Enabling Industrial Data Space Architecture for Seaport Scenario

David Sarabia-Jácome, Ignacio Lacalle, Carlos E. Palau, Manuel Esteve
Department of Communications
Universitat Politècnica de València
Valencia, Spain
dasaja@upv.es, iglaub@upv.es, cpalau@dcom.upv.es, mesteve@dcom.upv.es

***Abstract*—The evolution of seaports to smart ports is the most important revolution they must face in order to meet the high requirements of efficiency, economy, and security. Achieving this evolution is a task full of challenges. One of these challenges is the interoperability of the stakeholders' information systems. Interoperability has been widely studied but there is still some reluctance to adopt the solutions because the stakeholders are afraid of sharing their data. Recently, the Industrial Data Space (IDS) reference architecture proposes several recommendations and specifications to solve this issue based on data sovereignty concept. This architecture has not been studied nor implemented for the seaport use case yet. For this reason, this paper presents a seaport data space based on IDS architecture to share information in a secure and interoperable manner. Along with this, a big data architecture is added to exploit the shared data. The seaport data space is implemented using the FIWARE IoT platform. The seaport data space is tested by sharing the data of the port authority to the port terminal and analyzing the large shared data through descriptive analysis. The results show that by using the seaport data space, the shared data allow improving the decision making in the planning operations of the seaport.**

***Index Terms*—Interoperability, IDS, Industry, Big Data, seaport, Internet of Things**

I. Introduction

Seaports are a transport and logistics infrastructure that have a significant economic impact on world trade. The transport of goods or raw materials depends on the seaport as a central axis in intercontinental trade. In recent years, European ports traffic has seen an increment along with the operational burden of the seaport. According to the European Union (EU), inwards movements of goods to the main EU ports has increased by 1.4% and outwards movements by 1.3% in the 4th quarter of 2016 [1]. In the coming years, the expected increase in port operations will make it difficult to manage the port's resources and comply with the daily operational load requirements, which will result in unproductive operations. Thus, there is expected the transformation of the seaport following the guidelines of the new industrial revolution, Industry 4.0, in the coming years.

Industry 4.0 is the general concept to refer to the digitalization and automation of industrial processes through the use of emerging technologies such as the Internet of Things (IoT), cloud computing, big data, among others, to improve them and make them more efficient. In a seaport scenario, the synergy of these emerging technologies will help to a seaport become into a smart port or Seaport 4.0. The smart port concept is based on two fundamental pillars: the automation of port operations and equipment, and the interconnection of the key actors involved in the port logistics chain [2]. Currently, the automation of port operations and equipment follows reference industry 4.0 architectures (e.g., RAMI 4.0, OPC UA) to implement Cyber-Physical Systems (CPS) that are employed to monitor and provide information on the machinery of seaports, such as trucks, cranes, among others [3] [4]. However, the imminent heterogeneity of the IoT platforms, suppliers, technologies, and protocols employed by the stakeholders in the development of vertical solutions brings about isolation from the rest of the port operations process.

978-1-5386-4980-0/19/$31.00 ©2019 IEEE

This interoperability issue has been tackled by several industrial consortiums (IIC-IIoT), standards development organizations (IEC, OPC Foundation) and research projects (Inter-IoT) [5] [6]. Some interoperability proposals follow particular requirements in each use case, while some others have to deal with the data ownership issue. As a result, these proposals are difficult to use them in other applications domains, and some case the business is some reluctant to employ them because they are afraid of losing control over the data. This issue has aroused great political concern. Consequently, the data sovereignty concept arises, which is defined as the ability of the data owner to decide itself how to share and use its data.

Recently, the Industrial Data Space (IDS) standard initiative proposes a reference architecture model to cover the industrial requirements of trust, security and data sovereignty, standardized interoperability, value adding, apps, and data markets [7]. By doing so, the IDS facilitates the exchange and linkage of data by creating a trusted virtual data space. The international data space association (IDSA) is in charge of the reference architecture and interfaces design for setting up an international standard. However, the implementation details are out of the scope of the reference architecture. Instead, the IDSA motivates the interaction among researchers and business to actively develop the standard by mean of use cases implementations [8]. In the seaport use case, the IDS architecture has not been implemented or tested yet in the current literature. Also, the integration of big data analytics to the architecture IDS was out of the scope of IDS approach. Extracting valuable information is relevant to the active development of the smart port and should be considered.

This paper presents the implementation details and test of the IDS architecture as a solution to overcome the interoperability among stakeholders and promote a smart port. For this aim, the IDS architecture is implemented employing the FIWARE IoT platform. Also, the paper provides a big data analytics architecture to link to the IDS architecture for exploiting the shared data among stakeholders and bridging the gap among the IDS reference model and the big data analytics. The scenario is tested using the datasets of Valencia port authority and terminal containers. The sharing of data using the seaport data space is employed to extract useful information through a data descriptive analysis for improving decision making. The data analysis results are employed to develop Key Performance Indicators (KPIs) to be shown on a web dashboard. To sum up, the IDS architecture provides a trusted environment to share data that can be transformed into useful information for improving seaport operation efficiency.

This paper is structured as follows. Section II presents an overview of related works and IDS architecture. Section III describes the seaport data space architecture. Section IV describes the Valencia Port use case implementation. Section V presents the results of the big data analysis and its use to exploit the port data spaces. Finally, Section VI presents our conclusions and future improvements.

II. BACKGROUND

In this section, related works are detailed, and the IDS reference architecture is introduced briefly to provide a better understanding.

A. Related Works

In the current literature, there are several proposals to overcome the information sharing issue between the port stakeholders or port community. One of these proposals was the release of the Port Community Systems (PCS) which are information systems to improve coordination of merchandise movements between port stakeholders and improve control over the port route [9]. Along with this, the e-maritime initiative promotes the interoperability of systems in the EU maritime to reduce the complexity of the stakeholders networking to facilitate information sharing. The PCS platform has improved the information interchange and bring a competitive advantage [10]. Another approach is the intelligent management of maritime traffic proposed by the Sea Traffic Management (STM) through the Port Collaborative Decision Making (PortCDM) initiative [11]. In this case, the main goal is cooperation among different information transport systems to share information about the delay in the arrival of a cargo ship. By doing so, the logistics transport company can be informed of the delay, and this information is useful for the best planning of the containers loading and unloading operations. However, these interoperability approaches focused on specific port activities, and it is not well-suitable for business to business (B2B) environment in where the data privacy and security are very relevant in the sharing data process.

The IDS architecture proposes a trusted and interoperable environment in which the owner keep control over how its data is being used. However, IDS architecture requires an interaction between research and business for its development. For example, in [12] the key security requirements and security architecture for the IDS is presented. Also, an ontology-based information model has been proposed by [13] to describe the static and dynamic properties of the IDS entities. Since the IDS architecture does not provide implementation details, in [14] the authors provided relevant details to implement the main roles of the IDS architecture by using the FIWARE IoT platform. The proposed architecture was validated in a real manufacturing use case to improve the maintenance and operation of milling machines. According to [8], many use cases such as railway transport, manufacturing 3D printing, manufacturing, and energy removable business, among others, have been implemented successfully following the IDS architecture recommendation. However, there are not technical information about their implementation nor a seaport data space has been proposed, studied and implemented yet. Also, the IDS architecture did not consider big data analytics to exploit the data in the proposed data space. For these reasons, this paper applies the IDS reference architecture as a blueprint to share information in a secure environment for a seaport scenario, as well as its integration with a big data architecture to extract insight from seaport big data and improve the decision-making.

B. IDS Architecture Overview

The IDSA reference architecture is developed to meet the requirements of trust, security and data sovereignty, the ecosystem of data, standardized interoperability, value-adding apps, and data markets. To do so, the architecture establishes entities and functional roles along its five layers (business, functional, information, process, system). The first layers are considered in this proposal due to they describe the entities and their roles in the IDS architecture, as well as their interactions. The business layer describes the key roles which take part of the data exchange process (data owner, data provider, data consumer, data user) and roles intermediate which are used for administrative and security process (clearing house, identity provider, app store, store vocabulary, broker service provider). These roles are adopted by the entities and are shown in Fig. 1, and described below.

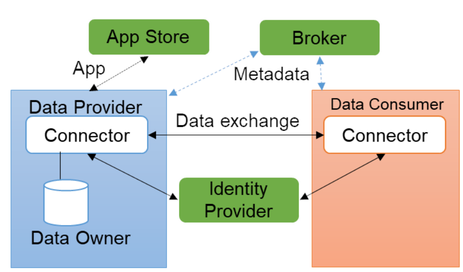

Figure 1. Industrial Data Space Architecture

- *IDS Connector*: the connector is the key entity of this architecture. These are capable of interconnecting among them and exchange data, forming a distributed network. Each connector has a unique identification and is certificated to make a trust ecosystem. The connector is located on the client infrastructure. To exchange data between them, they use a push and pull mechanism. Also, they are capable of executing applications so close to the data source to ensure privacy and keeps control of the use of the data. This kind of applications must be certificated to ensure the data. Since the execution of the application must be monitored to ensure data sovereignty, the IDS connector uses a Policy Enforcement Point PEP proxy.
- *IDS Broker*: This entity is in charge of managing the information of data sources in the space. To do so, the IDS broker stores the metadata, semantics information, pricing or usage policies, and provides an interface to data consumer to look up for available data in the space.
- *Identity Provider*: this entity is capable of providing identity manager service by storing, managing and validating the IDS connector to identify information. This way, the IDS ecosystem provides for trust and security.
- *App Store*: This entity is a repository of services and applications which are used on the data processing in the IDS connector.

The IDS reference architecture states some recommendations to ensure the security environment. For example, authentication and authorization is a key characteristic of IDS architecture. The IDS connector must have a valid X.509 certificate which is important to verify the identity of the participants. Also, the usage policies and usage enforcement are used to keep the data traceability and are defined by the data owner. The policies can include restrictions on transferring and processing data to other parties. More, secure communication among entities is established using encrypted and integrity protected. IDS connectors operators must be identifiable and manageable [7].

III. Seaport Data Space Architecture Overview

The proposed architecture takes advantage of the IDS architecture, which provides trust and an interoperable information sharing environment, to enable a seaport data space. Also, the architecture integrates a big data analytics architecture for exploiting the shared data and extracting valuable insight. The high-level architecture is depicted in Fig. 2.

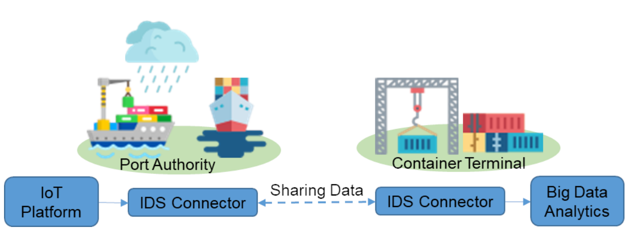

Figure 2. High-level Architecture Overview

A. Big Data Architecture Overview

The goal of this big data architecture is to exploit the data exchanged among IDS connectors. The architecture is composed of 4 modules which are a set of features to provide data store, processing, and visualizing. The architecture overview is described below.

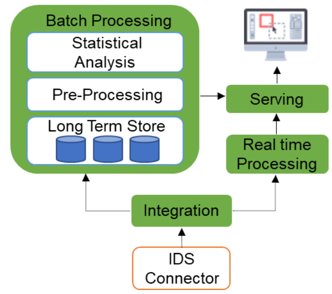

Figure 3. Big Data Analytics Architecture

- *Integration Module*: enables the integration of the big data architecture with the IDS connector. Mainly, the module employs the publish/subscribe IDS connector mechanism to push data into the next modules.
- *Batch Module*: provides two functionalities the long-term data storage and data processing. The raw data is stored using a NoSQL database system due to its benefits to provide data immutability, integrity, scalability, high service available, and high-frequency data writing. Also, the batch module uses the map-reduce paradigm to process large amounts of data efficiently. Also, the batch processing functionality provides high-level libraries for performing a data pre-processing (clean data) stage and processing (extract insight) stage through statistical data analysis. As a result, the descriptive statistics are handled by the serving module to publish in the application user interface (UI).
- *Real-time module*: provides a near real-time data processing to update the batch processing data and reduce the time to takes run again a batch processing job. The real-time module connects to the integration module and uses the merged data by using data aggregation inside a sliding window. Also, this module provides for a real-time platform and libraries to design easily stream pipelines. The results of the real-time processing are handled by the serving module to publish updated information in the UI.
- *Serving module*: provides a publish/subscribe type connection so that the services and applications layer can access the results of the data analysis and present this data to the user.

IV. Valencia Port Use Case Implementation

The port of Valencia was considered as the main Mediterranean port of Spain in large container traffic in 2015, handling 4.7 million movements of Twenty-foot Equivalent Units (TEU). The large container traffic increased by 1.77 million TEU in 2016 [15]. Accordingly, the imminent growth of the

Valencia port traffic requires to be supported by an ecosystem to integrate all the sources of information and managing resources efficiently. Valencia port operations are involved by a set of stakeholders (port authority, terminal container, road hauler companies, shipping lines, importers and exporters, among others), but two actors are considered in this use case (port authority, and the NOATUM terminal container) [6].

The port authority is in charge of arrivals and departures of vessels, train, trucks to and from the seaport. In this case, the port authority provides information on the location of the vessels using the AIS (Automatic Identification System). The port and vessels are exchanging AIS messages which contain information about vessels' name, length, operation type, position, course, and speed, among another navigation status. Similarly, the port authority uses a weather station to sense the weather condition in the port area. The weather data conditions considered in this case were temperature, humidity, and wind speed and direction because they are factors that affect or interrupt normal operations. These sources are compiled into the FIWARE IoT platform following [16]. FIWARE provides Generic Enablers GE which are modular and reusable resources that implement functionalities. The main GE is the Orion context broker which manage the context information lifecycle. To do so, the Orion provides REST API interfaces (NGSI9, and NGSI10) to enable a publish/subscribe communication mode, among the context producers (weather station sensors, vessels) and context consumer (big data analytics).

Meanwhile, the NOATUM terminal container is in charge of load/unload containers from the vessels, train, and trucks. The terminal container provides for information about the time takes to load/unload containers from vessels in each terminal. Also, the terminal container must know the information about the historical vessels traffic to plan its terminal operations. Information such as the vessels traffic seasonality, weekly, or daily is relevant in the development of resources planning to improve the terminal port efficiency. More, the information about the weather and its correlation to the load/unload containers process is important to consider in the efficient process. However, this information is management by the authority traffic and is scarcely shared with the terminal container. These issues are overcome by enabling a data space.

The testbed is implemented in a controlled environment (development environment) to test the IDS architecture for the seaport operations. Fig. 4 depicts the interaction among the entities implemented to enable the seaport data space.

A. *IDS connector implementation*

Two IDS connectors are employed in the proposed use case. The first one is employed onto the port authority organization, and the second one is employed onto the port terminal container organization. Since the port authority is using a FIWARE-based system, we have taken advantage of this fact and the well-developed and diverse GE to implement the IDS connector. Mainly, the Orion Context Broker GE is the core of each IDS connector. The Orion Context Broker provides a publish/subscribe mechanism using the NGSI API

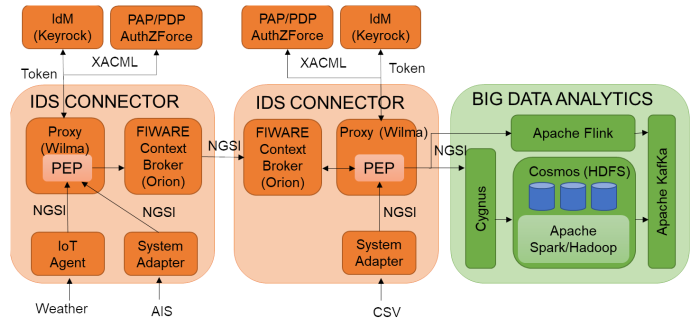

Figure 4. Seaport data space implementation

to exchange data among IDS connectors. The HTTPS NGSI API is activated to provides secure communication among connectors. Moreover, the FIWARE architecture provides an associated information model (entity, attribute, and metadata) that is employed in both connectors to ensure interoperability. Complementarily, the IoT Agents proposed in the FIWARE architecture are used to connect to CSV data from the container terminal. Also, the Wilma GE is used to provide for PEP-proxy functionalities and ensure the usage policies and usage enforcement. The authentication is in charge of the Identity Manager-Keyrock GE. The IdM Keyrock is employed to authenticate users by mean of the Oauth2 protocol and keep track of the IDS connector operators logging. Meanwhile, the Policy Administration and Decision Point-AuthZForce GE are used to enable access control. The PDP/PAP (AuthZForce) GE provides the authorization logic for defining access control policies to applications and services. This GE uses the eXtensible Access Control Markup Language (XACML) to allow the definition of fine-grained policies, and to the request/response interactions during an authorization decision. In this case, the policy implemented allows subscription to receive updates notification from weather and AIS data.

Once the IDS connector is implemented, the sharing process is done by using the Context Broker federation. In this case, the push mode federation is employed to send the *notifyContext* from the port authority IDS connector to the NOATUM terminal container. This is possible by creating a subscription in the port authority IDS connector to send updates to the NOATUM terminal container connector. In this case, the URL information is about the NOATUM IDS connector and it is going to receive notifications about the temperature from the port authority, as Fig. 5 illustrates. By using this mechanism, the weather data and vessels positions are sharing between IDS connectors.

B. *Big Data Architecture Implementation*

The Context Broker only stores the last value, so that persistent storage is required. To overcome this limitation, the Cygnus GE is employed to collect the data from the Orion Context Broker to the long-term storage. The Cygnus is based on Apache Flume and follows the same configuration style to configure the source and sink. In our case, the Hadoop Distributed File System (HDFS) is employed as the sink

```
curl 172.16.1.2:1030/v2/subscriptions -s -S -H 'Content-Type:
application/json' -d @- <<EOF
{
  "subject": {
    "entities": [
      {
        "id": "S17892",
        "type": "sensor"
      }
    ],
    "condition": {
      "attrs": [
        "temperature"
      ]
    }
  },
  "notification": {
    "http": {
      "url": "http://172.18.1.2:1031/v2/op/notify"
    }
  }
}
EOF
```

Figure 5. Subscription to vessel entity in port authority connector

(*cygnus-ngsi.sinks* = *hdfs-sink*) to store the data due to it provides high availability and scalability for data storage. A Parquet file is saved in HDFS for each entity subscribed and updated for each NGSI context notification received. Along with this connector, the Big Data Analysis Cosmos GE is used to provide batch and stream processing. The GE is composed of an Apache Hadoop cluster in which the platform Apache Spark is running using the YARN mode. Apache Spark is selected to run batch processing jobs due to its high-speed in-memory processing performance by using the Resilient Distributed Dataset (RDD). The SparkSQL high-level language provided by Apache Spark is used to perform a descriptive analysis from the data store in the HDFS. Also, the Big Data Analysis Cosmos GE enables Apache Flink processing Engine for real-time processing. The Orion-Flink connector is employed to push data from the Orion Context Broker as a source to the Flink engine. The stream data is used to aggregate the statistical data analysis by using the Flink engine. The processing results are sent to Apache Kafka sink. From the Kafka broker, the descriptive data analysis is pull using the publish/subscribe mechanism to be shown in the terminal operation dashboard. The terminal dashboard is implemented using the Nodejs libraries (express, kafka-node, socket.io).

V. Results

The results show how the architecture takes advantage of the IDS connector to exploit the big data for extracting useful information and improving decision making. Several descriptive data analysis were carried out with three datasets (weather, AIS, and terminal container operations) by using Big Data architecture. The description of the datasets is detailed in Table I. The performance indicators of the port operation are developed through the descriptive analysis of the data.

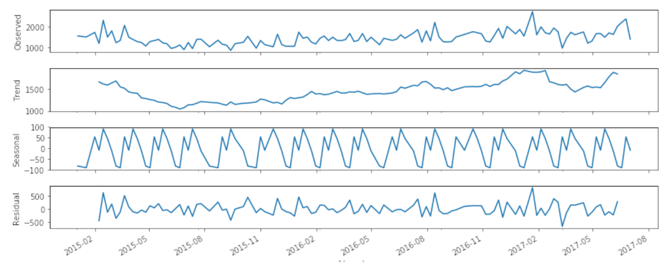

Figure 6. Vessel time per week Seasonality decomposition

This data analysis facilitates the development of relevant KPIs. There are set two KPIs for seaport operations. The first one provides information for the average time spent per vessel in the terminal, while the second one the terminal occupancy.

The trend, season, and residual data characteristics are evaluated from the vessels load/unload dataset, as the Fig. 6 shows. As we can see, the average time per vessel on the terminal container has an upward trend in the last years. This trend is justified with the increment of TEUs that the Valencia seaport has supported recently [15]. Also, there is a regular pattern at an interval of 2 months approximately, as part of a seasonality. The maximum and minimum are supposed to be variable on time because of this seasonality and trend. As a result, the maximum and minimum values are difficult to calculate. To overcome this, the maximum and minimum values are determined by using an average of the minimum values per week as well as an average of the maximum values. These values are considered as thresholds to determine the efficiency in the coming week operations. These KPI values are summarized in the Table II.

For the second KPI, aggregate operations are employed to determine the months and days of the week on which the terminal is the most occupancy. The results are depicted in Fig. 7. August is the month with the most terminal occupancy, supporting each week 35 vessels average. Meanwhile, Saturday is the day with the most terminal occupancy, supporting 6 vessels average. The mean, maximum and minimum values are calculated using the common statistical approach because these time series data are stationary. The results KPI are summarized in Table II. Also, the shared weather information is used to evaluate the correlation factor between the external operations factors values (i.e., temperature, humidity, wind speed, and precipitations) and the terminal load/unload time. To do so, the Pearson correlation is employed. The results suggest that the terminal time load/unload is correlated positively to temperature and humidity, and negatively to wind speed. In

Table I
Seaport datasets summary

Dataset	Size	Period
Weather	28 MB	2014/01/01 - 2017/08/31
AIS	10 GB	2016/01/01 - 2016/03/31
Terminal Operation	52 MB	2014/01/01 - 2017/08/31

Table II
Seaport KPIs summary

Parameters	*KPI1 [hour]*	*KPI2 [vessels]*
Average per week	23,3065	28
Max per week	64,2782	10
Min per week	1.8544	52

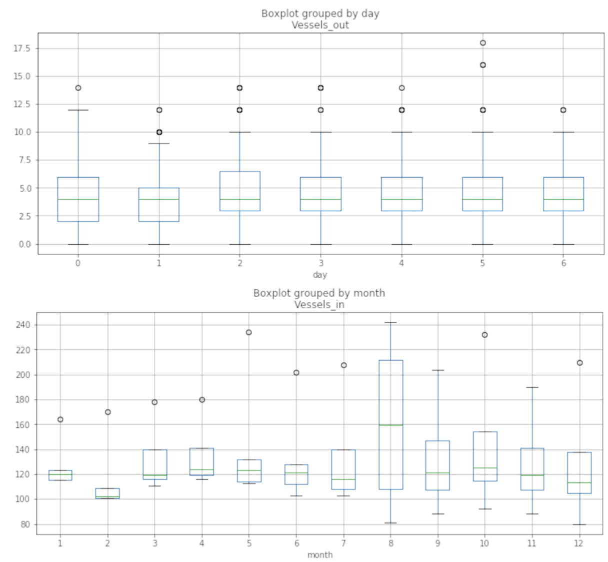

Figure 7. Days of Week and Months of Year Terminal Occupancy

other words, when the temperature and humidity are high, the load/unload time is high too. This is related to workers heat stress, cranes components failure (i.e., cranes designed for -45 °C to 45 °C), among others. This information is useful to implement strategies to plan workers and crane maintenance journals. Finally, the AIS information is employed to provide a context awareness of vessels that are coming to the port terminal. All this information are visualized on the web UI dashboard operation, as the Fig. 8 show.

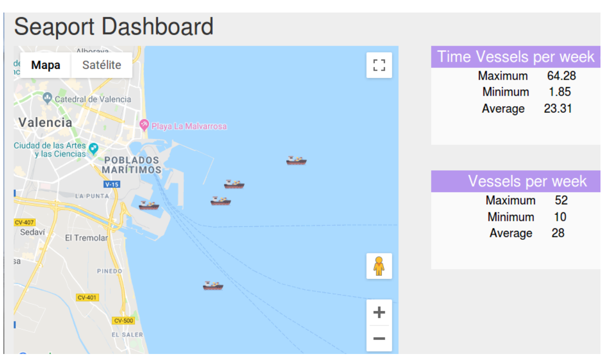

Figure 8. Web UI dashboard seaport terminal operations

VI. Conclusion

This paper has shown a seaport data space to overcome the interoperability among stakeholders' systems. To do so, an IDS-based architecture was implemented by employing an open source IoT platform (FIWARE). Along with this, a big data analytics architecture was proposed to exploit the data shared among data providers. Two KPIs were developed employing descriptive data analysis to evaluate the performance of the port operations. The results of the big data analytics showed the advantage of employing IDS architecture to share data between stakeholders and improve the decision making for increment the seaport operation efficiency. As future improvements, other stakeholders data (hauliers company, and terminals passengers) will be added to the seaport data space and tested employing other KPIs. Also, the rest of the IDS entities that are not considered in this paper will be implemented to enable a complete IDS architecture.

Acknowledgment

This research was supported by the Ecuadorian Government through the Secretary of Higher Education, Science, Technology, and Innovation (SENESCYT) and has received funding from the European Union's "Horizon 2020" research and innovation program as a part of the PIXEL project under Grant Agreement No. 769355.

References

[1] "Publications - eurostats: Maritime trasport of goods quarterly data." [Online]. Available: http://ec.europa.eu/eurostat/statistics-explained/index.php
[2] C. I. Liu, H. Jula, K. Vukadinovic, and P. Ioannou, "Comparing different technologies for containers movement in marine container terminals," in *Intelligent Transportation Systems, 2000. Proceedings. 2000 IEEE*. IEEE, 2000, pp. 488–493.
[3] F. Zezulka, P. Marcon, I. Vesely, and O. Sajdl, "Industry 4.0 an introduction in the phenomenon," *IFAC-PapersOnLine*, vol. 49, no. 25, pp. 8–12, 2016.
[4] M. Weyrich and C. Ebert, "Reference architectures for the internet of things," *IEEE Softw.*, vol. 33, no. 1, pp. 112–116, 2016.
[5] Y. Lu, "Industry 4.0: A survey on technologies, applications and open research issues," *Journal of Industrial Information Integration*, vol. 6, pp. 1–10, 2017.
[6] A. Belsa, D. Sarabia-Jacome, C. E. Palau, and M. Esteve, "Flow-based programming interoperability solution for IoT platform applications," in *2018 IEEE International Conference on Cloud Engineering (IC2E)*, IEEE. IEEE, 2018, pp. 304–309.
[7] B. Otto, S. Lohmann, S. Auer, J. Cirullies, A. Eitel, T. Ernst, C. Haas, M. Huber, J. Jrjens, C. Lange, C. Mader, N. Menz, R. Nagel, H. Pettenpohl, J. Pullmann, C. Quix, J. Schon, D. Schulz, J. Schte, M. Spiekermann, and S. Wenzel, "Reference architecture model for the industrial data space," p. 91, 2017.
[8] "International data spaces association." [Online]. Available: https://www.internationaldataspaces.org/
[9] M.-L. Baron and H. Mathieu, "PCS interoperability in europe: a market for PCS operators?" *The International Journal of Logistics Management*, vol. 24, no. 1, pp. 117–129, may 2013.
[10] V. Carlan, C. Sys, and T. Vanelslander, "How port community systems can contribute to port competitiveness: Developing a cost–benefit framework," *Research in Transportation Business & Management*, vol. 19, pp. 51–64, 2016.
[11] M. Lind, T. Andersen, M. Bergmann, R. T. Watson, S. Haraldson, M. Karlsson, M. Michaelides, J. Gimenez, R. Ward, N. Bjørn-Andersen *et al.*, "The maturity level framework for portcdm," 2018.
[12] G. S. Brost, M. Huber, M. Weiß, M. Protsenko, J. Schte, and S. Wessel, "An ecosystem and iot device architecture for building trust in the industrial data space," in *Proceedings of the 4th ACM Workshop on Cyber-Physical System Security - CPSS '18*, 2018, pp. 39–50.
[13] Á. Alonso, A. Pozo, J. Cantera, F. de la Vega, and J. Hierro, "Industrial data space architecture implementation using FIWARE," *Sensors*, vol. 18, no. 7, p. 2226, 2018.
[14] J. Pullmann, N. Petersen, C. Mader, S. Lohmann, and Z. Kemeny, "Ontology-based information modelling in the industrial data space," in *2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, 2017, pp. 1–8.
[15] "Valencia port report 2016." [Online]. Available: https://www.valenciaport.com/wp-content/uploads/Boletin-EstadisticoDiciembre-2016.pdf
[16] P. Fernández, J. Santana, S. Ortega, A. Trujillo, J. Suárez, C. Domínguez, J. Santana, and A. Sánchez, "Smartport: a platform for sensor data monitoring in a seaport based on fiware," *Sensors*, vol. 16, no. 3, p. 417, 2016.