A Maritime Big Data Framework Integration in a Common Information Sharing Environment

Zdravko Paladin*, Nexhat Kapidani*, Žarko Lukšić*,
Tommaso Nicoletti**, Marios Moutzouris*** , Alexis Blum****, Alkiviadis Astyakopoulos*****
* Administration for Maritime Safety and Port Management, Bar, Montenegro
** Engineering I.I., S.p.A., ENG, Milano, Italy
*** SATWAYS Ltd., Athens, Greece
**** Secrétariat général de la mer - SGMER, Paris, France
***** Center for Security Studies, Ministry of Citizen Protection, Athens, Greece
* {zdravko.paladin, nexhat.kapidani, zarko.luksic}@pomorstvo.me, ** tommaso.nicoletti@eng.it,
*** m.moutzouris@satways.net, **** alexis.blum@pm.gouv.fr, *****a.astyakopoulos@kemea-research.gr

***Abstract* - Ensuring a high level of vessel traffic surveillance and maritime safety is determined by exploiting innovative ICT technologies and international cooperation among maritime authorities. Therefore, initiatives for maritime surveillance, global and regional integrations are realised through a collaborative, cost-effective and interoperable Common Information Sharing Environment (CISE). Consisting of the institutional network of maritime authorities that cooperate on various domains like safety, border control, environmental and rescue missions at sea, CISE enables the efficient transfer and economic exchange of maritime data and information via different interoperable systems using modern digital technologies. The ever-increasing amount of data received from heterogeneous data sources requires specific processing through the adoption of a Big Data framework which hosts, manages and distributes data to maritime users, contributing with great overall benefits to the CISE network core functionality. Specifically, this paper analyses the advantages of the Data Lake infrastructure, including its processes, techniques, tools and applications used to enhance maritime surveillance and safety across the CISE network. This part contains the deployment and interoperability achieved through the components of the participating command and control (C2) systems.Last, as a case study, an overview of the EU project EFFECTOR is provided which aims to demonstrate an end-to-end interoperability framework of data-driven solutions for maritime situational awareness at strategic and tactical operations.**

Keywords - Big Data; CISE; Data Lake; maritime safety

I. INTRODUCTION

The growth of needs for timely operational information, its fast sharing and distribution is commonly joint with an increasing amount of data received from different and heterogeneous sources. This indicates that existing patterns of communication and information-based collaboration among maritime authorities must be improved to meet efficiency and economics requirements. Considering that a high level of data management and control of information flows is provided by modern ICT technologies, the maritime sector gets a great deal of potential for exploitation. It tends to capitalize on applying and combining the latest technological achievements to ensure stable networking among relevant stakeholders [1].

During the last decades, the national maritime authorities have faced lot of challenges due to enormous global, social, economic, and demographic changes, as well as digital transformation with significant technological advancements. To this end, the following essential issues have been arisen:

1. Finding the appropriate method to handle and manage the increased amount of data (generated by intensive activities in the maritime transport industry) and;
2. Increasing the level of international collaboration among maritime authorities, respecting the economic principles and technology standards.

These issues are particularly addressed through the maritime vessel traffic services (VTS) for surveillance and monitoring, as well as through the exchange of navigational data among national, regional and international safety agencies. Its purpose is the readiness for joint operation in maritime safety, border security, search and rescue, environmental and emergency missions. In particular, a large amount of data that cannot be processed and managed with regular devices, and which require greater processing engines is commonly determined under the concept of Big Data. Also, the term "maritime big data" is related to a vast volume of various data received by maritime sensors with specific relevance to maritime authorities/companies ([2], [3]).

Referring to maritime safety and surveillance, the exploited assets provide essential information about ships and vessel types, exact positions, navigational routes, weather etc., by exploiting essential data sources such as: static and dynamic data of Automatic Identification System (AIS), Satellite AIS (S-AIS), Long-range Identification and Tracking (LRIT), coastal radars, CCTV/thermal cameras, electrical and optical devices and weather data, collected from their own or international platforms. Specifically, in order to integrate the above-mentioned datasets and provide a general maritime picture of the area of responsibility, the maritime authorities have introduced the IMO directive for Vessel Traffic Monitoring and Information Systems (VTMIS), designed according to the needs of a particular authority [4]. The VTMIS, being a collector and processor of maritime Big Data, acts as an excellent surveillance legacy system for integration into regional cooperation among maritime and safety agencies.

The paper unfolds as follows: Chapter II elaborates the Common Information Sharing Environment, Chapter III refers to the Big Data Framework and Infrastructure, while Chapter IV gives a short overview of EFFECTOR project case study.

II. Maritime CISE Initiative and Information Exchange

Recognizing the essential importance of cross-sector and cross-border cooperation and information exchange among national authorities (maritime safety agencies, customs, fishery, security, border control and defense), the EU has adopted the initiative to establish the Common Information Sharing Environment (CISE) in October 2010. In order to facilitate the cost-effectiveness and resilience in the collaboration among the involved maritime authorities, it has been emerged a need to develop a decentralized network of maritime civilian-military agencies' collaboration on the basis of: "*need-to-know*" and "*responsibility-to-share*".

The main aim of the CISE is to ensure the proper maritime situational awareness resulting in a fast response of national agencies in case of crisis, threats, emergencies and other risk events at sea. Therefore, the CISE structure comprehends the different information layers, determined by international platforms (SafeSeaNet, Thetis, LRIT) and participating agencies' particular ICT systems/sensors. All these assets are integrated according to the specific CISE Data and Services Model, which can fill existing information gaps across the EU member states, and it is able to provide a Common Operational Picture (COP) to all networked users ([4], [5]). Having defined the operational needs of the User Community, the European Maritime Safety Agency (EMSA), as a governing EU body for CISE, has provided the technical framework for implementing CISE infrastructure at member state level. There are several governance models for participating in CISE stakeholder group [6]:

1. One CISE node – one adaptor,
2. One CISE node – more than one adaptor,
3. One country with more than one CISE node,
4. National node connected to the CISE node.

Specifically, concerning on model 4, as represented in Figure 1 (adapted according the [6],[7],[8]), the CISE technical infrastructure comprises five basic building blocks: the member state Legacy Systems (LS), EU/national/regional CISE Node, CISE Adaptor, CISE Node/Gateway and CISE Network. Other blocks related to Data Lake are developed within EU project EFFECTOR, with more detailed description in the respective Case study chapter. To begin with LS, these ICT systems represent the existing assets used by national maritime authorities, e.g. VTMIS, National Maritime Single Window (NMSW), Unmanned Aerial Vehicles (UAV or UxV), Command, Control and Communication Systems (C2/C3i), able to share the data and information received by maritime surveillance sensors (AIS, radars, VMS, METOC, LRIT, etc.).

In general, to join the national/regional/EU node, a maritime authority needs to link its own LS with the national Node via specific LS-to-CISE adaptor, which translates the data in IVEF or NMEA communication protocols to Node component. Furthermore, the CISE Adaptor translates CISE data and services models to LS specific data format and intermediates between authority and CISE Node/Gateway. Finally, stakeholders are

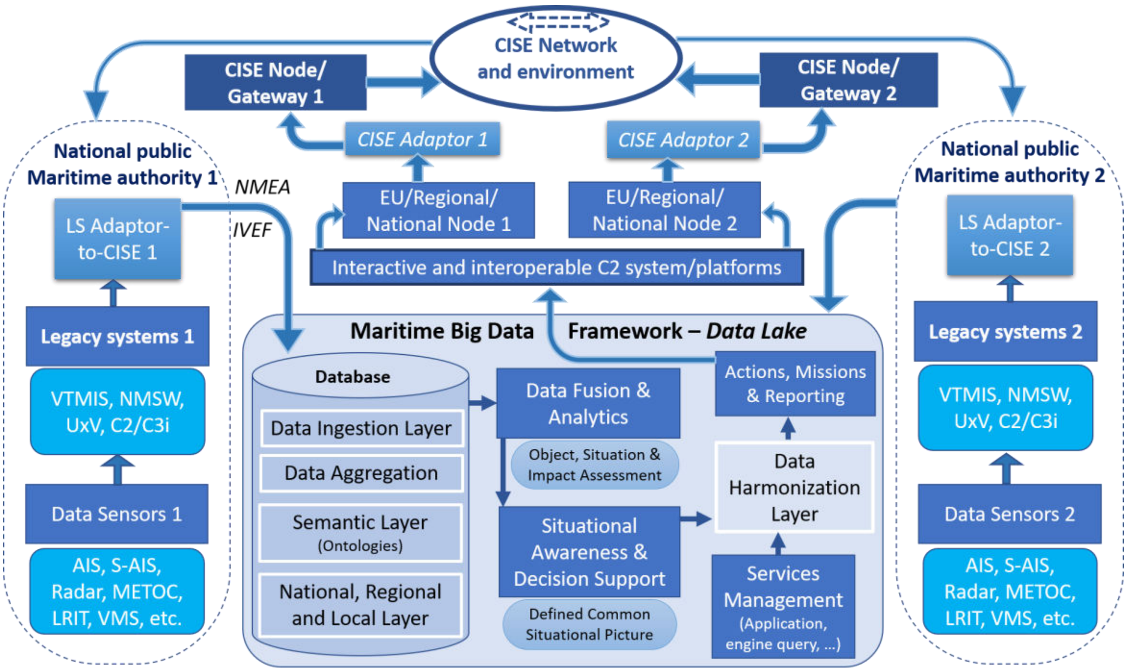

Figure 1. General CISE Hybrid Architecture with integrated Maritime Big Data Framework (authors adapted according the [6],[7],[8])

enabled to access the CISE network (fully secured by VPN and IPSEC) via CISE Node/Gateway core, advanced and common communication services (web and application security, auditing, collaboration and administration console for status tracking). The access is realised by both available entry points of REST (Representational State Transfer) and SOAP (Simple Object Access Protocol) protocols, that operate based on CISE Data and Services Model ([7], [9]).

The data part of this model is the common language for sharing information among sectors and borders. It has defined specific vocabulary that involves seven core entities and relationships among them, such as: *Agent, Object, Location, Document, Event, Risk and Period*. Moreover, the model comprehends eleven auxiliary entities: *Vessel, Cargo, Operational Asset, Person, Organization, Movement, Incident, Anomaly, Action, Unique Identifier* and *Metadata* [10].

The service part of the Model concerns high-level abstraction of specific working capacity, which implements the following communication patterns: pull (request for information provision), multicast pull (for more requests sent to authorities), push (sending the information upon request), subscribe (continual automatic sharing of data in both directions), publish, and acknowledgment (confirmation on receipt of sent/received CISE message) [11]. Specifically, the services are defined by attributes Service ID, Service type, Profile, Capabilities, and the most significantly, all provided services delivered by participating authorities' LS are listed in CISE Service Registry in order to inform all stakeholders about services which are provided by a particular agency to the CISE Network. For instance, mostly exchanged services include collaboration services, vessel details, location and voyage, intervention assets, incident history/event notification, identified risks, cargo and distributed search services. The CISE Data & Services Model is based on XSD (XML Schema Definition) accompanied by UML diagrams (Unified Modelling Language), providing the models for classes and subclasses, related to enumeration of authority and message and service core entities ([11], [12], [13]).

III. BIG DATA FRAMEWORK AND INFRASTRUCTURE

A. *Data Lake Infrasturucture and features*

In last several years, maritime activities and initiatives such as CISE and VDES (Very high frequency Data Exchange System) [4], have been facing the era of Big Data, with the amount of available data that is growing rapidly thanks mainly to the introduction of a collaborative positioning system, namely AIS, and thanks to the widespread of satellites' use. The increase of the availability of data in the maritime sector favors the growth of a technological research area and machine learning, which is particularly dependent on Big Data. The most effective and complete way to organize a data infrastructure in order to manage Big Data is the Data Lake.

In general, the Data Lake is a system capable of storing data at any scale. The data can be stored in raw, semi-structured or structured way, allowing you to perform various tasks such as reporting, data visualization or machine learning. The core idea of a Data Lake is to minimize all input transformations in a "store when data arrives" way, but also the idea about adding a new data source, which should be easy to manage and to expose to the rest of the system making the newcomer data available. These phases are covered in the Data Lake by two main components: data acquisition and aggregation, that guide the flow of data before the archiving phase.

The data ingest layer [14] processes the incoming data, prioritizing sources and validating the data. While, the data extraction can be done in a single large batch or it can be split into several smaller batches. The resulting data will be forwarded to the data module aggregation. Adding a new data source requires traditionally the revision of the whole system, as processes do not easily adapt themselves to the new data sources. This specification is then easily exposed to the rest of the system, making the newly arrived data available. The major advantage of a Data Lake, contrary to databases, is the facility of integration of new data sources as there is hardly any or no transformation of it at all. The drawback associated to this is the difficulty of querying the different data stored.

To support the Big Data technologies, an aggregation layer [15] brings the capability to get a variety of data from very different data sources, it admits analysis of large quantities of heterogeneous data related to maritime surveillance. Furthermore, this huge data, thanks to the aggregation layer will ensure data quality, avoiding data redundancies. After the acquisition phase, the storage phase follows. In Figure 2. the Data Lake storage subsystem could be composed of three modules in which the data can be stored depending on use cases: the Operational Database module, serving transactional operations; the Data Warehouse module, serving analytical operations; the Search Engine module, indexing the data for search operations.

The need of multiple storage system comes with the need of managing data at any scale and it is the same for the Big Data. With sufficient volume, no database can provide sufficient performances in both throughput and latency, leading to the need of multiple databases solutions [16]: online analytical processing database (OLAP) and online transactional processing database (OLTP) sometimes complemented with a hybrid transactional analytical processing database (HTAP) to fill

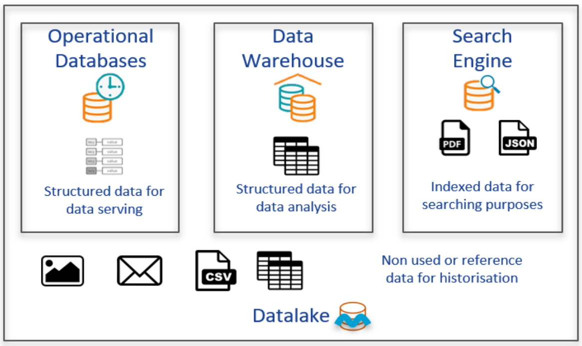

Figure 2. Data Lake storage subsystem (authors' adaptation)

the gap and enable analytics on the fast data. In the maritime domain, datasets traditionally used are well structured. Even semi-structured data sources are far away from being that heterogeneous as Web data would be. This eases the data integration process as datasets are homogeneous. Although the datasets are generally homogeneous, they are actually heterogeneous referring to each other. Multiple data sources produce more consistent, accurate and useful information than any other provided by any individual data source. Thanks to the data fusion and analysis layer, it is possible to integrate the data that comes from multiple data sources continually; to catalogue the available systems, assets and sensors to be used, and finally to identify which of those produce data.

Some of a data provided from data fusion and analysis module are the result of analysis over a raw dataset. In such cases the time needed to analyze the raw data can cause a time delay between the acquisition of the raw data and the availability of the processed data. The aim of this module in the architecture is to send enhanced data to the application layer (interactive and interoperable C2 system). Big Data Analytics characterize processing unstructured information from sensors generated content which can be transformed into valuable border security information, using computational techniques to unveil trends and patterns between datasets. The Big Data emphasizes application and the adoption of new technology, for instance using of past/historical data to predict future trends [17]. This analysis helps to identify patterns and learn from historical data.

B. *Support of Data Lake for enhancing Command and Control capabilities*

Operators of Command-and-Control systems require easy and quick access to data from their sensors, surveillance systems, external third-party databases and operational information created during their operation [18]. As technology evolves, new sensors along with their protocols are available to enhance the maritime picture of an agency. Additionally, users want to be able to keep historical data and to query this data when needed. Managing these needs and catering for future expansion of storage and new data types are the requirements that can be successfully met by the implementation of a multi-layered Data Lake technologies, as described in the previous section. Data Lakes can provide an infinite pool of raw and processed information, enabling operators of C2s to obtain quick access to current information and to perform extensive analysis on historical information. A typical deployment that solves managing and accessibility is depicted in the diagram, Figure 3. Current information or near real-time information is needed to visualize where vessels are currently located, nowcast of weather, current ownership structures, vessel images and other information that might be needed by C2 operators. This maritime data value chain produces a vast data and it is of on-demand nature ([19], [20]). Data Lake technology can enable batch ingestion, dynamic ingestion, on-the-fly data processing, ontology management and querying of data beyond traditional relational data stores. Storage can be added easily to a Data Lake as data increases without any impact to accessibility (ease and speed).

Segmentation of a Data Lake into tactical and strategic levels ensures that the command structures have access to data in their area of responsibility and that data can be propagated to the right level when requested and based on the configured criteria. Incidents can be propagated up to higher levels, while incoming data from the CISE network can be pushed down to the specific department, whose responsibility is a certain area where e.g., the vessel is located or information can be used operationally (documents, risks, etc.). The segmentation allows for deployments to be distributed and not centralized. Each regional command structure or even a local command structure can have its own installation and information is moved automatically or by a user-action. If connectivity is broken, information is queued and once connectivity is restored, this queued information is propagated.

As time passes, historical information is accrued from sensors, external databases and from the day-to-day operation of a maritime agency. Access to this information is needed for various reasons. Understanding the patterns of life of vessels enables operators or third-party services to make decisions on whether the vessel is a risk or behaving normally. Services that use machine learning can constantly train their algorithms, as more historical data is available allowing for better decision support. Data Lakes provide the required capability of scaling using Big Data technologies and storing even more heterogenous data. Operators can query historical operational information for training purposes or for determining the best measures that were taken to combat the situation using available assets. It is not enough just to have access, but access in a timely manner. Responses to querying historical data sets might take seconds or minutes using relational storage, but it is with non-relational document storage and with proper indexing when this response can be under a second. This allows the C2 operators to explore historical data and not to be forced to look up for something specific as it would have taken longer [21].

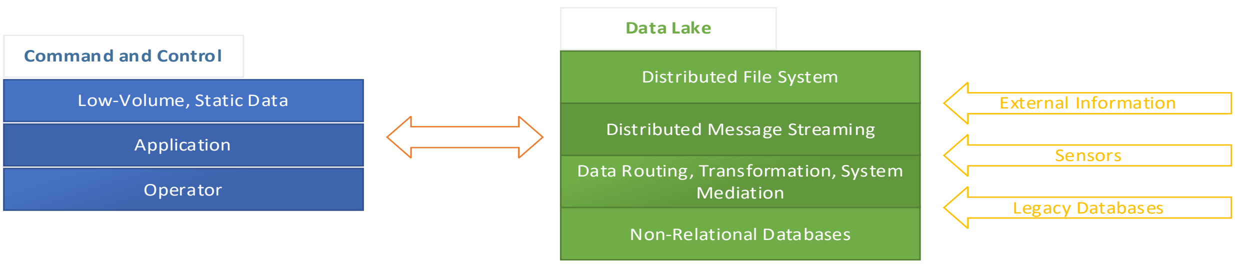

Figure 3. Typical deployment of Command and Control systems and Data Lake interaction (authors' adaptation)

The added value of a semantically enabled Data Lake allows the operator to infer new insights into existing data that otherwise lays hidden in the various data stores of agency. Information coming into the Data Lake can be converted into a semantic stream inferring quickly certain patterns of interest to the operator in a visual manner (e.g., heat maps). Semantics also have the added advantage of providing one "format" to Data Fusion services irrespective of the source enabling these services to be quickly integrated into the ecosystem of services inside a maritime agency.

Additionally, to gain the benefits described above, the command and control's operational methodology should be bound to the Data Lake and vice versa. Otherwise, it might not live up to its full potential in generating incremental operational value [22]. There must be taken care in order not to become a dumping ground for data, but rather curated, auditable and entity-aware. The Multi-layered Data Lake deployed as part of an EU project described in the next section has these in its core design principles.

IV. Case study: EU Project EFFECTOR

Following the goals of the European Commission for achieving maritime CISE network's full operability and building on outputs of previous projects related to surveillance and CISE topic, the project EFFECTOR (*an End to end interoperability Framework For maritimE situational awareness at strategiC and Tactical OpeRations*) is focused on exploitation of significant potentials of maritime surveillance assets and systems with enhanced interoperability framework implementation. The project assumes deployment of multilayered data lake platform with incorporated Big Data analytics, techniques and tools. Its aim is to improve decision support and information sharing process in order to reach high interoperability between maritime information systems (at national level, or cross-border) and cooperation among partners at strategic and tactical level [8].

Using the maritime surveillance assets and C2 systems (SeaMIS, ENGAGE and MUSCA) of the end users involved in the project and deploying Data Lakes, ontology design and data alignment, the project EFFECTOR provides improved solutions for data fusion and analytics, maritime anomaly - suspicious vessel behavior detection, and semantic trajectories fusion, as well. The set of software tools are also supported by AI modules (for alert and anomaly detection), Uniform Communication Gateway, as well as Early Collision Notification system based on fuzzy logic. The project gathers relevant partners from the IT industry, academia and end-user community of maritime authorities, which jointly work on provision of enhanced collaborative framework for maritime surveillance and control. Three national Data Lakes will be installed, tested and validated through 3 operational trials (respectively run in Portugal, France and Greece) covering the large spectrum of coast-guards' missions. Among the seven end-user participants spanning from Portugal and Bulgaria (and including European navies, coast-guards or police corps), Administration for Maritime Safety and Port Management of Montenegro (AMSPM), being the EU candidate country, contributes with provision of necessary AIS information for implementation of operational scenario related to the French Maritime Trial, shared in specific format respecting CISE Data and Service Model [8].

The CISE hybrid architecture, referred to Figure 1, assumes the specific interface between participating authorities substantiated in the component of mentioned interactive and interoperable C2 platforms for execution of actions and services and information exchange according to the provided entities in CISE Data Model. These C2 software provide full COP visualization, the entire control over actions taken by users, and services exchanged in joint missions based on communication patterns, as designed in CISE architecture. Moreover, in order to process a huge amount of data received by maritime authorities' sensors, it was necessary to involve the concept of Maritime Big Data Framework, in the form of Data Lake, which covers all necessary operations to enable the relevant data flow into C2. Through this interface the data are further transferred to EU/regional/national CISE Node, then to CISE adaptor, and over CISE Node/Gateway to CISE Network and Environment.

To be more specific, the repository databases of Data Lake, previously filled with data in raw formats, e.g. CSV, JSON, or RDF received from national maritime authorities' LSs over specific adaptors, consist of several operational layers which process the ingested data transferred over IVEF/NMEA protocols. After structuring and refining the data, the capabilities such as Data Fusion & Big Data Analytics, Situational Awareness & Decision Support Tools involving some artificial intelligence engines with Service Management, transfer the data to harmonization layer, which delivers the information in standardized format to the mission and reporting entity. Finally, the C2 platform receives processed and structured information as an output of Big Data Framework – data lake, in the form of straightforward actions, services, missions and reports, which are then transferred to other agencies in CISE Network, as described.

In detail, data is ingested from different data sources and communication protocols into the Data Lake, transformed, filtered, stored and routed where necessary to the various operational layers of involved maritime agencies (e.g., national, regional, local). Any associated metadata is also propagated throughout the various processes to provide context to the data item. This allows for data to be stored in its raw format and in other formats. Many third-party services use different data formats (e.g., AIS or IVEF or ASTERIX etc.) and by providing these various formats this allows for these services to be integrated easily without any further development.

By combining, processing and analyzing data messages from different sensors, the project EFFECTOR systems deal with a huge amount of streaming data, which could contain noise data, irregular time sampling or other kind of lacks that recurs in a specific data domain. The relevant data analysis helps in this purpose, and supports the final goal of enabling the end-user human operator to have an intelligent support for decision making. To extract relevant information from the data gathered, there are considered two types of data analysis: static that provides insight, and serve as a decision-making guide; and dynamic that enables to use data as a past, predictive, and live decision-making resource.

V. CONCLUSION

The paper provides an analysis of current collaborative environment for exchange of relevant maritime information based on CISE concept, according the communication patterns and services designed under CISE Data and Services Model. The highly developed interoperable platform, in the form of multi-layered Data Lake, that has capacity to fuse, structure, manage and store vast amount of data received by heterogeneous maritime sensors is here presented as an essential part of CISE network. For the purpose of providing greater decision support for maritime safety/security agencies, Data Lake deploys processing layers with big data techniques for analysis of vessel movements, behavior, risks, various maritime activities and other events at sea. Following the provided system design, the Big Data component and its interaction with C2 capabilities, represent key feature which enables higher level of situational awareness at sea. As a case study, it is presented the EU project EFFECTOR, that aims to further deploy interoperability framework of national agencies competent for maritime surveillance and foster collaboration among them on tactical and strategic level. Therefore, the integration of Maritime Big Data applications into information-sharing process significantly contributes to enhance CISE capabilities bringing many benefits to collaboration among member states. Finally, they are reflected in a higher quality of maritime COP, detailed coverage of monitoring areas, efficient and economic process of information distribution among stakeholders, more comprehensive and quicker responsiveness in surveillance and operations at sea, ensuring safer and more secured EU maritime domain.

ACKNOWLEDGEMENT

This work has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 883374 (project EFFECTOR). This article reflects only the authors' views and the Research Executive Agency (REA) is not responsible for any use that may be made of the information it contains.

REFERENCES

[1] N. Kapidani, S. Bauk, I.E. Davidson, "Digitalization in Developing Maritime Business Environments towards Ensuring Sustainability", *Sustainability* 2020, Vol. 12, pp. 9235, Published: 6th November 2020.

[2] M. Jović, E. Tijan, R. Marx, B. Gebhard, "Big Data Management in Maritime Transport", *Pomorski zbornik* Vol. 57, pp. 123-141., Croatia, 2019.

[3] E. Tijan, M. Jović, S. Aksentijević, A. Pucihar, "Digital transformation in the maritime transport sector", in *Technological Forecasting & Social Change*, Vol. 170, pp.1-15., 2021.

[4] B. Garnier, A. Napoli, "Exploiting the Potential of the Future "Maritime Big Data" ". Maritime Knowledge Discovery and Anomaly Detection Workshop, July 2016, Ispra, Italy. pp.24-27.

[5] Integrating maritime surveillance in Europe: Common information sharing environment (CISE), draft roadmap, 20th October, 2010. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52010DC0584&from=HU

[6] European Martime Safety Agency (EMSA): Practical Guide for joining the CISE network, v2.0, 13th January 2022. http://www.emsa.europa.eu/publications/leaflets/item/4639-practical-guide-for-joining-the-cise-network.html

[7] European Maritime Safety Agency (EMSA): CISE Technical Documentation, http://emsa.europa.eu/cise-documentation/

[8] EFFECTOR EU Project - "An End to end Interoperability Framework For Maritime Situational Awareness at Strategic and TacTical OpeRations", Grant agreement ID: 883374, 2020-2022, https://www.effector-project.eu/

[9] Z. Paladin, A. Mihailović, N. Kapidani, D. Merino, J.M. Grenner, G. Vella, M. Moutzouris, R. Leuzzi, "Augmenting maritime Command and Control over a regional Common Information Sharing Environment implementation: Montenegro Case", *NMIOTC Journal*, Vol. 22 No 1, pp. 20-29, Greece 2021.

[10] M. Riga, E. Kontopoulos, K. Ioannidis, S. Kintzios, S. Vrochidis, I. Kompatsiaris, "EUCISE-OWL: An Ontology-based Representation of the Common Information Sharing Environment (CISE) for the Maritime Domain", *Semantic Web*, vol. 12, no. 4, pp. 603-615, 2021.

[11] D. Berger, J. Hermida, F. Oliveri, G. Pace, "The Entity Service Model for CISE -Service Model Specifications", Technical Report, Joint Research Centre of the European Commission, 2017, online: https://webgate.ec.europa.eu/maritimeforum/en/node/4039

[12] A. Mihailović, N. Kapidani, E. Kočan, Ž. Lukšić, D. Merino, S. Antonopoulos, M. Moutzouris, "A Framework for Incorporating a National Maritime Surveillance System into the European Common Information Sharing Environment", The *25th International Conference on Information Technology* (IT), 2021, p.1-6. (a)

[13] A. Mihailović, N. Kapidani, E. Kočan, D. Merino, J. Rasanen, "Analysing the prospect of the maritime common information sharing environment's implementation and feasibility in Montenegro", in *Scientific Journal of Maritime Research*, No. 35, pp. 256-266., 2021. (b)

[14] T. Hlupić, J. Puniš, "An Overview of Current Trends in Data Ingestion and Integration", *The 44th International Convention on Information, Communication and Electronic Technology (MIPRO),* 2021. IEEE.

[15] D.P. Mandić, D. Obradović, A. Kuh, T. Adali, U. Trutschell, M. Golz, Ph. De Wilde, J. Barria, A. Constantinides, J. Chambers, "Data fusion for modern engineering applications: An overview", *International Conference on Artificial Neural Networks*. Springer, Berlin, Heidelberg, 2005.

[16] P. Mangaiyarkkarasi, "On Line Analytical Processing–An Overview", *International Journal For Research And Development In Technology* , 6 (2), September, 2016.

[17] J. Zakir, T. Seymour, and K. Berg, "Big Data Analytics." *Issues in Information Systems* 16 (2), 2015.

[18] J. Zhang, G. Wang, W. Gang, "Command and Control System Construction in Big Data Era", in *Journal of Physics*: Conference Series, 2019.

[19] J.M. Cavanillas, E. Curry, & W. Wahlster, "New horizons for a data-driven economy: A roadmap for usage and exploitation of big data in Europe". Berlin: Springer, 2016.

[20] J. Ferreira, C. Agostinho, R. Lopes, K. Chatzikokolakis, D. Zissis, M.-E., Vidal, "Maritime data technology landscape and value chain exploiting oceans of data for maritime applications", in *Proceedings of 2017 International Conference on Engineering, Technology and Innovation (ICE/ITMC)* pp. 1113–1122.

[21] F. Huang, "Managing Data Lakes in Big Data Era: What's a data lake and why has it became popular in data management ecosystem", *The 5th Annual IEEE International Conference on Cyber Technology in Automation, Control and Intelligent Systems*, June 8-12, 2015, Shenyang, China.

[22] S. Kumar, "Four Common Mistakes That Can Make For A Toxic Data Lake", Forbes, Nov 25, 2014, Retrieved Feb 4, 2022: https://www.forbes.com/sites/ciocentral/2014/11/25/four-common-mistakes-that-make-for-toxic-data-lakes/