▪ EpiGateAccess
	• Apache Web Server

	○ Description: Web server hosting the *accesscards* application ○ Provider: Apache Software Foundation ○ License: Apache License 2.0 ○ Version: 2.2.22 ○ Infrastructure: Operations ○ Used by the services/applications: ▪ *Accesscards* Web Application • Windows Server 2012 R2 ○ Description: The GOS server ○ Provider: Microsoft ○ Versions: Server 2012 R2 ○ Infrastructure: Operations ○ Used by the services/applications: Gate Operating System • GOS – Gate Operating System ○ Description: A suite of applications, working with GOS Gate devices (OCR, cameras, etc.) to process and store all information collected ○ Provider: CAMCO ○ License: Private ○ Versions: 3.5.7 ○ Infrastructure: Operations • Windows 2012R2 Server ○ Description: The main TOS Server, hosting the TOS Database ○ Provider: Microsoft ○ Versions: Windows 2012R2 Server ○ Infrastructure: Operations ○ Used by the services/applications: TOS, NT Services (XML Server and XML Document Server) • SQL Server Database ○ Description: Database server, hosting the central TOS Database ○ Provider: Microsoft ○ Version: SQL Server 2012 (11.0.7493) ○ Infrastructure: Operations ○ Used by the services/applications: ▪ GOS (Gate Operating System) ▪ FRETIS (TOS)

Table 19 – THPA_02_TRUCK_ENTRY/EXIT_CONTROL_FREE_ZONE process in Thessaloniki's Port

3.1.3.3 THPA_03_TRUCK_ENTRY/EXIT_CONTROL_FREE_ZONE

ID	THPA_03 truck entry/exit control (at Container Terminal)
Short description	Truck entry/exit control (at Container Terminal)

Process description	This process describes the entry/exit control of a truck at the Container Terminal as performed by the dedicated components of TOS. The process is for container pick up and is composed of the following steps: - The truck vehicle arrives empty at the entrance gate. - With the arrival at gate, the responsible personnel checks in his/her application the existence of the container to be picked up at the terminal, the vehicle number is registered (OCR) as well as the truck driver id. - The truck driver is guided to the parking place (the position is printed and handed to the driver on paper) and waits to get loaded. - As soon as the truck is loaded, it moves to the exit control. - The responsible personnel checks (physically and through TOS) the accompanying documents and if everything is fine then the exit is allowed. In a different action a U-turn takes place and corresponding corrective actions. - All actions taken are registered in TOS (automatically or via a UI). The application, like all of the modules of the TOS, runs locally on a Windows workstation and communicates with the main TOS database via an ODBC connection.
Department(s) involved	Container Terminal
Type	Operational
Involved technologies	• Windows 2012R2 Server o Description: The main TOS Server, hosting the TOS Database o Provider: Microsoft o Versions: Windows 2012R2 Server o Infrastructure: Operations o Used by the services/applications: TOS, NT Services (*XML Server* and XML *Document Server*) • SQL Server Database o Description: Database server, hosting the central TOS Database o Provider: Microsoft o Version: SQL Server 2012 (11.0.7493) o Infrastructure: Operations o Used by the services/applications: ▪ NT Services (XML Server and XML Document Server) ▪ TOS

Table 20 – THPA_03_TRUCK_ENTRY/EXIT_CONTROL_FREE_ZONE process in Thessaloniki's Port

3.1.3.4 THPA_04_TAS

ID	THPA_04_TAS
Short description	Truck Appointment System / Bookings

Process description	This process describes the booking process for container pickup and delivery. Stakeholder (e.g., trucking company), logs in the Web Application https://tas.thpa.gr using a standard web browser. The connection is served over SSL (HTTPS). User is presented with a web form, where they enter their preferred timeslot to drop-off or pick-up containers, and the container IDs. After submitting, the information is sent via an HTTP POST request to the web server, where PHP scripts validate and processes the data. If all checks are successful, the booking details (user, slot, containers) are stored in an SQL Server database. The booking gets a unique ID, which is emailed to the stakeholder. The option of editing or cancelling a booking is also provided in corresponding forms. In these cases, the booking details are queried from the SQL database by the corresponding PHP scripts and presented to the client (browser). Any modifications are again processed by PHP scripts and stored to the SQL database.
Department(s) involved	Container Terminal
Type	Operational

Involved technologies	• Windows 2012R2 Server
	  ○ Description: The main TOS Server, hosting the TAS Database
	  ○ Provider: Microsoft
	  ○ Versions: Windows 2012R2 Server
	  ○ Infrastructure: Operations
	  ○ Used by the services/applications: TAS, TOS
	• SQL Server Database
	  ○ Description: Database server, hosting the central TOS Database
	  ○ Provider: Microsoft

Table 21 – THPA_04_TAS process in the Port of Thessaloniki

3.1.3.5 THPA_05_COREOR

ID	THPA_05_COREOR submission /issue
Short description	COntainer RElease ORder (COREOR) submission /issue
Process description	This process refers to the submission of COREOR in case of container pick up an is composed of the following steps: Following the completion of unloading activities and customs obligations, the clients of the container terminal can proceed to action to pick up the container. The COREOR request can be submitted electronically via the e-Documents module of TOS or by the physical presence at the terminal. As the latter is the most common one, the emphasis is given on this.

	The client of the container terminal submits in paper a legal document (called "diataktiki") that shows the name of the person who is liable to pick up the container and contains information about the container ID, number of IFSCUM et cans asks for COREOR. The responsible personnel checks a) the "diataktiki" (stamps, dates, container ID etc.) in his/her applications in TOS, b) the invoice and if everything is ok then issues the COREOR. It has a unique reference number. This reference number is provided by TOS. COREOR contains information about customs classification (EU, third countries etc.), the person that is liable to pick up (name of the company VAT number etc.) that must be the same with those declared in "diataktiki" and information about container id, size, load/empty and contents. The application, like all of the modules of the TOS, runs locally on a Windows workstation and communicates with the main TOS database via an ODBC connection.
Department(s) involved	Container Terminal
Type	Operational

Involved technologies	• Windows 2012R2 Server
	  ○ Description: The main TOS Server, hosting the TOS Database
	  ○ Provider: Microsoft
	  ○ Versions: Windows 2012R2 Server
	  ○ Infrastructure: Operations
	  ○ Used by the services/applications: TOS, NT Services (*XML Server* and XML *Document Server*)
	• SQL Server Database
	  ○ Description: Database server, hosting the central TOS Database
	  ○ Provider: Microsoft
	  ○ Version: SQL Server 2012 (11.0.7493)
	  ○ Infrastructure: Operations
	  ○ Used by the services/applications:
	  ▪ NT Services (XML Server and XML Document Server)
	  ▪ TOS
	
	In case the client wants to submit electronically the COREOR then the following systems are also involved.
	• Windows 2003 Server
	  ○ Description: The server on which the NT services *XML Server* and *XML Document Server* run
	  ○ Provider: Microsoft
	  ○ Versions: 2003 Server Std Edition Service Pack 2
	  ○ Infrastructure: Operations
	  ○ Used by the services/applications: NT Services (*XML Server* and XML *Document Server*)
	• NT Services (*XML Server* and *XML Document Server*)

	○ Description: Custom software for accepting the XML messages ○ Provider: TREDIT S.A. ○ License: Private ○ Versions: 1.6 (build 0.101) ○ Infrastructure: Operations ○ Used by the services/applications: TOS

Table 22 – THPA_05_COREOR process in the Port of Thessaloniki

3.1.3.6 THPA_06_STATISTICS

ID	THPA_06_Statistics
Short description	Statistics used for internal purposes and official reporting
Process description	This process refers to the creation of statistics and KPIs. ThPA has well established procedures of collecting statistical data, obliged to be reported in several Authorities and Organisations like Ministry of Marine and Aegean, EL. STAT (Hellenic Statistical Authority), ESPO, Harbour Master. Furthermore, when there is a requirement /need from the Port Management for collecting of KPIs in relation to a specific port activity or a group of processes then ad hoc actions are taken. TOS/TAS and other systems generate operational data in various formats, targeted actions are introduced in order to process this data and extract interesting statistical results. Depending on the KPIs that need to be calculated, additional efforts are required to process the data for their definitions. Based on its capacity, ThPA publishes selected statistics in its web site at the end of each year for standard dissemination and promotion purposes. The custom application *Statistics* is used, a client-server application, where information entered by Users is stored (via ODBC connection) to an SQL Server database.
Department(s) involved	Container Terminal
Type	Operational
Involved technologies	• Windows Server 2016 Std ○ Description: The server hosting the Statistics Database ○ Provider: Microsoft ○ Versions: Windows Server 2016 Std ○ Infrastructure: Operations ○ Used by the services/applications: Statistics • SQL Server Database ○ Description: Database server, hosting the Statistics Database ○ Provider: Microsoft ○ Version: SQL Server 2017 (14.0) ○ Infrastructure: Operations ○ Used by the services/applications: Statistics • Statistics

	○ Description: Custom application for statistics ○ Provider: Reform ○ License: Private ○ Versions: 1.161 ○ Infrastructure: Operations

Table 23 – THPA_06_STATISTICS process in the Port of Thessaloniki

3.1.4 Posidonia Port Solutions

The deployment and usage of this modular solution for the port and maritime operations management implies the following processes related with this project pilots.

3.1.4.1 POS_01_OPE

ID	pos_01_ope
Short description	Process data of vessel events and maritime services
Process description	A Java process called AIS Parser connects to an AIS service or antenna called AIS Server to receive AIS messages in raw format (NMEA, XML, JSON, etc...). These messages are transformed into a JSON internal object by AIS Parser and published in the RabbitMQ component (messaging middleware) with a topic called AIS Centre. The messages grouped under this topic are managed, on the one hand, by the Event Detector component, a component subscribed to the RabbitMQ AIS Centre topic and, on the other hand, by a Java process called Event Dispatcher published on an XMPP (Extensible Messaging and Presence Protocol) chat server, also subscribed to this topic, which chats this information so that the Posidonia Operations client application, also developed in Java, can display the ships in a graphic interface. The event detector is composed of one or more Drools processes (Business Rules Management Systems) called CEP (Complex Event Processing) that use a system of production rules for each port/area managed by a Port Authority. Each CEP processes only the AIS messages corresponding to its port/area. From the message information, the CEP generates events that are published as JSON messages in RabbitMQ under the topic AIS Event. All these communications are done through a TCP channel. These events are managed by an integration engine, a Java application subscribed to the RabbitMQ AIS Event topic. This Java application has a defined workflow for each event. These workflows are stored in the Oracle IPOMS (Integrated Port Operations Management System) database. Among the actions that these workflows can contain there is a step consisting of registering the event in the IPOMS database through the web services (EventDB) developed in Java. In addition to this action, there may be other actions consisting of obtaining and updating information from Posidonia Management, through its web services via the HTTP channel. The managed events are: • First vessel detection in range of AIS stations

	• Crossing by point, line or zone of passage or control (*Vessel crossing waypoint or control area*)
	• Approach to port determined by entering the pilotage service area (*Port approach / departure*)
	• Start of anchoring in the anchoring area
	• End of anchoring in the anchoring area
	• Start pilotage (*Pilot stage start / end*)
	• Start towing
	• Start berthing
	• Docked vessel begins to move without leaving the dock yet
	• End of docking. There is evidence of the vessel has left the dock and the initial moment in which it began to move is maintained
	• End of pilotage (*Pilot stage start / end*)
	• End of towing
	• Port departure determined by the departure of the pilotage service area (*Port approach / departure*)
	• Vessel out of range of AIS
	• Start bunkering (*Bunkering service start / stop*)
	• End of bunkering (*Bunkering service start / stop*)
	• Anchorage vessel begins to move without leaving of anchorage area yet (*Anchorage stop start / end*)
	• Start anchoring in the restricted area of anchoring (*Vessel anchored in forbidden area*)
	• Amended vessel berth. It does not use movement thresholds greater than one length
	• AIS detection off (previously on)
	• AIS detection on (previously off)
	• Frequency of sending AIS messages too low, not complying with the regulations
	• Vessel detection in an anchorage area, in which it had previously anchored without having left it
	• Approach of a vessel to the dock (it can dock or be passing through)
	• Entry in control area
	• Control area output
	• Maximum speed exceeded
	
	These messages are reported to an Integration engine. The Integration engine processes the notified message based on a predefined workflow for each event type. This process makes calls to the Posidonia Management web services.
	
	The provided services are:
	
	• Simulador WS
	  • denegarEstadia
	  • planificarEstadia
	  • autorizarEstadia (*Visit authorization*)
	  • pasarAIniciadaEstadia
	  • updateAisAtaAtd
	  • updateVideoAtaAtd
	  • inicioFinAtraqueViaVI
	  • finalizarEstadia (*Visit start / end*)

	• iniciarEstadia (*Visit start / end*) • pasarASolicitadaEstadia • testMareaEstadia • modificarEstadia (*ETA / ETD change*) • validateTide • Operations WS • getVesselByImoOrMmsi • finishAnchorage • getVesselByMmsi • startAnchorage • getANPBerthing • getCallNumber • updateAisAtaAtd • updateVideoAtaAtd • getPendingBerthings • getRoRoShips • getVesselByImo • getDock • getCallData • startBerthing • getVesselBy • finishBerthing • getBerthingByDock • getCallNumberExtended
Department(s) involved	Systems
Type	Operational
Involved technologies	• AIS Server ○ Description: AIS ASM (*A*pplication *S*pecific *M*essages) Server from Port Authority that provide AIS messages to AIS Parser ○ Provider: Unknown for being an external product ○ License: Unknown for being an external product ○ Versions: Unknown for being an external product ○ Infrastructure: Operations ○ Used by the services/applications: AIS parser • Ubuntu Server ○ Description: Operating System for the Posidonia Operations Services ○ Provider: Canonical Ltd / Ubuntu Foundation ○ License: GPL ○ Versions: 14.04 64 bits ○ Infrastructure: Operations ○ Used by the services/applications: All processes • Oracle Database

- Description: IPOMS (Integrated Port Operations Management System) Database server. This database contains information of the vessel visit automatization management
- Provider: Oracle
- License: Privative
- Products:
  - Oracle Database XE
  - Oracle Database Standard
  - Oracle Database Edition
- Versions:
  - Oracle 10g
  - Oracle 11g
- Infrastructure: Operations
- Used by the services/applications:
  - Web Application
  - Web Services
    - Services/database operations
    - EventDB

- RabbitMQ Server
  - Description: Messaging middleware for the management of messages generated by AIS Parser and those generated by Event Detector
  - Provider: Pivotal Software
  - License: Mozilla Public License
  - Versions: 3.5.2
  - Infrastructure: Operations
  - Used by the services/applications:
    - AIS Parser
    - Event Detector
    - Integration Engine
- Apache HTTP Server
  - Description: Web/Application Server
  - Provider: Apache Software Foundation
  - License: GPL
  - Versions: 2
  - Infrastructure: Operations
  - Used by the services/applications:
    - Client Application
- Jetty Server
  - Description: Web/Application server for hosting the web services responsible for managing the AIS messages in the database
  - Provider: Fundación Eclipse
  - License: Apache License 2.0, Eclipse Public License 1.0

	○ Versions: Not specified ○ Infrastructure: Operations ○ Used by the services/applications: ▪ Web Services (EventDB) • Apache Tomcat ○ Description: Web server for hosting the Posidonia Operations application and web services ○ Provider: Apache Software Foundation ○ License: Apache License 2.0 ○ Versions: 7.67 ○ Infrastructure: Operations ○ Used by the services/applications: ▪ Web Application ▪ Web Services (services / database operations) • OpenFire ○ Description: Messaging server for the management of AIS messages managed by AIS Parser ○ Provider: Jive Software ○ License: Apache License 2.0 ○ Versions: 3.10.3 ○ Infrastructure: Operations ○ Used by the services/applications: ▪ Event dispatcher

Table 24 – POS_01_OPE process in Posidonia Port Solutions

3.1.4.2 POS_02_MNG

ID	pos_02_mng
Short description	Process incoming EDI messages
Process description	The shipping agents or ship consignees send the documentation relating to a call to the Port Authority via the latter's EDI station in EDIFACT[37] format so that this station can transform them into XML format, or via XML message sent to Posidonia EDI's MensajesEDI web services so that these can be deposited at the EDI Station. The XML messages of the EDI station are sent to the web services EDICargaMensajesws of Posidonia Management. These services register the messages received in the Oracle database of Posidonia Management. The messages recorded in the database are processed, from time to time, by Posidonia EDI batch processes (scale and manifest). These processes

[37] EDIFACT (*Electronic Data Interchange for Administration Commerce and Transport*) is a global standard approved by the Unit Nations defining a set of data structured according to agreed message standards for transmission by electronic means defined in ISO 9735.

	syntactically and semantically validate the messages and execute business rules defined in files or in the database. As a result of executing these rules a response message must be issued, the messages are sent to the web services MensajeEDI of Posidonia EDI so that these services forward it to the EDI station and this in turn to the message recipient. As an example of this flow, some processes related to port management are described below. For example, in the case of the beginning of a call, the ship's agent or consignee sends a BERMAN[38] message to the Port Authority. This message can be sent directly to the Port Authority's EDI station in EDIFACT format so that this station can translate it into an XML format, or indirectly to the EDI station by sending an XML message to the Posidonia EDI messages web services, to be deposited later at the EDI station. The XML message is sent by the EDI station to the web services EDICargaMensajeWS of Posidonia Management. These services are in charge of registering the message in the Oracle database of Posidonia Management, in addition to other functions such as the denormalization of the information received in order to register it in the aforementioned database. Once the messages have been registered in the Oracle database, the EDI Escala batch process of Posidonia EDI, from time to time, processes the messages, validates them syntactically and semantically, and executes the business rules corresponding to the type and content of the messages, among which it generates an APERAK response message of acceptance (with scale number) or rejection of the scale request in XML format. This reply message is sent to the MensajeEDI web services of Posidonia EDI to be forwarded to the EDI station which in turn forwards it to the recipient. Once the port of call has been accepted, the Directorate General of the Merchant Navy authorises/discourages the request for a port of call (docking)[39] via the web application. As a result, an APERAK message is generated and sent to Shipping Agent. This message follows the same process described above, i.e.; it is sent to the web services of MensajesEDI of the Posidonia EDI so that it can be sent back to the EDI station which in turn sends it to the recipient. After the first sending of a call request, the Ship's Agent or Consignee can send new BERMAN messages to request a cancellation, replacement, modification and request for new berths/anchorages, as well as new services and operations related to the previously sent message. After the call is accepted, the Ship's Agent or Consignee sends an IFCSUM message corresponding to a Cargo Summary Declaration/Manifest

[38] The EDIFAC BERMAN message is also used for:

- The registration of new berths/anchors and of services and operations, replacements, modifications and cancellations
- Communication by the Port Authority of the ATA/ATD (Actual time of arrival/Actual time of departure)
- Message sent by the Port Authority to communicate the change of authorised consignee for a call

[39] When a request for a call includes several berths, there is an APERAK for the assignment of the call number and the same number of APERAK messages for each berth authorization.

	(Information regarding a ship and the cargo carried to be unloaded/loaded) to the Port Authority following the shipping procedures described above. In the case of the Summary Declaration of unloading, this is activated when the Port Authority sends a CUSREP message to Customs, called "Pilot Notice Notification". The procedure for sending this message follows the dispatch flow described above. The Customs will accept or reject the Summary Declaration and will communicate by sending a CUSRES message to the Port Authority's EDI Station, the result of its management, and this in turn, the result to Consignee by forwarding the message to the Posidonia EDI message web services, and these in turn, the forwarding to the EDI station for the later to forward it to the consignee. In the case of a Cargo Manifest, this is activated when the Port Authority sends a CUSCAR message to Customs. As with unloading summary declarations, Customs shall inform the Port Authority the acceptance or rejection of the manifest by sending a CUSRES message to the Port Authority's EDI station, which in turn sends the result of the Shipping Agent by resending the message. In the case of Cargo Manifest, the Batch EDI manifest processes are responsible for processing the corresponding messages. Once all operations on the ship have been carried out, the Port Authority, through the web application, sends a BERMAN message with the actual date of departure to communicate the end of the call. The Director General Merchant Navy is the one that authorises the departure clearance for a call request by means of an APERAK message.
Department(s) involved	Systems
Type	Operational

Involved technologies	• Oracle Database Server
	  o Description: Database server deployed on Port Authority with different schemas like Posidonia EDI and Posidonia Management. It registers the messages information.
	  o Provider: Oracle
	  o License: Privative
	  o Products:
	  ▪ Oracle Database XE
	  ▪ Oracle Database Standard
	  ▪ Oracle Database Edition
	  o Versions:
	  ▪ Oracle 11g
	  o Infrastructure: Management
	  o Used by the services/applications:
	  ▪ Web Application Posidonia
	  ▪ Web Services
	  • EDICargaMensajeWS

	• Batch EDI Escala • Batch EDI Manifiestos • Batch Invoicing • JBoss Application Server o Description: Java EE Application server used for the deployed of the Posidonia Management web application and EDICargamensajes web services o Provider: Red Hat o License: LGPL o Versions: JBoss 4.2.3 o Infrastructure: Management o Used by the services/applications: ▪ Posidonia Web Application ▪ Posidonia EDICargaMensajesWS web services • Ubuntu Server o Description: Operating System for the Posidonia Management. The EDI batches responsible of processing EDI messages are displayed in this system o Provider: Canonical Ltd / Ubuntu Foundation o License: GPL o Versions: 14.04 64 Bits o Infrastructure: Management o Used by the services/applications: All services and applications • Apache HTTP Server o Description: HTTP/HTML server by deployed Posidonia Management o Provider: Apache Software Foundation o License: GPL o Versions: 2 o Infrastructure: Management o Used by the services/applications: ▪ Client Application • Apache Tomcat o Description: Application server Java EE for deployed of Posidonia EDI web services o Provider: Apache Software Foundation o License: Apache License 2.0 o Versions: 7.67 o Infrastructure: EDI o Used by the services/applications: ▪ Web Services (Posidoniaedi)

Table 25 – POS_02_MNG process in Posidonia Port Solutions

3.2 BENCHMARKING AND COMPARISON

3.2.1 Thessaloniki Port

The technologies used in Thessaloniki Port Authority can be categorised into two main categories: a) purchased by a provider and b) developed internally. In the first case, the technologies used are evaluated following the internal evaluation process of the organisation for selecting a product. In the second case, it is a decision of the IT department considering the capabilities of the personnel to be involved, integration requirements with other systems in place but also the technological trends. Therefore, the analysis below will be used to compare and evaluate existing systems in terms of operating system, databases and web/application servers, providing ThPA point of view in this activity.

Where applicable, a statement regarding the selection of ThPA is provided.

3.2.2 Valencia Port

The technologies and systems described in this document for the port of Valencia are owned by different companies, so each of them carried out their own benchmarking analysis. In the last years, the Port Authority used mainly Microsoft products, so most of the software components included in their systems are from Microsoft and related technologies such as .Net.

For other solutions, like the industrial port platform, the IT department carried out an analysis of the solutions in the market to select the product that better fits to the Port Authority needs.

3.2.3 Posidonia Port Solutions

The technologies used in the different products that make up Posidonia are determined, on the one hand, by the needs of the client for whom these products were developed and, on the other hand, by the type of technology used in most Spanish ports. Within the first aspect the Java programming language is included and, consequently, the Linux operating system. Regarding the second aspect, Oracle is included, as it is the most used database manager by Port Authorities in the Spanish territory.

Regardless of the reasons why the technologies described in Posidonia products are being used, a comparative analysis of these technologies with those that can be considered as its main competitors has been carried out. In this analysis, each of the technologies has been evaluated from different aspects and scored according to the advantages offered.

Where applicable, a statement regarding the selection of Prodevelop is provided.

3.2.4 Common Technologies and Ad Hoc Scoring

In this section are compared and scored some of the technologies present in the different pilot scenarios, that currently provider the required functionalities for performing the already described processes.

3.2.4.1 Operating System

A comparison is made with the main operating systems for servers on the market. For each aspect, a score has been established according to the added value.

3.2.4.1.1 Cost

This aspect analyses the acquisition cost of the system and whether it is independent of the hardware. The lower the cost, the higher the score. The fact that the operating system is included in the hardware has been penalized, by making the system dependent on the hardware.

	Free	Paid	Included in hardware	Total
	3	2	1	
AIX	○	○	●	1
Linux	●	○	○	3
Windows Server	○	●	○	2

Table 26 – OS Costs

3.2.4.1.2 License

This table describes the type of license of the analysed operating systems. The score is determined by the copyright degree of the license.

	GPL	BSD	MPL	Total
	3	2	1	
AIX	○	○	●	1
Linux	●	○	○	3
Windows Server	○	○	●	1

Table 27 – OS Licenses

3.2.4.1.3 Supported Architectures

This table describes the types of architecture supported by the analysed operating systems. The established score is determined by the number of general-purpose registers, the clock frequency and the consumption.

	x86-64	PowerPC	x86	Total
	3	2	1	
AIX	○	●	○	2
Linux	●	●	●	6
Windows Server	●	○	●	4

Table 28 – OS Supported Architectures

From the results of the different comparative analysis, corresponding to the main server-oriented operating systems on the market, it can be deduced that Linux is the operating system that offers the best advantages, followed by Windows Server and, finally, AIX.

Linux is used by all components of the architecture.

3.2.4.1.4 Posidonia Port Solutions

The operating system used in Posidonia products is Linux. As mentioned above, the use of this system has been determined by the requirements imposed by the customer for whom the first Posidonia solution was

developed. Linux is mainly characterized by its robustness, efficiency, stability and cost, since it is free. On the other hand, it is the majority option used in web servers, internet services, data centres, etc.

3.2.4.1.5 Port of Thessaloniki

The operating system used in Posidonia products is Linux. As mentioned above, the use of this system has been determined by the requirements imposed by the customer for whom the first Posidonia solution was developed. Linux is mainly characterized by its robustness, efficiency, stability, and cost, since it is free. On the other hand, it is the majority option used in web servers, internet services, data centres, etc.

3.2.4.2 Databases

The following tables make a comparison of the main database managers on the market. Each feature reviewed has been rated based on the benefits it offers.

3.2.4.2.1 License

This table analyses the main licenses of the selected databases. The scoring level is determined by the copyright degree of the license.

	GPL	BSD	MPL	Total
	3	2	1	
DB2	○	○	●	1
MySQL	●	○	●	4
Oracle	○	○	●	1
PostgreSQL	○	●	○	2
SQL Server	○	○	●	1

Table 29 – DB Licenses

3.2.4.2.2 Supported Operating System

This table describes the operating systems compatibility with the different databases analysed. The higher the number of supported operating systems, the higher the score.

	Linux	Mac	Windows	Total
	1	1	1	
DB2	●	●	●	3
MySQL	●	●	●	3
Oracle	●	○	●	2
PostgreSQL	●	●	●	3
SQL Server	○	○	●	1

Table 30 – DB Supported OS

3.2.4.2.3 Security

This table describes the different levels of security offered by the analysed databases. The higher the security offered, the higher the score.

	User groups / Roles	Row Level security	Grant on column level	Total
	1	1	1	
DB2	●	●	●	3
MySQL	●	○	●	2
Oracle	●	●	●	3
PostgreSQL	●	●	●	3
SQL Server	●	●	●	3

Table 31 – DB Security

3.2.4.2.4 Programming Language

This table describes the programming languages that are supported by the analysed databases. The greater the number of supported languages, the higher the score.

	PL/SQL	PHP	Java	.NET	XML	T-SQL	C/C++	Pascal	PL/PgSQL	Total
	1	1	1	1	1	1	1	1	1	
DB2	○	●	●	●	○	○	●	○	○	4
MySQL	○	●	○	○	○	○	●	●	○	3
Oracle	●	●	●	●	●	○	○	○	○	5
PostgreSQL	○	○	●	○	○	○	●	○	●	3
SQL Server	○	○	○	○	○	●	○	○	○	1

Table 32 – DB Programming Languages

3.2.4.2.5 Storage

This table describes the size limitation of the analysed databases. The score is determined based on whether the size is limited or not.

	Unlimited	Limited	Total
	2	1	
DB2	○	●	1

MySQL	●	○	2
Oracle	●	○	2
PostgreSQL	●	○	2
SQL Server	○	●	1

Table 33 – DB Storage

3.2.4.2.6 Posidonia Port Solutions

The type of database used in Posidonia is Oracle. As mentioned above, this type of database is being used due to the requirements set by the customer for whom the first version of Posidonia was developed.

In the technical list of this technology, several products of this type of database have been defined, as well as versions, compatible with Posidonia. The choice is determined based on the customer's needs.

Once the different characteristics have been analysed, it is obtained that the best valued database is MySQL, followed by Oracle and PostgreSQL with the same score, then DB2 and finally SQL Server. However, today Oracle is still the type of database most used by the Port Authorities of the Spanish territory.

Oracle database is used in the following Posidonia Operations components:

- Web Services EventB
- Web Services (services/ipoms)
- Web Application.

3.2.4.2.7 Port of Thessaloniki

The previous comparison and conclusions are in line with ThPA technologies.

3.2.4.3 Queueing Systems

Unlike the technologies discussed so far, when it comes to queue management systems, RabbitMQ was selected internally for its advantages over other technologies on the market. Proof of this is the result of the following comparative analysis with the main queue management systems on the market.

RabbitMQ is open-source message management software that works as messaging middleware. This type of software stores received messages in exchange queues, so that clients can retrieve them later.

Next, different aspects of this type of software are analysed and compared with the main products on the market today.

3.2.4.3.1 License

This table describes the main licenses used by the selected queue management systems. The scoring level is determined by the copyright degree of the license.

	Mozilla	Apache 2.0	Total
	2	1	
Flume	○	●	1
Kafka	○	●	1

RabbitMQ	●	○	2

Table 34 – Queueing Systems' Licenses

3.2.4.3.2 Volume

This table describes the volume of events per second that the analysed systems are capable of handling. The higher the number of events, the higher your score.

	100k events/seg	20 events/seg	Total
	2	1	
Flume	○	●	1
Kafka	●	○	2
RabbitMQ	○	●	1

Table 35 – Queueing Systems' Volume

3.2.4.3.3 Federated Queues

This table describes whether the analysed systems allow queue federation.

	Yes	No	Total
	1	0	
Flume	○	●	0
Kafka	○	●	0
RabbitMQ	●	○	1

Table 36 – Queueing Systems' Federated Queues

3.2.4.3.4 Complex Routing

This table describes whether the selected systems provide complex routing.

	Yes	No	Total
	1	0	
Flume	○	●	0
Kafka	○	●	0
RabbitMQ	●	○	1

Table 37 – Queueing Systems' Complex Routing

3.2.4.3.5 Event Replication

This table describes whether the selected systems allow event replication.

	Yes	No	Total
	1	0	
Flume	○	●	0
Kafka	●	○	1
RabbitMQ	●	○	1

Table 38 – Queueing Systems' Event Replication

The analysis carried out shows that the highest rated queue management system is RabbitMQ, followed by Kafka and finally Flume.

3.2.4.3.6 Posidonia Port Solutions

RabbitMQ is used in the following Posidonia Operations components:

- AIS Parser
- Integration Engine
- Event Detector.

3.2.4.4 Web Application Servers

Apache, the technology used in Posidonia is due to the requirements of the client for whom the first version of this product was developed. The Apache HTTP server is an open-source server compatible with the Unix, Microsoft Windows and Macintosh platforms, among the most important. It is characterized by being modular, multiplatform and extensible. Until recently it was the most used server by the world's websites, however, in recent years it has been losing market share to the detriment of other servers.

Below is a comparative analysis of the leading web servers on the market today. For each characteristic analysed, a score is established that is determined by the advantages it offers or the number of compatible elements.

3.2.4.4.1 Function

This feature lists and evaluates the different functions offered by the analysed web servers. The greater the number of functions, the higher the score.

	Web Server	Proxy Sever	Email Server	Load balancer	Total
	1	1	1	1	
Apache	●	●	○	○	2
IIS	●	●	●	●	4
Ngnix	●	●	●	●	4

Table 39 – Web Application Servers' Functions

3.2.4.4.2 Operating System

This table describes and evaluates the different operating systems supported by the analysed web servers. The greater the number of compatible systems, the higher the score obtained.

	Unix	FreeBSD	Linux	Solaris	AIX	UX	OS	Windows	Total
	1	1	1	1	1	1	1	1	
Apache	●	○	○	○	○	○	○	●	2
IIS	○	○	○	○	○	○	○	●	1
Ngnix	○	●	●	●	●	●	●	●	7

Table 40 – Web Application Servers' OS

3.2.4.4.3 License

This table describes the main licenses used by the selected web servers. The scoring level is determined by the copyright degree of the license.

	Apache v2.0	BSD	Owner	Total
	3	2	1	
Apache	●	○		3
IIS	○	○	●	1
Ngnix	○	●		2

Table 41 – Web Application Servers' Licenses

3.2.4.4.4 Software Architecture

This table describes how the workload is managed on the different selected web servers. The scoring level is determined by the number of processes managed.

	Sub-processes/events	Processes/threads	Total
	2	1	
Apache	●	●	3
IIS	○	●	1
Ngnix	●	○	2

Table 42 – Web Application Servers' Architectures

3.2.4.4.5 Contents

This table describes the different content that the selected web servers can offer. The score is determined by the amount of content offered.

	Static web	Dynamic web	Total
	1	1	

Apache	●	●	2
IIS	●	●	2
Ngnix	●	○	1

Table 43 – Web Application Servers' Contents

3.2.4.4.6 Configuration

This table describes the types of settings available for the selected web servers. The score is determined by the number of settings allowed.

	Centralized	Decentralized	Total
	1	1	
Apache	●	●	2
IIS	●	○	1
Ngnix	●	○	1

Table 44 – Web Application Servers' Configurations

The result of the comparative analysis carried out shows that Apache has been displaced from the first position by Ngnix. This type of server is currently used by many websites such as Netflix, WordPress, Facebook, GitHub. However, Apache is still one of the most widely used web servers today.

3.2.4.4.7 Posidonia Port Solutions

Apache serves is used in the following Posidonia Operations components:

- Client Application.
- Web Application.
- Web Services services/ipoms.
- Web Services EventDB.

3.2.4.4.8 Port of Thessaloniki

All current web applications inside ThPA were developed long before Nginx started getting momentum in the global web server market share, or even toppled the traditionally dominant apache.

3.2.4.5 Development Language

Like most of the technologies analysed above, the development language used in Posidonia is due to the requirements of the client for whom the first version of this product was developed.

Java is one of the most popular programming languages that is fast, secure and reliable. It is used by more than 9 million developers worldwide. Java is the global standard for developing and distributing mobile and embedded applications, enterprise software, web content, and games.

The following is a comparison of this programming language with some of the most widely used web development languages today.

3.2.4.5.1 Execution of the Source Code

This feature evaluates how the source code is executed. In compiled languages, the source code must be

compiled (converted to machine language) before it is executed. In interpreted languages, the source code is compiled at runtime. Both types have their advantages and disadvantages so your choice will be determined by different factors.

	Compiled	Interpreted	Total
	1	1	
C# (*)	●	●	2
Java (*)	●	●	2
JavaScript	○	●	1
PHP	○	●	1
Python	○	●	1

Table 45 – Development Languages' Source Codes

(*) The source code is converted into an intermediate code (virtual machine) to be later interpreted and converted into the machine language

3.2.4.5.2 Type System

The type system used in a programming language affects its execution speed, as well as the number of execution errors. On the one hand, in the static typing the type checking is done during the compilation, while in the dynamic typing it is done during the execution, which implies a cost. On the other hand, checking the types during compilation ensures that the program does not have typing errors at runtime, not so in the dynamic typing. Taking into account these aspects, the following evaluation has been established.

	Static	Dynamic	Total
	2	1	
C#	●	○	2
Java	●	○	2
JavaScript	○	●	1
PHP	○	●	1
Python	○	●	1

Table 46 – Development Languages' Type Systems

3.2.4.5.3 Typing Fortress

As with the previous feature, this feature also has an impact on the speed of execution and the number of errors. In the strong typed every variable must have a type when it is declared, while in the weak typed this requirement is not necessary, which may lead to errors at runtime to be able to assign a value to a variable whose type differs from the type previously assigned. On the other hand, with the strong typing, a higher execution speed is ensured because the type inference is done at compile time. Taking into account these aspects has established the following assessment.

	Strong	Weak	Total
	2	1	
C#	●	○	2
Java	●	○	2
JavaScript	○	●	1
PHP	○	●	2
Python	●	○	2

Table 47 – Development Languages' Typing Fortresses

3.2.4.5.4 Learning Curve

This feature evaluates the time it takes to learn a programming language. The higher the curve, the less time has to be invested in learning the language, time that is reflected in productivity.

	High	Medium	Low	Total
	3	2	1	
C#	○	○	●	1
Java	○	●	○	2
JavaScript	○	●	○	2
PHP	●	○	○	3
Python	●	○	○	3

Table 48 – Development Languages Learning Curves

From the analysis carried out, it is clear that Java is the most valued programming language, followed by C#, PHP and Python, closing the list with JavaScript. However, Java has been losing ground in the last decade in favour of JavaScript and Python. An example of current technology trends can be found in the 2020 developer Survey[40].

3.2.4.5.5 Posidonia Port Solutions

Java is used in the following Posidonia Operations components:

- Client Application
- Web Application
- Web Services (services/ipoms)
- AIS Parser
- Integration Engine
- Event Detector

[40] https://insights.stackoverflow.com/survey/2020#most-popular-technologies

4 REQUIREMENTS ELICITATION

4.1 INTRODUCTION

Requirement's analysis is a key step in software design and engineering, as it determines the conditions that need to be met by the system to be developed. In agile software development processes as followed by DataPorts, the list of requirements can change, in accordance with the evolvement of the architecture design and specification. To this end, since the beginning of the project, all partners were requested to provide requirements based on internal technical analysis, end-users' expectations, experience from previous projects etc. This elicitation process resulted in a very broad set of initial requirements. As the project vision became more and more concrete, this set was being constantly updated. In addition, given the first stable version of the DataPorts architecture, test cases / acceptance criteria were specified for each one requirement to evaluate their completion towards the end of the project. Following many revisions, the final tables of the requirements are presented in this deliverable. The document collects requirements using the MoSCoW method [29], to prioritise the features needed from the DataPorts Platform.

This section is structured as follows: the table format for the requirements gathering is presented in subsection 4.2, subsection 4.3 includes the analysis of the requirements, and finally subsection 4.4 contains a glossary explaining the meaning of some terms mentioned in the requirements description.

All the technical requirements derived from WPs 2-4 along with the technical and business ones from WP5 are listed in Annex 1 (Section 7). Nevertheless, impact and business requirements are included and analysed in the D6.3 deliverable followed by D6.8 and D6.9 as its corresponding iterations.

4.2 TERMINOLOGY

To have a homogeneous reading of the project's requirements, each one of them is specified by the contents of the following fixed format table:

Requirements Fields	
ID	For WPs 2-5: **WP number.counter** (e.g. the first requirement of WP2 is 2.1 etc.)
Type	This field determines whether the requirement is **Functional** or **Non-Functional** [30] [31]: • **Functional** requirements are the ones that specify a function that a system or system component must be able to perform. These requirements should be action oriented and should describe the tasks or activities that the system performs during its operation • **Non-Functional** requirements provide a description of a property or characteristic that a software system must exhibit or a constraint it must respect, other than an observable system behaviour. In a nutshell, they describe not what a software will do, but how the software will do it. These requirements include several aspects, such as development constraints, business rules, external interfaces, and quality attributes

Category	This field illustrates the category(ies) of the requirement. The following list presents indicative categories, particularly for the non-functional requirements [32] [30] [33] [31]: • **Interoperability**: ability of two or more systems or components to exchange information and to use the information that has been exchanged. Interoperability can be further specified as technical, syntactic, and semantic • **Compatibility**: degree to which the solution operates effectively with other components in its environment, such as one process with another • **Functionality**: degree to which the solution functions meet user needs, including aspects of suitability, accuracy, and interoperability • **Maintainability**: ease with which a solution or component can be modified to correct faults, improve performance or other attributes, restore system to a state in which it can perform its required functions or adapt to a changed environment • **Performance Efficiency**: degree to which a solution or component performs its designated functions with minimum consumption of resources. Can be defined based on the context or period, such as high-peak, mid-peak or off-peak usage • **Portability**: ease with which a solution or component can be transferred from one environment to another concerning both the aspects of transportation and adaptation • **Scalability**: degree with which a solution is able to grow or evolve to handle increased amounts of work • **Security**: degree of protection of information and data so that unauthorized persons or systems cannot read or modify them and authorized persons or systems are not denied access to them. Properties of security are also considered Authentication, Integrity, Non-repudiation, Confidentiality, Availability, Accountability and Authorization • **Privacy**: privacy can be distinguished as hard privacy which refers to the minimization of data that is shared with third parties, and soft privacy that refers to data security in terms of processing data with specific purpose and consent by means of policies, access control and audit • **Usability**: ease with which a user can use the solution, including the effort required to learn, operate, prepare input, and interpret the output of the system • **Certification**: constraints on the solution that are necessary to meet certain standards or industry conventions • **Compliance**: regulatory, financial, or legal constraints, which can vary, based on the context or jurisdiction

Requirements Fields	
	• **Service Level Agreements**: constraints of the organization being served by the solution that are formally agreed to by both the provider and the user of the solution • **Extensibility**: ability of a solution to incorporate new functionality or storage capacity without loss of functionalities or requirements • **Verifiability:** ease of a solution to verify its performance • **…**
Source	This field identifies the **source** of the requirement (e.g., Internal Technical Analysts, End-Users, Experience from previous projects, etc.)
Priority	Based on the **MoSCoW** method [29]: • **MUST:** Describes a requirement that must be satisfied in the final solution for the solution to be considered a success • **SHOULD:** Represents a high-priority item that should be included in the solution if it is possible. Tis is often a critical requirement but one which can be satisfied in other ways if strictly necessary • **COULD:** Describes a requirement which is considered desirable but not necessary. Tis will be included if time and resources permit • **WON'T:** Represents a requirement that stakeholders have agreed will not be implemented in a given release, but may be considered for the future
Description	This field contains the specification of the requirement and it must be aligned with the priority label that is defined in the previous field. Moreover, it will be used to determine the completion of the requirement. A way that could be used is the **user story syntax**: As a <*type of user*>, I want <*some goal*>, so that <*some reason*> That syntax represents a small, concise statement of functionality or quality needed to deliver value to a specific stakeholder
Rationale	This field describes the need that the specific requirement is covering. It also shows the background reason for the requirement, which removes much of the ambiguity
Responsible Task(s)	This field indicates the specific **Task(s)** that the requirement belongs to
Additional Comments (optional)	Extra comments that could be used in order to further describe the specific requirement

Requirements Fields	
Test Case / Acceptance Criteria	This field describes the way to test the requirement, to ensure that it is developed and thus, fulfilled. This information will be the base for the verification & validation procedures A way that could be used is the **Given-When-Then** formula

Table 49 – Requirements Fields

4.3 ANALYSIS

In this subsection, the requirements of the project are presented on a higher level and the corresponding statistics define the overall picture of the results that are expected to be achieved.

Firstly, most of the project's requirements have been marked as mandatory (MUST) (Figure 9), while it is suggested that a 20% of them should be fulfilled (SHOULD). Lastly, a 9% of the requirements are proposed (COULD) but would not affect the project's outcomes and achievements if not fulfilled.

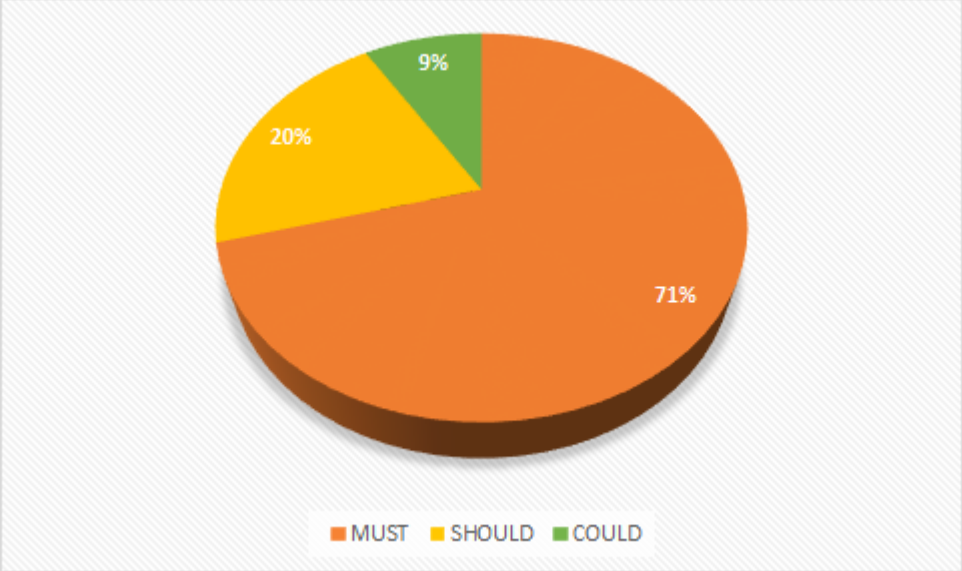


Figure 9 – Requirements' priorities

According to the terminology described in subsection 4.2, functional requirements depict the basic facilities that the platform should offer to its users, while non-functional requirements are quality attributes and constraints that it should comply to. Figure 10 shows that there is a balance between functional and non-functional requirements that have been defined in the context of the DataPorts project.

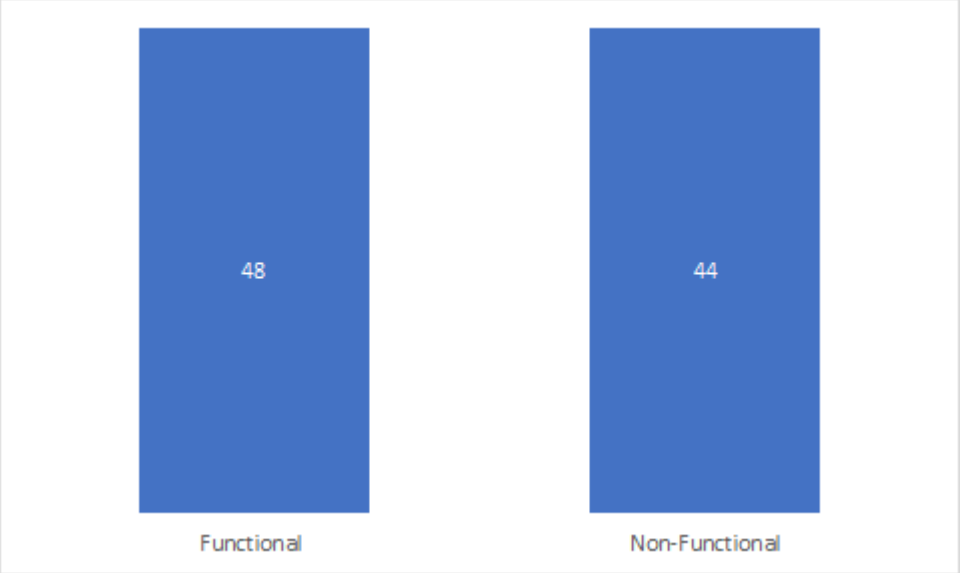


Figure 10 – Requirements' types

However, on WP level (Figure 11), only the requirements of WP3 are functional in their majority. This fact is expected since this WP addresses the implementation and functionalities of the DataPorts Platform. On the other hand, most of the requirements in WP2 and WP4 are non-functional. Indeed, WP2 requirements mainly describe the constraints and properties of the DataPorts architecture design, while WP4 deals with significant aspects of the DataPorts Platform, such as security and privacy. As far as WP5 is concerned, most of the requirements are business and describe the expectations of the end-users in terms of both the functionalities and the characteristics of the platform. Therefore, a balance between functional and non-functional requirements is anticipated.

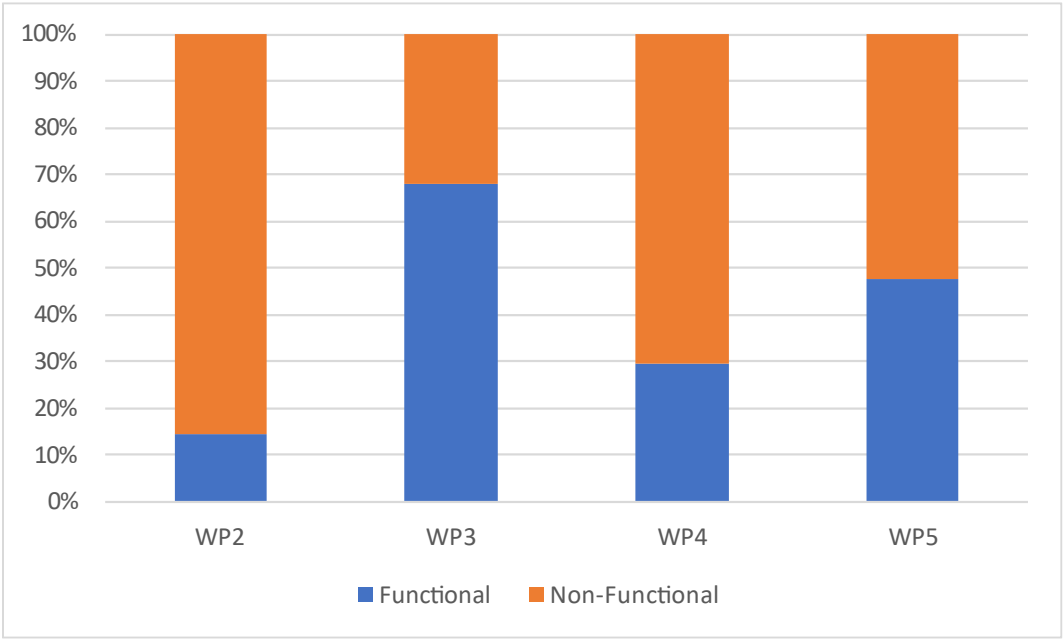


Figure 11 – Requirements' types per WP

Regarding the categories (Figure 12) in which the requirements belong, the most frequent are interoperability, functionality, security, and privacy. In fact, the DataPorts Platform is expected to function as a semantic interoperability framework with the introduction of a new global ontology that is going to

improve collaboration and common data representation among ports. Furthermore, one of the project's objectives is to offer to the involved stakeholders a trusted and secure environment to share their data, based on the Blockchain technology along with many supported functionalities. Consequently, the most frequent categories are well justified.

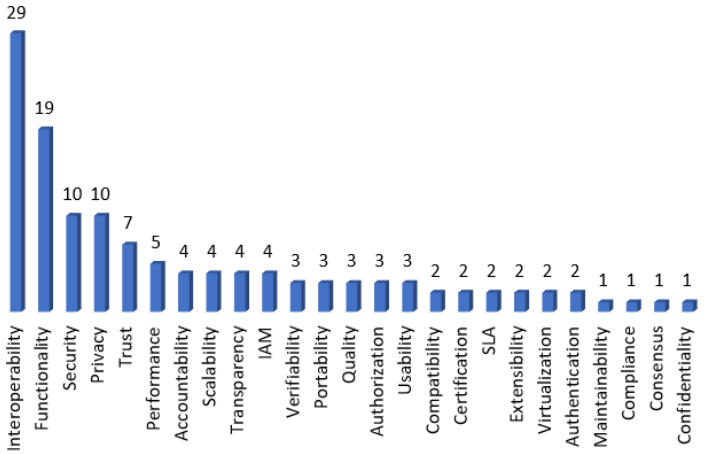


Figure 12 – Requirements per category

Lastly, each source group focused on different aspects of the DataPorts Platform and provided the respective requirements. More specifically, academic partners put special emphasis on interoperability, performance, scalability, and quality features of the platform. For instance, they underline -with the defined requirements- the necessity of interoperability regarding the data sources that are going to be integrated into the platform and provide a uniform API for data access. The pilots -as users of the platform-, apart from indicating several important functionalities based on their needs, they have also specified requirements concerning the platform's security, privacy, accountability, authorization, and authentication mechanisms. Industry partners on the other hand focused mainly on interoperability, privacy, trust, and identity access management requirements.

The complete list of the requirements separated per WP can be found in ANNEX 1: REQUIREMENTS AT WP LEVEL.

4.4 GLOSSARY

Term	Description
Data Provider	An organization that produces data or metadata
Data Consumer	An organization that uses data as input for further processing
Data Prosumer	An organization that produces and uses data or services
Blockchain Organization	An organization participating in a Blockchain network. This organization usually (but not necessarily) holds a copy of the ledger and run business logic (chaincode) and might participate in consensus on the network (decisions on ordering of the blocks of transactions on the ledger). Organizations can issue identities to their users so that if the users submit transactions on Blockchain, every transaction's source is clear and identifiable. Organizations share their data on the ledger in authorized manner. For each pilot a Blockchain network, including a list of participating organizations is included in the network design

Term	Description
Blockchain channel	On the same Blockchain network infrastructure, multiple channels can be configured and run. Channel is defacto a separate ledger. Therefore, if in the same network some participants want to share data in a subset of this network only, channels can be used for this purpose
Endorsing transactions	Running business logic on organizational peer to validate and "agree upon" transaction result with other peers in the network
Smart contracts/chaincode	The business logic running on peers (Blockchain node) to validate and agree upon transaction's results with other peers, and committing those transactions to the Blockchain ledger
End-User	A person or organization that uses a platform/application/program rather than the person or organization that developed it
System Administrator	A person who is responsible for managing and maintaining computer systems, monitors the implementation of the established information assurance policy and provides efficient utilization of the system resources, taking also into account security aspects
Port Community System (PCS)	An open and neutral electronic platform that allows safe and smart information exchange between public and private agents, in order to improve the competitive position of a Port Community
Container	Metal box of internationally standardized types and dimensions as well as equipped with hooks or rings to facilitate loading and unloading by cranes, which is used to transport goods over long distances
Terminal	An assigned area in which containers are managed; e.g. they are being prepared for loading into a vessel, truck, or train
Terminal Operator	A company that operates container handling activities on a terminal
Truck	Heavy automotive vehicle used to transport containers

Table 50 – Requirements Glossary

5 CONCLUSIONS

This deliverable lays the foundations that the DataPorts Platform is going to be constructed upon. It includes a thorough analysis and investigation of the current technological reality as far as the ports ecosystem and maritime applications are concerned. The analysis is performed on multiple levels and includes the review of the respective scientific literature, the contemporary technologies, and commercial products, as well as the related EU research projects that could contribute towards a potential collaboration. Furthermore, in the present document is included the documentation of the technologies, processes, and tools currently used by the ports of Valencia and Thessaloniki, as well as the benchmarking of these technologies. This survey creates the baseline technological framework that the DataPorts Platform is going to enhance and will enable the documentation and review of the results and improvements that are achieved through the outcomes of this project.

Resulting from the State-of-the-art analysis in combination with the recording of the already existing port practices and used technologies, have emerged the requirements of the DataPorts project. After the establishment of the requirement elicitation methodology, the consortium defined an adequate number of requirements that address different aspects of the DataPorts Platform, along with their test case and acceptance criteria that will be used to verify and validate their fulfilment. These requirements outline on a higher level the platform's functionality and are going to serve as input to multiple subsequent tasks and particularly to the architecture design. The requirements analysis also depicts the distribution of the implementation of functionalities at each technical task of the project, as well as the additional needs that are imposed by the consortium on an industry, end-user, and academic level. Overall, most of the requirements are considered mandatory for the DataPorts project to be considered successful and address primarily the interoperability, functionality, security, and privacy attributes of the DataPorts Platform.

In conclusion, this deliverable is expected to play a detrimental role not only in the development of the DataPorts Platform and the establishment of its position amongst State-of-the-Art and research solutions, but also in the evaluation and benchmarking against the current tools and processes of the end-users in the ports of Valencia and Thessaloniki.

6 REFERENCES AND ACRONYMS

6.1 REFERENCES

[1] D. Z. L. C. a. G. A. L. M. Millefiori, “A distributed approach to estimating sea port operational regions from lots of AIS data,” in *IEEE International Conference on Big Data (Big Data)*, Washington, DC, 2016.

[2] A. J. N. M. A. S. S. J. B. J. M. Rabe, “A decision support tool for port planning based on monte carlo simulation,” in *Winter Simulation Conference (WSC)*, 2018.

[3] M. V. F. O. a. J. A. I. Lytra, “A big data architecture for managing oceans of data and maritime applications,” in *International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, 2017.

[4] “The five V's of big data,” [Online]. Available: https://www.bbva.com/en/five-vs-big-data/. [Accessed 9 11 2020].

[5] M. F. C. M. a. E. A. S. K. Douaioui, “Smart port: Design and perspectives,” in *4th International Conference on Logistics Operations Management (GOL)*, Le Havre, 2018.

[6] P. &. P. M. Anitha, “A Review on Data Analytics for Supply Chain Management: A Case study,” in *International Journal of Information Engineering and Electronic Business*.

[7] R. &. H. P. Addo-Tenkorang, “Data Applications in Operations/Supply-Chain Management: A Literature Review. Computers & Industrial Engineering,” *Computers & Industrial Engineering,* vol. 101, 2016.

[8] S. Mohan, “Big Data: Transforming Logistics and Supply Chain,” *International Journal of Pure and Applied Mathematics,* vol. 117, no. 20, pp. 911-916, 2017.

[9] E. Hofmann, “Big data and supply chain decisions: the impact of volume, variety and velocity properties on the bullwhip effect,” *International Journal of Production Research,* vol. 55, pp. 1-19, 2015.

[10] R. a. H. P. Addo-Tenkorang, “Big Data Applications in Operations/Supply-Chain Management: A Literature Review,” *Computers & Industrial Engineering,* vol. 101, 09 2016.

[11] T. M. Systems, “Use of Big Data in the Maritime Industry Report,” Port technology International.

[12] J. S. L. L. Çağatay Iris, “A review of energy efficiency in ports: Operational strategies, technologies and energy management systems,” *Renewable and Sustainable Energy Reviews,* vol. 112, pp. 170-182, 2019.

[13] L. S. Hannes Johnsona, “Increased energy efficiency in short sea shipping through decreased time in port,” *Transportation Research Part A: Policy and Practice,* vol. 71, pp. 167-178, January 2015.

[14] C.-C. a. J. C. Chang, “Reducing speed and fuel transfer of the Green Flag Incentive Program in Kaohsiung Port Taiwan,” *Transportation Research Pard D: Transport and Environment,* vol. 46, pp. 1-10, 07 2016.

[15] Y. a. C. Q. a. L. J. S. L. a. X. Y. a. C. J. X. Du, “Modeling the Impacts of Tides and the Virtual Arrival Policy in Berth Allocation,” *Transportation Science,* vol. 49, 02 2015.

[16] G. a. I. C. a. K. C. a. L. A. Venturini, “The multi-port berth allocation problem with speed optimization and emission considerations,” *Transportation Research Part D: Transport and*

Environment, vol. 54, pp. 142-159, 07 2017.

[17] J. a. H. Y. a. Y. W. He, “Yard crane scheduling in a container terminal for the trade-off between efficiency and energy consumption,” *Advanced Engineering Informatics,* vol. 29, 09 2014.

[18] J. a. Z. W. a. H. Y. a. Y. W. He, “A simulation optimization method for internal trucks sharing assignment among multiple container terminals,” *Advanced Engineering Informatics,* vol. 27, pp. 598-614, 10 2013.

[19] G. a. G. K. a. G. M. Chen, “Reducing truck emissions at container terminals in a low carbon economy: proposal of a queueing-based bi-objective model for optimizing truck arrival pattern,” *Transportation Research Part E: Logistics and Transportation Review,* vol. 55, pp. 3-22, 08 2013.

[20] M. a. W. G. Aacciaro, “Energy efficiency in maritime logistics chains,” *Research in Transportation ,* vol. 17, 11 2015.

[21] G. a. Z. A.-K. a. F. J. a. M. A. Wilmsmeier, “Energy consumption and efficiency: emerging challenges from reefer trade in South American container terminals,” 2014.

[22] T. M. A. M. A. L. V. M. Christos-Antonios Gizelis, “Towards a Smart Port: The Role of the Telecom Industry,” in *Artificial Intelligence Applications and Innovations. AIAI 2020 IFIP WG 12.5 International Workshops*, 2020.

[23] M. V. M. P. K. D. F. A. a. M. H. R. M. S. Ali, “Applications of Blockchains in the Internet of Things: A Comprehensive Survey,” *IEEE Communications Surveys & Tutorials,* vol. 21, no. 2, pp. 1676-1717, 2019.

[24] Z. a. X. S. a. D. H.-N. a. C. X. a. W. H. Zheng, “An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends,” in *6th IEEE International Congress on Big Data*, 2017.

[25] F. a. d. S. a. K. S. a. M. T. a. D. S. Restuccia, “Blockchain for the Internet of Things: Present and Future,” *IEEE Internet Of Things Journal,* 01 2018.

[26] M. M. a. S. N. G. Gourisetti, “Blockchain for smart grid resilience: Exchanging distributed energy at speed, scale and security,” in *Resilience Week (RWS)*, Wilmington, DE, 2017.

[27] Z. a. W. W. a. L. G. a. L. X. a. H. J. a. H. G. Li, “Toward open manufacturing: A cross-enterprises knowledge and services exchange framework based on blockchain and edge computing,” *Industrial Management & Data Systems,* vol. 118, pp. 303-320, 02 2018.

[28] C. DataPorts, “D2.3 Blockchain Design Specification,” September 2020.

[29] I. I. o. B. Analysis, A Guide to the Business Analysis Body of Knowledge (BABOK guide), 2nd ed., 2009.

[30] K. M. Adams, “Non-functional Requirements in Systems Analysis and Design,” Springer, 2015.

[31] D. Z. N. N. Dewi Mairiza, “An Investigation into the Notion of Non-Functional requirements,” in *ACM Symposium on Applied Computing*, 2010.

[32] I. I. o. B. Analysis, A Guide to the Business Analysis Body of Knowledge (BABOK guide), 3rd ed., 2015.

[33] M. W. K. S. R. e. a. Deng, “A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements,” *Requirements Eng 16,* p. 3–32, 2011.

6.2 ACRONYMS

Acronyms List	
ACID	Atomicity, Consistency, Isolation and Durability
ACL	Access Control List
AI	Artificial Intelligence
AIS	Automatic Identification System
API	Application Programming Interface
ARIMA	Auto Regressive Integrated Moving Average
BDI	Business Data Intelligence
CA	Certificate Authority
CDM	Collaborative Decision Making
CEP	Complex Event Processing
CGM	Compagnie Générale Maritime (General Maritime Company)
CLI	Command Line Interface
CMA	Compagnie Maritime d'Affrètement (Maritime Freighting Company)
COREOR	COntainer RElease ORder
COTS	Commercial off-the-shelf
CPU	Central Processing Unit
CQL	Contextual Query Language
CSA	Cloud Security Alliance
DAPS	Dynamic Attribute Provisioning Service
DB	Database
DBaaS	Database as a Service
DBMS	Database Management System
DDoS	Distributed Denial of Service
DoS	Denial of Service
DSP	Data Sharing Platform
DTD	Document Type Definition
EDI	Electronic Data Interchange
eFTI	electronic Freight Transport Information
ERP	Enterprise Resource Planning
ETA	Estimated Time of Arrival
ETD	Estimated Time of Departure
EU	European Union
GA	Grant Agreement
GDPR	General Data Protection Regulation
GOS	Gate Operating System
GPL	General Public License
HTML	Hyper Text Markup Language

Acronyms List	
HTTP	Hyper Text Transfer Protocol
HTTPS	Hyper Text Transfer Protocol Secure
IAM	Identity and Access Management
ICT	Information Communication Technology
IDSA	International Data Spaces Association
IDS RAM	International Data Spaces Reference Architecture Model
IIS	Internet Information Services
IMO	International Maritime Organization
I/O	Input/Output
IoT	Internet of Things
IPOMS	Integrated Port Operations Management System
IT	Information Technology
JSON	JavaScript Object Notation
JSP	Java Server Pages
KPI	Key Performance Indicator
LGPL	Lesser General Public License
LRU	Least Recently Used
LSTM	Long Short-Term Memory
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
MoSCoW	MUST, SHOULD, COULD, WON'T
MPL	Mozilla Public License
MQTT	Message Queuing Telemetry Transport
MSC	Mediterranean Shipping Company
MSP	Membership Service Provider
NASA	National Aeronautics and Space Administration
NIST	National Institute of Standards and Technology
NMEA	National Marine Electronics Association
NOAA	National Oceanic and Atmospheric Administration
NODC	National Oceanic Data Centre
OCR	Optical Character Recognition
ODBC	Open Database Connectivity
OS	Operation System
OWASP	Open Web Application Security Project
PaaS	Platform as a Service
PCS	Port Community System
PHP	Hypertext Pre-processor
PMS	Port Management System
PoC	Proof of Concept
RAM	Random-access memory

Acronyms List	
RDBMS	Relational Database Management System
RDF	Resource Description Framework
REST	Representational State Transfer
RFID	Radio Frequency Identification
RMSE	Root Mean Square Error
SA	System Administrator
SLA	Service Level Agreements
SMS	Short Message Service
SOAP	Simple Object Access Protocol
SOLAS	Safety Of Life At Sea
SotA	State-of-the-Art
SSL	Secure Sockets Layer
TAS	Truck Appointment System
TBD	To Be Defined
TCP	Transmission Control Protocol
TEU	Twenty-foot Equivalent Unit
TLS	Transport Layer Security
TOS	Terminal Operating System
UI	User Interface
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
USB	Universal Serial Bus
VAT	Value Added Tax
VGM	Verified Gross Mass
WP	Work Package
WS	Web Service
XACML	eXtensible Access Control Markup Language
XML	eXtensible Markup Language

Table 51 – Acronyms

7 ANNEX 1: REQUIREMENTS AT WP LEVEL

The following tables presents the complete list of the requirements separated per WP.

7.1 WP2 (DATA PLATFORM DESIGN)

ID	2.1
Type	Non-Functional
Category	Extensibility, Interoperability
Source	Internal Technical Analysts
Priority	MUST
Description	The DataPorts Platform must be interoperable to be used for alternative architectures and dynamic systems
Rationale	Any other stakeholder could integrate the DataPorts Platform to their existing system
Responsible Task(s)	T2.5
Additional Comments	Consider common architecture standards
Test Case / Acceptance Criteria	Several data sources and organizations (together with their corresponding systems) will interact and share data with the DataPorts Platform. Test the platform by sharing/exchanging data between different data sources and organizations

ID	2.2
Type	Non-Functional
Category	Security
Source	Internal Technical Analysts
Priority	MUST
Description	The DataPorts Platform must provide a secure interface framework for data exchange between itself and the potential data sources

Rationale	Any platform's component to fulfil minimum security and privacy constraints
Responsible Task(s)	T2.5
Additional Comments	Consider security standards (NIST, CSA etc.)
Test Case / Acceptance Criteria	The DataPorts Platform (according to the defined scenarios) share/exchange data in two secure/privacy-protected approaches: off-chain data through IDSA connectors (peer-2-peer), and on-chain data were verifiable and immutable data is stored in Blockchain

ID	**2.3**
Type	Non-Functional
Category	Certification
Source	Internal Technical Analysts
Priority	MUST
Description	The definition of the architecture must be aligned with the last version of the Industrial Data Space Architecture Reference Model
Rationale	This specific requirement is explicitly expressed in the T2.5 description and introduced in the architecture section of the GA (1.3.2)
Responsible Task(s)	T2.5
Additional Comments	
Test Case / Acceptance Criteria	Verify if the major roles of the IDS RAM are implemented and included in the DataPorts Platform. Try to interact with the existing connectors, data sources of the IDS environment

ID	**2.4**
Type	Non-Functional

Category	Certification
Source	Internal Technical Analysts
Priority	MUST
Description	The architecture must follow a federated data approach avoiding the “data-centric model”
Rationale	Described in the ambition section of the GA, T3.1 and aligned with the approach proposed by IDSA (requirement 2.3)
Responsible Task(s)	T2.5
Additional Comments	
Test Case / Acceptance Criteria	Apart from temporal data stored for integration or data analytics, no data is stored in the platform, and the platform is completely decentralized

ID	**2.5**
Type	Non-Functional
Category	Portability
Source	Internal Technical Analysts, End-Users
Priority	MUST
Description	The DataPorts Platform must be deployed in different infrastructures (cloud, on premise, etc.)
Rationale	Aligned with IDSA and requested by the Valencia Port pilot End-Users
Responsible Task(s)	T2.5
Additional Comments	
Test Case / Acceptance Criteria	Choose at least one of the pilot scenarios and test the platform in their ecosystem (the platform is on prototype level)

ID	2.6
Type	Non-Functional
Category	Portability
Source	Internal Technical Analysts
Priority	SHOULD
Description	DataPorts should follow the Big Data Analytics as a Service (BDAaaS) paradigm, providing a level of abstraction to application developers about the implementation and set-up details of the data platform, thus simplifying deployment
Rationale	From the architecture section of the GA (1.3.2) and T3.4 description. The idea of building cognitive services on top of the DataPorts Platform is also aligned with this paradigm
Responsible Task(s)	T2.5
Additional Comments	
Test Case / Acceptance Criteria	With different data sets as input, test the functionality of the data analytics as a service in the pilot scenarios

ID	2.7
Type	Functional
Category	Usability
Source	Internal Technical Analysts
Priority	MUST
Description	The DataPorts Platform must provide resources (user guides, documentation, etc.) to support the provided functionalities of the platform
Rationale	In order to assist the usage and increase the adoption of the DataPorts Platform
Responsible Task(s)	T2.5

Additional Comments	
Test Case / Acceptance Criteria	Review there exist links with information, instructions and guides, they are accessible, and can be downloaded/read by any user

7.2 WP3 (DATA PLATFORM IMPLEMENTATION AND SERVICES)

ID	3.1
Type	Functional
Category	Interoperability
Source	Internal Technical Analysts
Priority	MUST
Description	The DataPorts Platform must homogenize the input data so that it can be retrieved in the same format and data model regardless of the data source
Rationale	Due to the big amount of data sources, DataPorts must transform the data to a homogenized format and data model. This process takes care of the structural and content translation to unify its format
Responsible Task(s)	T3.1
Additional Comments	
Test Case / Acceptance Criteria	Check that the data stored in the context broker from the data access agents meets the structure and requirements of the defined data model

ID	3.2
Type	Functional
Category	Interoperability
Source	Internal Technical Analysts
Priority	SHOULD

Description	The Data Access component should be able to make use of the interface of each data source connected to the DataPorts Platform in order to retrieve the data in its original format. Supported data interchange formats should be at least: XML and JSON. The Data Access component should also support the security mechanisms implemented in each data source
Rationale	Each data source interface needs a network connectivity to the DataPorts infrastructure that matches the security requirements of the data exposed by this interface. Secure and encrypted link should be preferred. The applications require common data formats and application programming interfaces (APIs) so data can be accessed and combined as needed
Responsible Task(s)	T3.1
Additional Comments	
Test Case / Acceptance Criteria	Check that the Data Access component has connectivity to the data sources APIs and the data retrieved is secured under the DataPorts Platform security components

ID	**3.3**
Type	Functional
Category	Interoperability
Source	Internal Technical Analysts
Priority	COULD
Description	Each data source that is going to be connected to the DataPorts infrastructure through an API could provide version management for its API, in order to allow a proper identification of the API and the data formats
Rationale	Data sources interfaces may not have a unique implementation or they may offer different versions. Therefore, it is necessary to consider all these factors in order to detect and anticipate

	format change issues and achieve a correct development of the agents to connect with the data sources
Responsible Task(s)	T3.1
Additional Comments	
Test Case / Acceptance Criteria	Check that every data source connected to DataPorts through an API provides version management

ID	3.4
Type	Functional
Category	Interoperability
Source	Internal Technical Analysts
Priority	SHOULD
Description	Each data source that is going to be connected to the DataPorts infrastructure through an API should provide documentation about it (such as an OpenAPI/Swagger definition)
Rationale	This API documentation would help building the acquisition connector. A standard definition of the data source API would ease the development of the connector
Responsible Task(s)	T3.1
Additional Comments	
Test Case / Acceptance Criteria	Check that every data source connected to DataPorts through an API provides documentation about how to connect and gather data

ID	3.5
Type	Functional

Category	Interoperability
Source	Internal Technical Analysts
Priority	MUST
Description	The DataPorts Platform must provide an ontology that describes the concepts from the data models of all the connected data sources so that all the input data can be represented following a common ontology
Rationale	The O3.2 of WP3 is to define ontologies, mechanisms and enablers to provide semantic interoperability with data platforms, IoT devices, robots and other data sources, and develop the semantic-based tools needed to facilitate generation of interfaces for sensing and actuation required in the DataPorts data platform. For that reason, this requirement is needed in order to provide a target ontology for all the available data sources
Responsible Task(s)	T3.2
Additional Comments	
Test Case / Acceptance Criteria	**Given** a set of data sources with heterogeneous data format/data model/ontology, **when** I want to provide a target ontology that covers this heterogeneous data sources, **then** the target ontology must define and describe the concepts and relationships of the data sources information

ID	**3.6**
Type	Functional
Category	Interoperability
Source	Internal Technical Analysts
Priority	MUST
Description	The DataPorts Platform must be able to obtain data from platform/sensors/sources with and without

	explicitly defined ontology in order to support as many potential data sources as possible
Rationale	The data sources can be classified into the following categories: a) data sources that provide their own ontologies or use a standard one, b) data sources that have the possibility of defining an ontology or custom data model, c) data sources that have no ontologies. In order to handle as many sources as possible, the agents must be able to support all of these cases
Responsible Task(s)	T3.1
Additional Comments	
Test Case / Acceptance Criteria	**Given** a message/data produced by a data source, **when** I want to perform the translation to the common ontology using the agent, **then** the agent must be able to translate the data independently of the existence of an ontology for the data source

ID	**3.7**
Type	Functional
Category	Interoperability
Source	Internal Technical Analysts
Priority	MUST
Description	The DataPorts Platform must provide publish/subscribe operations to internal components in order to provide access to business real-time data from the connected sources
Rationale	In order to provide interoperability and integration of data collected from multiple sources for enabling real-time decisions and control of assets at seaport level and along the whole supply chain
Responsible Task(s)	T3.2
Additional Comments	
Test Case / Acceptance Criteria	**Given** a message/data produced by a data source, **when** I want to send the data to the internal

	components subscribed to that topic, **then** the information must arrive only to the subscribed internal components

ID	3.8
Type	Functional
Category	Interoperability
Source	Internal Technical Analysts
Priority	MUST
Description	The Data Access component must provide the proper connectors to communicate with the available data sources
Rationale	The O3.1 of WP3 is to identify different data sources to be integrated in the DataPorts data platform, including the mechanisms to store and facilitate data management. The interfaces of these data sources may be categorized depending on the target datasets, data platforms and stakeholders, in order to define a common approach to provide a homogeneous set of access tools
Responsible Task(s)	T3.1
Additional Comments	
Test Case / Acceptance Criteria	**Given** a data source, **when** I want to access it to retrieve data, **then** the framework for agent creation developed in T3.1 must provide the necessary tools to communicate with the desired data source

ID	3.9
Type	Functional
Category	Interoperability
Source	Internal Technical Analysts

Priority	SHOULD
Description	The Data Access component should support the most common communication protocols in order to be able to communicate with as many potential data sources as possible
Rationale	In order to facilitate the inclusion of new data providers in the DataPorts Platform and handle as many data sources as possible in an easy way
Responsible Task(s)	T3.1
Additional Comments	
Test Case / Acceptance Criteria	**Given** a data source accessible using an established communication protocol, **when** I want to access it using the DataPorts Platform, **then** the framework for agent creation developed in T3.1 should provide the necessary tools to facilitate the creation of the agent for the data source

ID	**3.10**
Type	Functional
Category	Interoperability
Source	Internal Technical Analysts
Priority	MUST
Description	The Semantic Interoperability layer must support semantic modelling in order to provide an unambiguous meaning of the data
Rationale	Semantic interoperation defines the rules for understanding the meaning of the content of information and creates a domain specific information model, known as semantic model
Responsible Task(s)	T3.2
Additional Comments	
Test Case / Acceptance Criteria	**Given** a message in the data source format, **when** I want to send it to components belonging to the upper layer of the DataPorts Platform, **then** the

	message must be translated to the semantic target model defined in DataPorts

ID	3.11
Type	Non-Functional
Category	Privacy (Transparency), Trust
Source	Internal Technical Analysts
Priority	MUST
Description	As a Blockchain organization, I want to be able to interconnect with other organizations within my network in an agreed upon manner, to facilitate trust and transparency
Rationale	To facilitate data exchange between different untrusting organizations, an infrastructure providing transparent and trusted interconnection of such organizations must be created
Responsible Task(s)	T3.5
Additional Comments	
Test Case / Acceptance Criteria	Setup of Blockchain network based on Fabric Blockchain technology, including immutable ledger, orderer and peer nodes to ensure consensus, CA nodes to allow certification and channels to allow data privacy

ID	3.12
Type	Functional
Category	Privacy (Transparency), Trust
Source	Internal Technical Analysts
Priority	MUST

Description	As a Blockchain organization, I would like to be able to onboard an existing Blockchain network and participate as a validating member
Rationale	An organization must be able to join a new/existing Blockchain network, have Blockchain network components such as endorsing/validating/anchor peers, add an orderer node if needed, use a CA to provide its member components and end-users and applications with appropriate certificates
Responsible Task(s)	T3.5
Additional Comments	
Test Case / Acceptance Criteria	Add a new organization to existing Blockchain network: configure a new CA node, add the organization with appropriate access roles to network configuration, join organizational peer to channels

ID	**3.13**
Type	Functional
Category	Consensus, Transparency, Trust
Source	Internal Technical Analysts
Priority	MUST
Description	As a participating Blockchain organization, I would like to be able to agree on a consensus algorithm which will provide pilot appropriate transaction validation and ordering functionality for the network
Rationale	Blockchain network's core is the consensus algorithm used to assure integrity and security of the network. The distributed nodes of the network need to agree on the order and validity of the transactions
Responsible Task(s)	T3.5
Additional Comments	

Test Case / Acceptance Criteria	Use a validated and known consensus algorithm in setting up ordering service of Blockchain network, i.e. RAFT

ID	3.14
Type	Non-Functional
Category	Security
Source	Internal Technical Analysts
Priority	MUST
Description	As a participating Blockchain organization, I would like to be able to securely sign in my network components (peers, CAs) and clients (end-users/application clients) to a Blockchain network
Rationale	Blockchain network's transparency and trust are reliant on several factors: authorized participants for data sharing and data access, ability to handle malicious participants, ability to withstand security attacks from outside. Part of those abilities are provided by strict authentication and authorization protocols of the Blockchain technologies which need to be configured for each pilot
Responsible Task(s)	T3.5
Additional Comments	
Test Case / Acceptance Criteria	Design of CA and MSP configuration for each pilot ensuring proper certification, configure organization access roles in network and channel configuration, encode attribute-based access restrictions in pilot's chaincodes

ID	3.15
Type	Functional
Category	Scalability, Privacy

Source	Internal Technical Analysts
Priority	MUST
Description	As a participating Blockchain organization, I would like to be able to maintain my own copy of distributed ledger while being assured other organizations on the network have exact same replica of the ledger
Rationale	Organizations in a business network would like to have the ability to maintain a shared, distributed single source of truth for all the transactions within the network. Some of those organizations for various reasons of security, performance, regulations etc., need to maintain their own copy of the shared ledger
Responsible Task(s)	T3.5
Additional Comments	
Test Case / Acceptance Criteria	Create channels as appropriate for each pilot to ensure privacy of data within the Blockchain network. Add peers of organizations that need separate copy of the ledger to the channels

ID	**3.16**
Type	Functional
Category	Trust, Privacy (Transparency), Verifiability, Accountability
Source	Internal Technical Analysts
Priority	MUST
Description	As a participating Blockchain organization, I would like to be able to endorse transactions by executing agreed upon business logic specified as a bounding contract between network organizations
Rationale	Organizations in a business network would like to potentially execute agreed upon business logic which will be required to validate and sign on

	transactions, achieving transparency and accountability across the network
Responsible Task(s)	T3.5
Additional Comments	
Test Case / Acceptance Criteria	For organizations required to endorse transactions for a particular pilot, add endorsing peers to the appropriate channels

ID	**3.17**
Type	Non-Functional
Category	Trust, Verifiability, Accountability
Source	Internal Technical Analysts
Priority	MUST
Description	As a participating Blockchain organization, I would like to be assured of immutability and permanent availability of the shared data
Rationale	Organizations in a business network would like to assure that the information shared with other organizations in the network and stored in the ledger is immutable and untamperable. This allows for environment for trusted data exchange, easy dispute resolution, accountability and verifiability of business processes within the network
Responsible Task(s)	T3.5
Additional Comments	
Test Case / Acceptance Criteria	Use appropriate Blockchain technology (agreed upon Hyperledger Fabric) which ensures immutability and untamperability of the underlying ledger

ID	**3.18**

Type	Functional
Category	Interoperability
Source	Internal Technical Analysts
Priority	MUST
Description	The Data Abstraction and Virtualization component must have an open API in order for big vendors as well as new providers to be able to publish their services and components
Rationale	The Data Abstraction and Virtualization component must be easy to understand and use
Responsible Task(s)	T3.3
Additional Comments	
Test Case / Acceptance Criteria	Rely on the OpenAPI specification, which is a homogenous standardized solution for describing REST services

ID	**3.19**
Type	Non-Functional
Category	Interoperability
Source	Internal Technical Analysts
Priority	MUST
Description	The Data Abstraction and Virtualization component must be data-source independent
Rationale	The structure of the metadata format as well as the way that the metadata are saved must be orchestrated in a way that every future candidate data source could be supported
Responsible Task(s)	T3.3
Additional Comments	
Test Case / Acceptance Criteria	Data Abstraction and Virtualization component takes input from the Data Access / Semantic

	Interoperability layer, where the data has been already transformed into JSON format

ID	3.20
Type	Non-Functional
Category	Interoperability
Source	Internal Technical Analysts
Priority	SHOULD
Description	The data model specification should follow the semi-structured format
Rationale	In order to achieve proper message interchange and schema creation and describe different aspects of heterogeneous data sources
Responsible Task(s)	T3.3, T3.2
Additional Comments	
Test Case / Acceptance Criteria	The data model specification is created using JSON semi-structured markup language

ID	3.21
Type	Non-Functional
Category	Compatibility
Source	Internal Technical Analysts
Priority	MUST
Description	The notation language must be able to be parsed by multiple different programming languages
Rationale	In order to be programmable and used from multiple programming families
Responsible Task(s)	T3.3

Additional Comments	
Test Case / Acceptance Criteria	The selection of JSON language is vital cause of the ease of its implementation to the majority of the programming languages such as Python, Java, C# etc. JSON libraries are very easy to include in any of the aforementioned languages (JSON's library for Python is called... json)

ID	**3.22**
Type	Non-Functional
Category	Interoperability
Source	Internal Technical Analysts
Priority	SHOULD
Description	The notation language should be human readable, easy to script and to understand
Rationale	The notation language should hide the complexity of the architecture and components
Responsible Task(s)	T3.3, T3.2
Additional Comments	
Test Case / Acceptance Criteria	The data model is described using JSON schema

ID	**3.23**
Type	Functional
Category	Virtualization
Source	Internal Technical Analysts
Priority	COULD
Description	The components of the DataPorts Platform could be virtualized, in order to ease its deployment and portability

Rationale	This allows improving the compatibility, maintainability, isolation, portability and scalability of the developed components
Responsible Task(s)	T3.1, T3.2, T3.3, T3.4, T3.5
Additional Comments	
Test Case / Acceptance Criteria	**Given** a DataPorts Platform component, **when** I want to deploy and run a new instance, **then** it has to be able to be containerized and deployed in a Docker environment

ID	**3.24**
Type	Functional
Category	Virtualization
Source	Internal Technical Analysts
Priority	COULD
Description	The DataPorts Platform could offer mechanisms for the automatic deployment, maintenance and scaling of the developed components
Rationale	In order to facilitate the adoption of the DataPorts Platform
Responsible Task(s)	T3.1, T3.2, T3.3, T3.4, T3.5
Additional Comments	
Test Case / Acceptance Criteria	**Given** a DataPorts Platform component, **when** I want to run a new instance of this component, **then** it has to be automatically deployed and configured

ID	**3.25**
Type	Non-Functional
Category	Scalability

Source	Internal Technical Analysts
Priority	MUST
Description	The Semantic Interoperability component must support distributed real-time stream processing in order to avoid the overload of specific components when high data volumes are processed
Rationale	Data can grow exponentially and can quickly overload specific components based on the data volume
Responsible Task(s)	T3.2
Additional Comments	
Test Case / Acceptance Criteria	**Given** a considerable amount of real time data, **when** the performance of the platform could be affected, **then** a new instance of the data broker could be deployed in order to process the information without affect the work and performance of the whole DataPorts Platform

ID	**3.26**
Type	Functional
Category	Functionality
Source	End-Users
Priority	MUST
Description	As an end-user, I want the platform to provide cognitive services specific to ports requirements, so that I could improve my decision making processes and/or KPIs
Rationale	From the ambition section of the GA (1.4), but also an outcome from the pilot's use case descriptions
Responsible Task(s)	T3.4
Additional Comments	
Test Case / Acceptance Criteria	**Given** a cognitive service to predict the future value of a ports KPI or event, **when** I request the expected

	value for a specific time period (i.e. the next hour), **then** the service must report visually the prediction and the accuracy

ID	3.27
Type	Functional
Category	Performance Efficiency
Source	Internal Technical Analysts
Priority	MUST
Description	As a developer, I want an abstraction mechanism regarding the implementation and set-up details of the data sources connection, so that the deployment will be faster and easier
Rationale	Outcome from T3.4 and related with the Big Data Analytics as a Service (BDAaaS) approach, described in the architecture section. The idea is to reduce deployment time
Responsible Task(s)	T3.4
Additional Comments	
Test Case / Acceptance Criteria	**Given** a data source available in the DataPorts Platform, **when** I want to query such data, **then** the retrieved data must be the same as if I connect directly to the original data source

ID	3.28
Type	Functional
Category	Functionality
Source	End-Users
Priority	MUST
Description	As an end-user, I want software components based on State-of-the-Art Machine Learning (ML)

	algorithms, specifically customized to the ports domain, so that I could easily and automatically create models
Rationale	Mainly from the T3.4 goals and the ideas depicted in the ambition section of the GA (1.4). Usually, developers lack of expertise on ML algorithms. This requirement must provide ready-to-use (as a library or framework) model trainers with easy configuration
Responsible Task(s)	T3.4
Additional Comments	
Test Case / Acceptance Criteria	**Given** a port-oriented indicator that I want to predict, **when** I require to build a ML model to fulfil such task, **then** the DataPorts Platform must test several techniques (ARIMA, LSTMs, Random Forest etc.) and select the best model according to accuracy metrics (RMSE, MAPE etc.)

ID	**3.29**
Type	Non-Functional
Category	Scalability
Source	End-Users
Priority	MUST
Description	As an end-user, I want a distributed AI platform, so that huge data volumes and time consuming tasks could be achieved
Rationale	From the ambition section of the GA (1.4) and T3.4. This distributed platform will be based on current technologies (Spark, Hadoop, etc.)
Responsible Task(s)	T3.4
Additional Comments	
Test Case / Acceptance Criteria	**Given** a dataset which cannot be analyzed in a single computer, **when** I require to build a ML model using that dataset, **then** the DataPorts Platform will

	distribute such processing in a distributed infrastructure for improving performance

ID	3.30
Type	Functional
Category	Functionality, Performance Efficiency
Source	End-Users
Priority	SHOULD
Description	As an end-user, I want to use continuous data streams, so that the platform provides predictions in near real-time
Rationale	From T3.4 description
Responsible Task(s)	T3.4
Additional Comments	
Test Case / Acceptance Criteria	**Given** a data source in the DataPorts Platform that periodically generates new data, **when** I want to generate new predictions as new data arrives, **then** a ML model subscribed to such data publishes such predictions

ID	3.31
Type	Non-Functional
Category	Interoperability
Source	Internal Technical Analysts
Priority	MUST
Description	As a developer, I want the platform to deal with the original data sources heterogeneity, real-time (streaming) or persistent data, relational or non-relational databases (NoSQL), so that I can use the

	data without taking into account the underlying storage system
Rationale	From T3.3 description and overall architecture, regarding federated approach
Responsible Task(s)	T3.1
Additional Comments	
Test Case / Acceptance Criteria	**Given** a data source available in the DataPorts Platform, **when** I want to query such data, **then** the data must be retrieved without specifying technological implementation details

ID	**3.32**
Type	Functional
Category	Functionality, Interoperability
Source	Internal Technical Analysts
Priority	MUST
Description	As a data provider, I want to publish data in the DataPorts Platform, so that it is available for the platform users with the access rights
Rationale	In order to send my own data to other DataPorts instances (T3.1). The platform must provide an agent that receives data from the data source and processes it for every kind of shared data
Responsible Task(s)	T3.1, T3.2
Additional Comments	
Test Case / Acceptance Criteria	Data shared from the data sources is sent/gathered to/from the DataPorts Platform through the existing agents

ID	**3.33**

Type	Functional
Category	Functionality, Interoperability
Source	Internal Technical Analysts
Priority	MUST
Description	As a data consumer, I want to get the list of the available data sources and all the methods provided by the platform to subscribe or request data on demand
Rationale	In order to be aware of the data that is available through the use of the platform and the methods that should be called for retrieving it
Responsible Task(s)	T3.1, T3.2, T3.3, T3.4
Additional Comments	
Test Case / Acceptance Criteria	**Given** a data source available in the DataPorts Platform, **when** a data consumer wants to know the full list of available methods, **then** the platform must return all the information needed to retrieve the data

ID	**3.34**
Type	Functional
Category	Functionality, Interoperability
Source	Internal Technical Analysts
Priority	MUST
Description	As a data consumer, I want to subscribe to an available subscription provided by the DataPorts Platform
Rationale	In order to get updates from the data sources through the platform
Responsible Task(s)	T3.1, T3.2, T3.4
Additional Comments	

Test Case / Acceptance Criteria	**Given** a data source available in the DataPorts Platform, **when** a data consumer wants to get updates from it and subscribes through the provided methods, **then** the platform must create a subscription for the user

ID	**3.35**
Type	Functional
Category	Functionality, Interoperability
Source	Internal Technical Analysts
Priority	MUST
Description	As a data consumer, I want to be able to cancel a current subscription, so that I stop receiving data modifications
Rationale	In order to stop receiving data updates. The platform must provide an API that lets data consumers to delete their subscriptions to data sources
Responsible Task(s)	T3.1, T3.2, T3.4
Additional Comments	
Test Case / Acceptance Criteria	**Given** a data source available in the DataPorts Platform, when a data consumer wants to delete an active subscription, **then** the platform must delete the subscription and stop sending updates

ID	**3.36**
Type	Functional
Category	Functionality, Interoperability
Source	Internal Technical Analysts
Priority	MUST

Description	As a data consumer, I want to request data on demand from the data sources using the methods provided by the DataPorts Platform
Rationale	In order to get information of other DataPorts instances on demand. The platform must provide an API that lets the data consumer to request data through the available methods
Responsible Task(s)	T3.1, T3.2, T3.3
Additional Comments	
Test Case / Acceptance Criteria	**Given** a data source available in the DataPorts Platform, **when** a data consumer wants to get data from it and calls an on demand provided method, **then** the platform must return the data to the user

ID	**3.37**
Type	Functional
Category	Performance Efficiency
Source	Internal Technical Analysts
Priority	COULD
Description	The DataPorts Platform could reduce and process the data on the ports' side, before they reach the Data Abstraction and Virtualization component's repository (Virtual Data Repository – VDR) and become available to distributed resources
Rationale	Processing and reducing data before they are stored in a central location allows the distribution of task, I/O overhead, as well as the reduction of resources space, needed for the storage and the transmission of data
Responsible Task(s)	T3.2
Additional Comments	
Test Case / Acceptance Criteria	Process the semantic transformation as close to the source as possible / Deploy and run the Semantic

	Interoperability component on one of the pilots premises

ID	3.38
Type	Non-Functional
Category	Scalability, Performance Efficiency
Source	Internal Technical Analysts
Priority	SHOULD
Description	As a data consumer, I want low latency in order to gain quick access and precise feedback
Rationale	In order to improve scalability and efficiency of dealing with a large number of transactions
Responsible Task(s)	T3.5
Additional Comments	
Test Case / Acceptance Criteria	Design an hierarchy of smart contracts as appropriate for each pilot to ensure scalability and efficiency of transaction processing. Ensure separation of concerns when designing the smart contracts

ID	3.39
Type	Functional
Category	Functionality
Source	Internal Technical Analysts
Priority	MUST
Description	As a data provider, I want to retrieve the history of transactions, so that I am able to track the information flow
Rationale	In order to view assets intermediate transactions

Responsible Task(s)	T3.5
Additional Comments	
Test Case / Acceptance Criteria	As appropriate for each pilot, add invoke functions in chaincode to return a list of historical transactions

ID	**3.40**
Type	Non-Functional
Category	Interoperability
Source	Internal Technical Analysts
Priority	MUST
Description	DataPorts must provide a common data model schema to get the training data
Rationale	Related with the IDSA broker. A metadata repository is required for data sources discovery
Responsible Task(s)	T3.5, T3.2
Additional Comments	
Test Case / Acceptance Criteria	**Given** a data source available in the DataPorts Platform, **when** I want to query such data, **then** the data description (schema, columns, metadata) must conform to the DataPorts ontology

ID	**3.41**
Type	Functional
Category	Quality
Source	Experience from previous projects

Priority	SHOULD
Description	The DataPorts Platform should have (efficient) provisions for checking data quality (e.g., to detect concept drift, missing data, inconsistent data etc.)
Rationale	80% of development effort in previous projects went into data integration and quality
Responsible Task(s)	T3.3
Additional Comments	
Test Case / Acceptance Criteria	Outliers detection based on statistical measures like standard deviation or quantiles. Checking for missing values to be removed or replaced. Searching for duplicates that must be eliminated, based on statistical methods. Tests using unclean datasets will be taken place. Data Abstraction and Virtualization's (DAV's) efficiency will be evaluated based on how well the unclean datasets were "cleaned" and corrected. Cleaning will take place in DAV's Processing & Filtering Software, with the cleaned datasets being stored at Virtual Data Repository later on

ID	**3.42**
Type	Non-Functional
Category	Quality
Source	Experience from previous projects
Priority	SHOULD
Description	The DataPorts Platform should deliver cleansed, integrated etc. data to the analytics services
Rationale	Data quality is a concern for analytics ("garbage in - garbage out"). Significant further accuracy improvements of predictive algorithms only via better data and not anymore via better algorithms or hyperparam-tuning; e.g., in predictive process monitoring, ca. 5% improvement in accuracy considered good; but experiments with better data collection show up to 250% improvements!

Responsible Task(s)	T3.3
Additional Comments	
Test Case / Acceptance Criteria	Outliers detection based on statistical measures like standard deviation or quantiles. Checking for missing values to be removed or replaced. Searching for duplicates that must be eliminated, based on statistical methods. Tests using unclean datasets will be taken place. Data Abstraction and Virtualization's (DAV's) efficiency will be evaluated based on how well the unclean datasets were "cleaned" and corrected. Cleaning will take place in DAV's Processing & Filtering Software, with the cleaned datasets being stored at Virtual Data Repository later on

ID	**3.43**
Type	Non-Functional
Category	Compatibility
Source	Internal Technical Analysts
Priority	MUST
Description	Integration of legacy sources. The DataPorts Platform must be able to integrate with relevant port platforms
Rationale	In the port environment there are several data providers sharing data. There is a need to have a platform to share data in a secure way
Responsible Task(s)	T3.1
Additional Comments	
Test Case / Acceptance Criteria	Data from existing data sources is published into the DataPorts Platform through the Data Access component agents

ID	**3.44**

Type	Functional
Category	Interoperability
Source	Internal Technical Analysts
Priority	SHOULD
Description	Data correlation. The data from different sources should be correlated (virtual object) or in the same process
Rationale	In the port environment there are several data providers providing data from the same operation/object. It is needed to know that is the same operation/object and correlate the data
Responsible Task(s)	T3.3
Additional Comments	
Test Case / Acceptance Criteria	Using correlation algorithms such as Pearson's and Spearman's to summarize the strength of the linear relationship between two data samples. Correlation functions will be included in the Data Abstraction and Virtualization component. Correlated datasets' scores will be close to 1 or -1 (that is, positive or negative correlation). Uncorrelated datasets' scores will be close to 0

ID	**3.45**
Type	Functional
Category	Interoperability
Source	Internal Technical Analysts
Priority	MUST
Description	The DataPorts Platform must provide and API for developers in order to build specific applications or services
Rationale	New applications can be considered based on the data available and cognitive services from the

	platform, so an API is needed to allow the integration
Responsible Task(s)	T3.4
Additional Comments	
Test Case / Acceptance Criteria	**Given** the API of a specific component, **when** I request the documentation of such API, **then** it must be properly documented (for REST APIs we will use OpenAPI specifications) pointing out the input and output messages

ID	**3.46**
Type	Non-Functional
Category	Quality
Source	Internal Technical Analysts
Priority	SHOULD
Description	Data quality. Before correlate data from different sources, quality should be checked to guarantee the correct interpretation
Rationale	The data from some data sources may not have the needed quality. If these data is going to be correlated with other data, they need to have a minimum level of quality
Responsible Task(s)	T3.3
Additional Comments	
Test Case / Acceptance Criteria	Outliers detection based on statistical measures like standard deviation or quantiles. Checking for missing values to be removed or replaced. Searching for duplicates that must be eliminated, based on statistical methods. Tests using unclean datasets will be taken place. Data Abstraction and Virtualization's (DAV's) efficiency will be evaluated based on how well the unclean datasets were "cleaned" and corrected. Cleaning will take place in DAV's Processing & Filtering Software, with the cleaned datasets being stored at Virtual Data Repository later on

ID	3.47
Type	Functional
Category	Trust
Source	Internal Technical Analysts
Priority	SHOULD
Description	As a Blockchain organization, I would like to be informed about the name and number of the other participant organizations
Rationale	It would be useful to know about the name and the job sector of each participant
Responsible Task(s)	T3.5
Additional Comments	
Test Case / Acceptance Criteria	Deploy a fabric network and create a mechanism to get info about the participant organizations. It is needed to have the capability to query the network for participating organizations, in order to assign sharing and governance rules at an organization level

7.3 WP4 (DATA GOVERNANCE AND SECURITY)

ID	4.1
Type	Non-Functional
Category	Trust, Verifiability, Accountability
Source	Internal Technical Analysts
Priority	MUST
Description	As a participating Blockchain organization, I would like to be able to develop and run agreed upon business logic in a distributed manner and agree on results

Rationale	Organizations in a business network would like to execute agreed upon business logic in an automated manner to validate and sign on transactions, to achieve transparency and accountability across the network. The platform must provide execution runtime to allow execution of such business logic (aka smart contracts)
Responsible Task(s)	T4.2
Additional Comments	
Test Case / Acceptance Criteria	Deploy smart contracts executing pilot's business logic on the appropriate Blockchain network

ID	**4.2**
Type	Non-Functional
Category	Privacy
Source	Internal Technical Analysts
Priority	MUST
Description	As a participating Blockchain organization, I would like to share data in a privacy preserving manner with selected organizations within the business network
Rationale	In multiple business scenarios, organizations would like the ability to share data and/or business logic with subset of their business network, or alternatively share part of the data for public access within this network and hold some data private
Responsible Task(s)	T4.3
Additional Comments	
Test Case / Acceptance Criteria	Use Blockchain networks for each pilot with stakeholder organizations to share data in the network. Use channels between network participants who need to share data in a manner invisible to other participants

ID	4.3
Type	Non-Functional
Category	Security (Authentication, Authorization)
Source	Internal Technical Analysts
Priority	MUST
Description	As a participating Blockchain organization, I would like to enroll my client applications and end users to access and share data in my Blockchain network in a secure manner using appropriate authentication
Rationale	Organizations require the means to be able to connect organizational applications via secure APIs into Blockchain network to share and consume shared data. This must be done in a authenticated manner to ensure only the approved parties (such as client applications and end users) are allowed to access the Blockchain network
Responsible Task(s)	T4.3
Additional Comments	
Test Case / Acceptance Criteria	Certify end users, so they can be enrolled and registered as such to invoke Blockchain client applications. Use Fabric CA to provision the certificates and Fabric authentication mechanism for enrolling and registering those certificates

ID	4.4
Type	Non-Functional
Category	Security (Authentication, Authorization)
Source	Internal Technical Analysts
Priority	MUST
Description	As a participating Blockchain organization, I would like to enroll my client applications and end users to access and share data in my Blockchain network in

	a secure manner using appropriate authorization mechanisms
Rationale	Organizations require the means to be able to connect organizational applications via secure APIs into Blockchain network to share and consume shared data. This must be done in an authorized manner to ensure only the approved parties (such as client applications and end users) are allowed to see data according the specified roles and rules
Responsible Task(s)	T4.3
Additional Comments	
Test Case / Acceptance Criteria	Use Blockchain authorization mechanisms (resource level ACLs, for managing which orgs/users can add organizations to network, which can access channel data), smartcode level authorization (such as role based authorization for users determining who can read/write what data) to determine access to onchain data

ID	**4.5**
Type	Non-Functional
Category	Compliance, Security, Service Level Agreements
Source	Internal Technical Analysts
Priority	MUST
Description	As a data provider, I would like to have an agreement with data consumer about data manipulation
Rationale	Data provider and data consumer have achieved an agreement that is stored on-chain in which rules, obligations and prohibitions are specified
Responsible Task(s)	T4.2, T4.3, T4.4
Additional Comments	
Test Case / Acceptance Criteria	Deploy smart contracts that store rules, obligation and prohibitions about data manipulation

ID	4.6
Type	Non-Functional
Category	Security
Source	Internal Technical Analysts
Priority	MUST
Description	As a data provider, I would like not only to grant access to data consumer, but also to revoke access
Rationale	Data access changes dynamically
Responsible Task(s)	T4.2, T4.4
Additional Comments	
Test Case / Acceptance Criteria	Deploy a Hyperledger Fabric application to overview datasets and assigned permissions

ID	4.7
Type	Non-Functional
Category	Extensibility
Source	Internal Technical Analysts
Priority	SHOULD
Description	As a participant Blockchain organization, I want to set data governance rules, so that data access is specified
Rationale	In order to incorporate new business rules and enforce the existing ones
Responsible Task(s)	T4.2, T4.4
Additional Comments	
Test Case / Acceptance Criteria	Create Blockchain data governance usecase implementation network, which provides ability to

	add/query new data governance rules for off chain data. Ensure that onchain data for data sharing scenarios have policies and access roles in place to govern access to the data

ID	4.8
Type	Non-Functional
Category	Privacy
Source	Internal Technical Analysts - Privacy impact assessment guideline for RFID applications (BSI) - Privacy and data protection impact assessment
Priority	MUST
Description	The DataPorts Platform must ensure the privacy of personal data required for authentication purposes
Rationale	GDPR
Responsible Task(s)	T4.1
Additional Comments	User must not upload sensitive or personal data without a previous anonymization process
Test Case / Acceptance Criteria	Verify that personal data is adequate, relevant and limited to what is necessary, in relation to the purposes for which they are processed if needed and that this data is not used for any other purpose

ID	4.9
Type	Non-Functional
Category	Privacy
Source	Internal Technical Analysts - Privacy impact assessment guideline for RFID applications (BSI) - Privacy and data protection impact assessment
Priority	COULD

Description	The DataPorts Platform could ensure personal data minimization by authentication (stripping)
Rationale	GDPR
Responsible Task(s)	T4.1
Additional Comments	
Test Case / Acceptance Criteria	1. Verify that personal data must be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed. Essentially, it means that data cannot be processed, unless it is needed to process them, in order to achieve the above-mentioned purposes 2. Identify platform process and information/data managed on each one, guaranteeing that only needed and minimal information is used for its purposes

ID	**4.10**
Type	Non-Functional
Category	Privacy, Security (Confidentiality)
Source	Internal Technical Analysts
Priority	MUST
Description	The DataPorts Platform must control the usage of its services only to authorized subjects, guaranteeing the separation of roles and functions
Rationale	COM (2018) 279 final
Responsible Task(s)	T4.1
Additional Comments	
Test Case / Acceptance Criteria	1. Review that admin users or general only access to personal data considering specific process, and due to authorized and controlled activities, and only can modify data considering different privileges 2. Test access to services, using a) admin account with privileges b) user account with no privileges

ID	4.11
Type	Non-Functional
Category	Privacy, Usability
Source	Internal Technical Analysts
Priority	SHOULD
Description	The DataPorts Platform should provide the users the ability to manage and control their requirements concerning personal data
Rationale	GDPR
Responsible Task(s)	T4.1
Additional Comments	
Test Case / Acceptance Criteria	Login to manager portal with user credentials; Check the type of user logged; Test with different types of users based on privileges; Each user has privileges based on the range given once created; On the panel of the portal navigate to personal data; Check all the personal data that is stored; Different type of data is stored in personal profile; Check the possibility to change personal data; Test with all personal data fields; Check that personal data is successfully changed; Cancel the user account; Check that all personal data of the user have been removed

ID	4.12

Type	Non-Functional
Category	Privacy, Security
Source	Internal Technical Analysts
Priority	MUST
Description	The DataPorts Platform must support the security and governance recommendations from the IDS reference architecture in order to share and transfer data securely
Rationale	According to the GA (p. 140), the architecture of DataPorts will also follow the approach of the International Data Space (IDS), in the sense that it offers i) endless connectivity, **ii) trust between different security domains, and iii) governance for the data economy** At the same time, it also is aligned with IDSA's mission statement about **secure data exchange**
Responsible Task(s)	T4.1, T4.4
Additional Comments	IDS reference architecture establish several practices to ensure a trustful environment. The DataPorts Platform must ensure their deployment or at least provide an alternative approach
Test Case / Acceptance Criteria	1. Review the existence of a service identity provider 2. If available, review the DAPS, who must ensure that data governance rules are deployed and working 3. Verify, when transferring any type of information, which security protocol is in use and assess its secureness 4. Identify point-to-point encryption (between connectors), using an encrypted tunnel 5. Verify end-to-end authorization (authenticity and authorization based on actual communication endpoints; i.e., data Blockchain connectors) 6. Verify, regarding the trustiness between security domains, PKI structure to ensure identity, by reviewing the security protocols and processes within smart contracts

ID	4.13
Type	Functional
Category	Identity and access management
Source	Internal Technical Analysts
Priority	MUST
Description	Implementation of a solution for IAM, fully aligned with the scope of the DataPorts Platform and future challenges
Rationale	Accounts must be adequately created, and this creation must be performed based on an IAM process
Responsible Task(s)	T4.1, T4.4
Additional Comments	
Test Case / Acceptance Criteria	Once the IAM solution is implemented; Log in to the platform as SA; Go to Manage System -> User Roles; Check if the user has logged in the application as role previously selected (System Administrator)

ID	4.14
Type	Functional
Category	Identity and access management
Source	Internal Technical Analysts
Priority	MUST
Description	An existing account may be deleted, and this deletion must be adequately performed, aligned with the IAM solution
Rationale	In order to avoid non existing accounts in the platform

Responsible Task(s)	T4.1, T4.4
Additional Comments	
Test Case / Acceptance Criteria	Once the IAM solution is implemented; Log in to the platform as SA; Go to Manage User Accounts ->-User roles; Click in account or user to be deleted; Go to delete; Check in list of user access if selected user does not appear anymore

ID	**4.15**
Type	Functional
Category	Identity and access management
Source	Internal Technical Analysts
Priority	SHOULD
Description	The DataPorts Platform should have the capacity to create roles, and (de)assign roles to an existing account
Rationale	Accounts must be adequately created, and role related tasks should be performed based on the IAM process
Responsible Task(s)	T4.1, T4.4
Additional Comments	
Test Case / Acceptance Criteria	Once the IAM solution is implemented; Log in to the platform as SA; Go to Manage User Accounts -> User roles; Click in account or user to be modified; Add roles, or delete them, and submit; Check in list of user access if selected user has new role, or has been deleted; Go to Manage Roles -> User roles;

	Click in new role; Configure the role and submit; Check in list if the new role appears

ID	4.16
Type	Functional
Category	Identity and access management
Source	Internal Technical Analysts
Priority	MUST
Description	The DataPorts Platform must not reveal any sensitive information when a user fails to login into the platform
Rationale	This validates if the platform supports, by textual information or messages, during the login process, reveal non necessary information. The texts must be sanitized, must not include any information about the login error, and must include the links needed to reset or securely remind any login information, along with additional security info methods
Responsible Task(s)	T4.1, T4.4
Additional Comments	
Test Case / Acceptance Criteria	Access to the platform; Insert username or password wrong; Check if "Bad Credentials" message appears, not specifying if username or password is wrong; Click in "Forgot your password" button and check if new password or instructions are submitted

ID	4.17
Type	Functional

Category	Platform security
Source	Internal Technical Analysts
Priority	MUST
Description	The DataPorts Platform must be secured
Rationale	For ensuring this securitization, it must be performed a web application pentest with automated and manual tools to validate the robustness of the application using OWASP controls. In addition, it must be performed a security source code review with automatic tools to validate if secure code development best practices have been considered
Responsible Task(s)	T4.4
Additional Comments	
Test Case / Acceptance Criteria	Run platform pentesting and security analysis process

7.4 WP5 (DEPLOYMENT, PILOTING AND EVALUATION)

ID	**5.1**
Type	Functional
Category	Functionality
Source	Internal Technical Analysts
Priority	MUST
Description	DataPorts must collect metrics about the pilots (requirements satisfied, end-user satisfaction, data collected and processed etc.) to establish a benchmarking of the KPIs achieved by the project
Rationale	This procedure must be implemented in order to support task T5.5 and the Evaluation Plan deliverable
Responsible Task(s)	T5.5
Additional Comments	

Test Case / Acceptance Criteria	**Given** a pilot of the DataPorts project, **when** I require to report the Evaluation Plan, **then** the use cases must provide the established metrics in the Evaluation Plan

ID	**5.2**
Type	Non-Functional
Category	Security
Source	Internal Technical Analysts
Priority	MUST
Description	As a data provider, I want to be ensured for data privacy, so that there is no data leakage
Rationale	Evaluation of secure private data when using IoT devices
Responsible Task(s)	T5.1, T5.3, T5.4
Additional Comments	
Test Case / Acceptance Criteria	The DataPorts Platform ensures that connected IoT devices are not vulnerable malware and DoS/DDoS attack. Data transfer is achieved according to GDPR compliance. It is about ThPA use cases. Use case 1: Data driven application for strategic and real time decisions

ID	**5.3**
Type	Non-Functional
Category	Portability, Interoperability
Source	Internal Technical Analysts
Priority	MUST

Description	As a data prosumer, I want to have interoperable data models for my use case scenarios, so that they can be portable with an interconnected platform
Rationale	Use case scenarios must be useful and efficient and also applicable to more ports
Responsible Task(s)	T5.1, T5.3, T5.4
Additional Comments	
Test Case / Acceptance Criteria	Ensure all the functionalities which will be tested according to the pilots' needs that are suitable for the available infrastructures. In addition, global use case is a PoC to interconnected ports. For example, the DataPorts Platform should be efficient in case it has to be implemented in other ports with similar infrastructures. It is about ThPA use case 1: Data driven application for strategic and real time decisions

ID	**5.4**
Type	Functional
Category	Functionality
Source	Internal Technical Analysts
Priority	SHOULD
Description	As a data prosumer, I want intervention-mediation, in order to solve mobility issues
Rationale	It will be useful that e.g. passenger queues will be implemented by the DataPorts Platform
Responsible Task(s)	T5.1, T5.3, T5.4
Additional Comments	
Test Case / Acceptance Criteria	This is the necessary intervention-mediation for the successful completion of the scenario regarding the mobility queues. More specifically, it is about ThPA use case 2: Improve mobility of passengers, professional and visitors of the port

ID	5.5
Type	Non-Functional
Category	Interoperability
Source	eFTI Regulation proposal
Priority	COULD
Description	The data elements processed correspond to the common eFTI data set and subsets, and can be processed in any of the official languages of the Union
Rationale	COM (2018) 279 final
Responsible Task(s)	T5.2, T5.3, T5.4
Additional Comments	
Test Case / Acceptance Criteria	**Given** the pilots of DataPorts project, **when** involved entities share or subscribe to transport data, **then** the data have to be based on the eFTI standard when applicable

ID	5.6
Type	Non-Functional
Category	Service Level Agreements
Source	Internal Technical Analysts
Priority	SHOULD
Description	DataPorts should involve clients to get real data from different platforms
Rationale	The client is the owner of the data and we need their permission to access it
Responsible Task(s)	T5.2, T5.3, T5.4
Additional Comments	

Test Case / Acceptance Criteria	**Given** the pilots of DataPorts project, **when** I require data to test the functionalities, **then** the data owners should allow the data sharing even if they are not partners

ID	**5.7**
Type	Functional
Category	Functionality
Source	Internal Technical Analysts
Priority	MUST
Description	The DataPorts Platform must provide template agents for the Posidonia Operations data
Rationale	For the integration of Posidonia Operations and DataPorts
Responsible Task(s)	T5.4
Additional Comments	
Test Case / Acceptance Criteria	**Given** the pilots of DataPorts project, **when** a Posidonia Operations application is going to be integrated with the platform, **then** a template agent will be provided

ID	**5.8**
Type	Functional
Category	Functionality
Source	Internal Technical Analysts
Priority	MUST
Description	The DataPorts Platform must provide template agents for the Posidonia Management data

Rationale	For the integration of Posidonia Management and DataPorts
Responsible Task(s)	T5.4
Additional Comments	
Test Case / Acceptance Criteria	**Given** the pilots of DataPorts project, **when** a Posidonia Management application is going to be integrated with the platform, **then** a template agent will be provided

ID	**5.9**
Type	Functional
Category	Functionality
Source	Internal Technical Analysts
Priority	COULD
Description	The DataPorts Platform could provide template agents for the Posidonia PCS data
Rationale	For the integration of Posidonia PCS and DataPorts
Responsible Task(s)	T5.4
Additional Comments	
Test Case / Acceptance Criteria	**Given** the pilots of DataPorts project, **when** a Posidonia PCS application is going to be integrated with the platform, **then** a template agent will be provided

ID	**5.10**
Type	Non-Functional
Category	Performance Efficiency
Source	Internal Technical Analysts

Priority	MUST
Description	As a data prosumer, I want to exploit the capabilities of Big Data analytics
Rationale	The DataPorts Platform must process in a smart way (in combination etc.) data from various databases and provide specific analytics based on ThPA business needs e.g. historical data, patterns of containers pick up/delivery etc. (related to my business) for long and short term decision making
Responsible Task(s)	T5.3
Additional Comments	
Test Case / Acceptance Criteria	**Given** ThPA pilot, **when** data from different databases will be available to the DataPorts Platform, **then** the set of data to be produced will facilitate the long and short term decision making and planning. It can be related to patterns regarding a) the performance of ThPA community (e.g. the pattern that the containers of a specific company/ship leave the terminal) b) the performance of ThPA itself (e.g. the pattern of containers delivery/pick up (quantities, days, months) from a specific company/ship to better plan and manage my yard, my personnel (shifts) and equipment) and c) data from other ports that participate in DataPorts

ID	**5.11**
Type	Non-Functional
Category	Usability
Source	End-Users
Priority	MUST
Description	As terminal operator, I want to have a user friendly Dashboard for the presentation of results
Rationale	The visualization of results must be provided in a user friendly way, with filters etc., in order to be a handy tool

Responsible Task(s)	T5.3
Additional Comments	
Test Case / Acceptance Criteria	**Given** ThPA pilot, **when** the data will be processed and analysed by the DataPorts Platform, **then** it will be presented in a user friendly Dashboard with filters and selection criteria that will facilitate the user to work on. It will be organized by thematic sections for easy access

ID	**5.12**
Type	Functional
Category	Functionality
Source	End-Users
Priority	MUST
Description	As terminal operator, I want to have a model that can predict queues
Rationale	The model will facilitate the optimization of Truck Appointment System, improve the flows inside the terminal and the environmental conditions of the port and its area of influence
Responsible Task(s)	T5.3
Additional Comments	
Test Case / Acceptance Criteria	**Given** ThPA pilot, **when** the prediction model for creation of queues at land gate will be available, **then** the DataPorts Platform will support the optimization of Truck Appointment System by better exploiting the available slots and reducing the waiting time to be served

ID	**5.13**
Type	Functional

Category	Functionality
Source	End-Users
Priority	MUST
Description	As terminal operator, I want to have real time information about the progress of the appointment systems in place
Rationale	The DataPorts Platform must collect in real time status, bookings, preannouncements and after processing to provide the relevant info for reactive measures
Responsible Task(s)	T5.3
Additional Comments	
Test Case / Acceptance Criteria	**Given** ThPA pilot, **when** the DataPorts Platform collects data about land gates operation, **then** will process them and provide real time (or almost real time) information about the status of the land gates

ID	**5.14**
Type	Functional
Category	Functionality, Accountability
Source	Internal Technical Analysts
Priority	MUST
Description	As a system administrator, I want to have access to all logs of the internal platform components
Rationale	The DataPorts Platform must ensure the accountability, by providing a complete logging mechanism for the components deployed in the port ecosystem
Responsible Task(s)	T5.2, T5.3, T5.4
Additional Comments	

Test Case / Acceptance Criteria	**Given** the pilots of DataPorts project, **when** any action is performed by a component, **then** it is recorded in the logs

ID	**5.15**
Type	Non-Functional
Category	Authorization
Source	Internal Technical Analysts
Priority	MUST
Description	As data prosumer, I want to ensure that the data provided to/from the DataPorts will be accessible by only authorized users
Rationale	The DataPorts Platform must provide strong access rights (based on the business rules inside the pilot)
Responsible Task(s)	T5.3
Additional Comments	
Test Case / Acceptance Criteria	**Given** that ThPA will define the rules of "which user has access to what information" for its pilot, **when** the smart contracts will be implemented, **then** ThPA will check and ensure the validity

ID	**5.16**
Type	Functional
Category	Functionality
Source	Internal Technical Analysts
Priority	MUST
Description	As data prosumer, I want to have records of the business activities took place

Rationale	The DataPorts Platform must record the data exchange (e.g. booking request and confirmations)
Responsible Task(s)	T5.3
Additional Comments	
Test Case / Acceptance Criteria	**Given** the Blockchain component, **when** data exchange will take place between collaborators in ThPA pilot, **then** a full record of transactions will be available for business and legal purposes

ID	**5.17**
Type	Non-Functional
Category	Interoperability
Source	Internal Technical Analysts
Priority	MUST
Description	As data provider, I want to be able to easily provide data
Rationale	The DataPorts Platform must have a high degree of interoperability
Responsible Task(s)	T5.3
Additional Comments	
Test Case / Acceptance Criteria	**Given** the interoperability capabilities, **when** data must be provided for ThPA pilot, **then** a template agent will be provided

ID	**5.18**
Type	Non-Functional
Category	Interoperability
Source	Internal Technical Analysts

Priority	MUST
Description	As data consumer, I want to be able to easily get data
Rationale	The DataPorts Platform must have a high degree of interoperability
Responsible Task(s)	T5.3
Additional Comments	
Test Case / Acceptance Criteria	**Given** the interoperability capabilities, **when** data must be provided to ThPA pilot, **then** APIs will be provided

ID	**5.19**
Type	Non-Functional
Category	Interoperability
Source	End-Users
Priority	MUST
Description	As terminal operator, the technical requirements to use the platform must be simple (storage requirements, servers, network etc.)
Rationale	The DataPorts Platform must have simple and at low cost technical infrastructure requirements for the users
Responsible Task(s)	T5.3
Additional Comments	
Test Case / Acceptance Criteria	Easy adaptation of the DataPorts Platform to existing infrastructures

ID	**5.20**
Type	Non-Functional

Category	Maintainability
Source	End-Users
Priority	COULD
Description	As terminal operator, parametrization is important to provide flexibility to users
Rationale	The DataPorts Platform could be flexible to new services and applications
Responsible Task(s)	T5.3
Additional Comments	
Test Case / Acceptance Criteria	Changes and additions to existing use cases and business models should be easily made

ID	**5.21**
Type	Functional
Category	Functionality
Source	Internal Technical Analysts
Priority	SHOULD
Description	As a data consumer, I want to access statistics and predictions that include mobility or visitors
Rationale	In order to get specific result per day/week/month
Responsible Task(s)	T5.1, T5.3, T5.4
Additional Comments	
Test Case / Acceptance Criteria	The required statistics/predictions should be defined for each use-case in particular, and implemented in smart contracts as required