Multidisciplinary
SCIENTIFIC JOURNAL
OF MARITIME RESEARCH

University of Rijeka
FACULTY OF MARITIME STUDIES

Multidisciplinarni
znanstveni časopis
POMORSTVO

https://doi.org/10.31217/p.36.2.13

Maritime information sharing environment deployment using the advanced multilayered Data Lake capabilities: EFFECTOR project case study

Zdravko Paladin[1], Nexhat Kapidani[1], Piero Scrima[2], Georgios Vosinakis[3], Guillaume Hajduch[4], Marios Moutzouris[5], Christos Bolakis[6], Alkiviadis Astyakopoulos[6]

[1] Administration for Maritime Safety and Port Management in Montenegro, Street Maršal Tito No. 7, PO Box 14, 85000 Bar, Montenegro, e-mail: zdravko.paladin@pomorstvo.me; nexhat.kapidani@pomorstvo.me
[2] Engineering I.I., S.p.A., ENG, Palazzo D1/3, Centro Direzionale Milanofiori, Strada 2, 20090 Assago, Milano, Italy, e-mail: piero.scrima@eng.it
[3] Institute of Communication and Computer Systems (ICCS), No. 9, Iroon Politechniou Str. Zografou, Athens, Greece, e-mail: giorgos.vosinakis@iccs.gr
[4] Collecte Localisation Satellites (CLS), Bâtiment Le Ponant, avenue La Pérouse 29 280 Plouzané, Brest, France, e-mail: ghajduch@groupcls.com
[5] SATWAYS Ltd., 15 Megalou Konstantinou Street, Irakleio, Attica, 14122, Athens, Greece, e-mail: m.moutzouris@satways.net
[6] Center of Security Studies (KEMEA), Ministry of Citizen Protection, P. Kanellopoulou 4, 101 77, Athens, Greece, e-mail: c.bolakis@kemea-research.gr; a.astyakopoulos@kemea-research.gr

ABSTRACT

Establishing an efficient information sharing network among national agencies in maritime domain is of essential importance in enhancing the operational performance, increasing the situational awareness and enabling interoperability among all involved maritime surveillance assets. Based on various data-driven technologies and sources, the EU initiative of Common Information Sharing Environment (CISE), enables the networked participants to timely exchange information concerning vessel traffic, joint SAR & operational missions, emergency situations and other events at sea. In order to host and process vast amounts of vessels and related maritime data consumed from heterogeneous sources (e.g. SAT-AIS, UAV, radar, METOC), the deployment of big data repositories in the form of Data Lakes is of great added value. The different layers in the Data Lakes with capabilities for aggregating, fusing, routing and harmonizing data are assisted by decision support tools with combined reasoning modules with semantics aiming at providing a more accurate Common Operational Picture (COP) among maritime agencies. Based on these technologies, the aim of this paper is to present an end-to-end interoperability framework for maritime situational awareness in strategic and tactical operations at sea, developed in EFFECTOR EU-funded project, focusing on the multilayered Data Lake capabilities. Specifically, a case study presents the important sources and processing blocks, such as the SAT-AIS, CMEMS, UAV components, enabling maritime information exchange in CISE format and communication patterns. Finally, the technical solution is validated in the project's recently implemented maritime operational trials and the respective results are documented.

ARTICLE INFO

Original scientific paper
Received 14 September 2022
Accepted 7 December 2022

Key words:
CISE
Maritime surveillance & safety
Big Data
Data Lake
New technologies

1 Introduction

The exponential increase in versatility of data sources, datasets, surveillance assets, devices, has been followed by the growth of compliant technologies on high-level readiness for data processing, management, and sharing. This trend has been recognized and is of special importance in the maritime domain. The various challenges affecting its regular performance imposed the need for deployment of advanced technologies in each sphere of the maritime segment. Therefore, the networked approach comprehending relevant maritime stakeholders together with an effective system of new ICT technologies, brings many benefits to the broader community of participants, including a safe, secure and resilient framework for maritime activities. These activities need to be persistently monitored and controlled in order to face various challenging situations, the most frequent among which are the intensity of maritime traffic, vessel collisions in coastal areas, border control, environmental risks (oil spillage), law enforcement, irregular migrations, border- crossing, smuggling, illegal fishing, etc. Therefore, seamless and cost-effective collaboration between maritime authorities based on timely information collected from national legacy systems led to the establishment of a cooperative environment for maritime information and data exchange, known as the Common Information Sharing Environment (CISE). The infrastructure for the data sharing process for the purpose of raising the Maritime Situational Awareness (MSA) and creating the Common Operational Picture (COP), comprehends the following important elements:

- own data sources, assets and devices in the form of sensors (e.g. AIS, radar, UAV, meteo data, etc.),
- the integration platforms for all collected data from sensors and configuration of a system that is capable of processing, managing and coordinating of relevant actions (e.g. VTMIS, NMSW, PCS, etc.)
- Operationalized Big Data repositories – Data Lakes for storing huge volumes of maritime data,
- Advanced tools for Big Data Analytics with Artificial Intelligence (AI) Modules for decision support to national safety centers in missions at sea over Command, Control, and Coordination systems (C2 or C3i).

The aim and motivation of the paper are to demonstrate the key features of Data Lakes as Big Data repositories that together with integrated AI modules provide key important support to a more efficient CISE Network of maritime participants. The most important features of the Data Lakes and technologies discussed in the paper, contribute to adopting a consistent model for an integrated maritime surveillance framework. Also, extensive research was conducted within the EU Project EFFECTOR, leading to results and conclusions regarding the interoperable network of maritime agencies, supported with novel techniques in maritime surveillance and safety domain.

The structure of the paper is organized as follows. After the introductory part, Section 2 with the methodological approach, gives a broader insight into a maritime network for sharing relevant data and information through CISE, as well as the architecture of the data lake as a system component that collects and stores large data with the purpose of supporting a data and services sharing model. Section 3 gives a case study showing the most important features and plugins of multilayered Data Lake, deployed within the EU Research and Innovation project EFFECTOR. Section 4 provides the analysis of evaluation and validation surveys done with the aim to estimate the satisfaction of maritime end-users with developed technical solutions after testing and live demonstration of operational trials within EFFECTOR Project. The concluding remarks are given in Section 5.

2 Methodology

The proposed methodological framework derives from previously conducted research on advancing and optimizing information sharing processes in the maritime domain, mostly concerning the EU CISE Initiative and developing disruptive technologies in the field of Big Data, Analytics and AI, that support modern business processes and data management ([1], [2], [3]). For this purpose, the overall architecture of the framework is composed of three structural aspects: CISE Model for international maritime collaborations, Big Data Infrastructure for hosting, storing, distribution, and analytics of large data sets, and comprehensive Data Lake architecture with intelligent layers for data processing, querying and retrieval of relevant information to support data exchanges between maritime authorities within CISE Network, as illustrated on Figure 1.

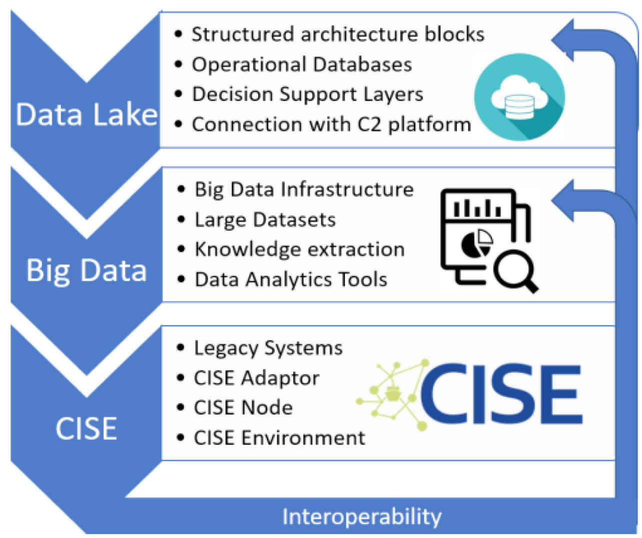


Figure 1 Proposed methodological approach

Source: Authors

2.1 The CISE Network and Maritime Surveillance systems

Becoming aware of the necessity to establish a more resilient and cost-effective solution for communication and interaction between maritime authorities that cooperate in various situations at sea, the European Commission (DG MARE) set the Roadmap for the introduction of a framework for collaboration among maritime stakeholders in EU (COM (2009) 538). The result is an initiative for creating the Common Information Sharing Environment (CISE) which was adopted in 2010 and is expected to be fully operational by 2023. The core idea of the CISE concept lies in a network of voluntarily interconnected maritime authorities, which in a more efficient, secure, and faster way collaborate during emergencies, maritime safety and security operations, and other activities at sea, providing timely and fast information and services to all participating partners based on a specifically designed data and services model. It is important to mention that CISE is not a system but rather a network of interconnected existing legacy systems (LS), data sources, and operating platforms for communications between partners in the maritime domain. The EU institution in charge of the CISE Model development and implementation of all provisions, infrastructure, and maintenance is the European Maritime Safety Agency (EMSA). In order to support the proposed way of collaboration, EMSA established the CISE Stakeholder Group (CSG) assembling the participating countries – Member States in maritime data and information sharing. Also, keeping in mind the specifics of each Member State, several different models of collaboration and governance were proposed [4]:

- One CISE node – one adaptor (M1),
- One CISE node – more than one adaptor (M2),
- One country with more than one CISE node (M3),
- National node connected to the CISE node (M4).

Figure 2 provides the layout of CISE Basic Architecture and models of CISE Collaborations and governance between involved public authorities (PA).

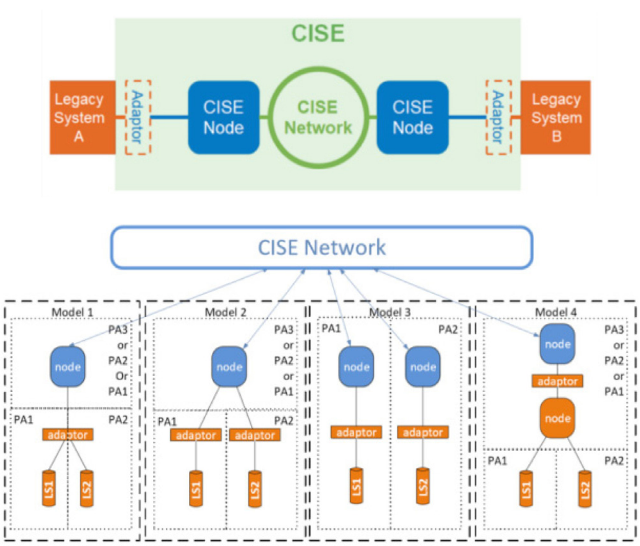


Figure 2 CISE Basic Architecture and governance models

Source: [4]

Regarding the technological specifics of CISE, we will focus on two main aspects, the main features of CISE building blocks and CISE Messages exchange patterns. As depicted in Figure 2, the CISE infrastructure requires the following main components: A legacy system of maritime authority (LS), a CISE adaptor for connection with other blocks, EU, national or regional Node, as well as the main CISE Node/Gateway for accessing the CISE network, and CISE Environment comprising all Member States. It is important to mention that LS is an existing operating ICT system managed by maritime agencies, which collects data, visualizes, and shares them in a specific format, being mostly connected with sensor devices and assets for any type of maritime data. These LSs generally concern Vessel Traffic Monitoring and Information System (VTMIS), National Maritime Single Window (NMSW), various UxVs, Command, Control, and/or Coordination and Information Platforms (C2/C3i), and Electro-optical devices (E/O), while the data sensors most frequently provide maritime data such as Automatic Identification System (AIS in the form of static, dynamic, voyage-, and safety-related data), MARES-AIS, Satellite AIS (SAT-AIS), Long Range Identification and Tracking (LRIT), SafeSeaNet (SSN) and CleanSeaNet (CSN), Meteorological and oceanographic data (METOC), Vessel Monitoring System for fisheries (VMS), VHF direction finder, coastal radar data, (thermal/optical) camera recordings, microwave links, EPIRB Detections, etc. These data are commonly distributed over communication protocols and formats like International VTS Exchange Format (IVEF), National Maritime Engineers Association (NMEA), Extended Markup Language (XML), HTTP, TCP/IP, VoIP, ASTERIX, NetCDF, GeoJSON, OGC standards, STANAG 4586, MAVLink, ONVIF, etc., which are all secured with VPN communication access.

The CISE adaptor is the component for translating the relevant data from LS in continual or on-demand mode to CISE Node, which are further consumed by other networked partners that require a piece of specific information. Having in mind that different LSs provide various types and formats of data, the CISE Adaptor translates the received data to CISE format and enables a unique and uniform pattern of sharing the information in the appropriate form for all participants. The main technical components of Adaptor are the CISE Service and a CISE client (for communication with the CISE Node), and an LS Custom Service and LS Custom client (for communication with the LS) [5].

Finally, the CISE Node/Gateway is organized as a node of regional, local, and EU-wide components for connecting LSs with other elements of the CISE Network, over a specific CISE Adaptor. The CISE Node, according to the governance models presented, is a common software that can be managed and hosted by Public Authority, which handles the routing between legacy systems and gives access to the European/Regional/National node. The CISE Node/Gateway provides Core, Common, and Advanced Services for data sharing and interaction between partners involved in the CISE network.

For the communication patterns between partners, several functionalities are available such as pull (CISE consumer requests a piece or of information from a CISE provider), multicast pull (a CISE consumer requests information to a group of CISE participants), push (a CISE provider sends a piece of information to CISE consumer using Push operation and message), multicast push (a CISE provider sends a piece of information to a group of CISE consumers), publish/subscribe (a CISE Consumer sends a request for the continual flow of information from other partners by subscribing for the particular type of information on CISE provider services), discovery service for pull/push unknown (using a registry of services, a CISE consumer is able to request/push information to one or several providers without knowing in advance which one can answer his request) [6].

The services the original CISE Data and Services Model provides represent the most important and relevant information such as: Collaboration services, Vessel details service, Vessel Voyage services, Intervention asset service, Incident/event notification (alert) service, Risk information service, Cargo service, Vessel location service, Incident history service, Registry of Authorities (Query), Distributed search.

The CISE Data and Services Model is determined by seven Core Vocabulary entities (Agent, Object, Location, Document, Event, Risk and Period) with auxiliary ones (Vessel, Cargo, Operational Asset, Person, Organization, Movement, Incident, Anomaly, Action, Unique Identifier and Metadata). The model is developed upon XSD (XML Schema Definition) which uses the UML diagrams (Unified Modelling Language) containing the models for Authority, Message and Service Core Entities, which are divided into enumerations and classes [6]. Specifically, the enhanced CISE Model (eCISE v.2.2.0) [7] was developed within Project ANDROMEDA (EU H2020) and extends the CISE features on the land domain adding some classes to specific events and categories related to land surveillance situations and activities ([8], [9]).

2.2 Data Lake Architecture

The maritime domain usually exploits well-structured data, using shared formats and standards. In any case, the standards and formats used are not few, and despite the efforts of the parties to find universal standards, maritime data often runs into old data structures or specific structures for subdomains. Data integration is difficult in these contexts, even at the data storage management level, which requires uniform data structures (XML, JSON, text ...). One way to tackle this problem is certainly a storage strategy that exploits the concept of the Data Lake.

The Data Lake is a storage system that minimizes all input transformations in a “store as data comes” way, it is called schema on read, i.e. only when the data is read during processing is it parsed and adapted into a schema as needed. This facilitates the data acquisition process, and also that of interchange. It is also much easier to acquire and add new data sources and new players. In a data lake, however, the use of datasets with different formats and schemes involves an increase in management complexity, and also data searches and queries become more complex. For dataset location, discovery mechanisms are used. Those mechanisms employ graph or semantic databases that are used to implement metadata management and governance systems, often associated with the use of data lakes ([10], [11], [12]). To simplify the queries, the solution proposed by the EFFECTOR Project uses a semantic layer which will also be described in section 3.2.

In addition to raw data storage, the data lake provides several stages (also called area or layer) that process and contain data in a more structured way. In the end, a complete data lake system contains all the technological solutions that allow data to be stored in different formats, from the raw origin format to the more structured one, in order to serve as a reference point for the data of an entire system. The EFFECTOR solution embraces this concept and in addition to the raw data storage capacity includes additional modules for storage, these modules implement complementary features: the operation databases (DB) and the data warehouse ([12],[13]).

The operational DB is the database that usually implements some transaction mechanisms and is a so-called (OLTP) even if some more recent DBs NoSQL allows dialogue with the transaction in a different way. In any case, their purpose is often to be used as application storage. For example, in EFFECTOR operational databases are used to provide storage mechanisms to C2 systems. The data warehouse, on the other hand, is a storage system used mainly for data analysis, it provides more efficient mechanisms for data reading, which is the main activity of the analysis. Additional modules, the data flow manager and the input/output module have also been prepared to support the data lake. The first deals with moving data from one layer to another according to suitably designed and defined schemes, the second with acquiring data external to the system and publishing them after they have been processed, in Figure 3 is shown a detailed modules diagram [13].

The EFFECTOR project used this generic architecture so far described, as an architectural baseline for the implementations of different instances of the same system. Each of these instances used different mechanisms, while maintaining its general features. In Figure 4 the general scheme of one of these specific applications is presented. The storage system in the figure includes several technologies that implement the general scheme described above. In particular, this instance specifically manages and acquires data from the AIS system. The data acquisition is delegated to Kafka, a message brokering system that allows the acquisition of streaming data coming from the outside (in this case AIS messaging).

The acquired data are routed and processed using NiFi, which will also be described in the next chapter. The routed raw data is saved on an HDFS data storage as is, in their original format. The stored data is also indexed on Apache Atlas, a governance metadata system, which allows users to search for information between different databases. In addition, the data stored on HDFS can be accessed via the Hadoop web platform. After the raw storage, the data are also transformed through NiFi processors to be further

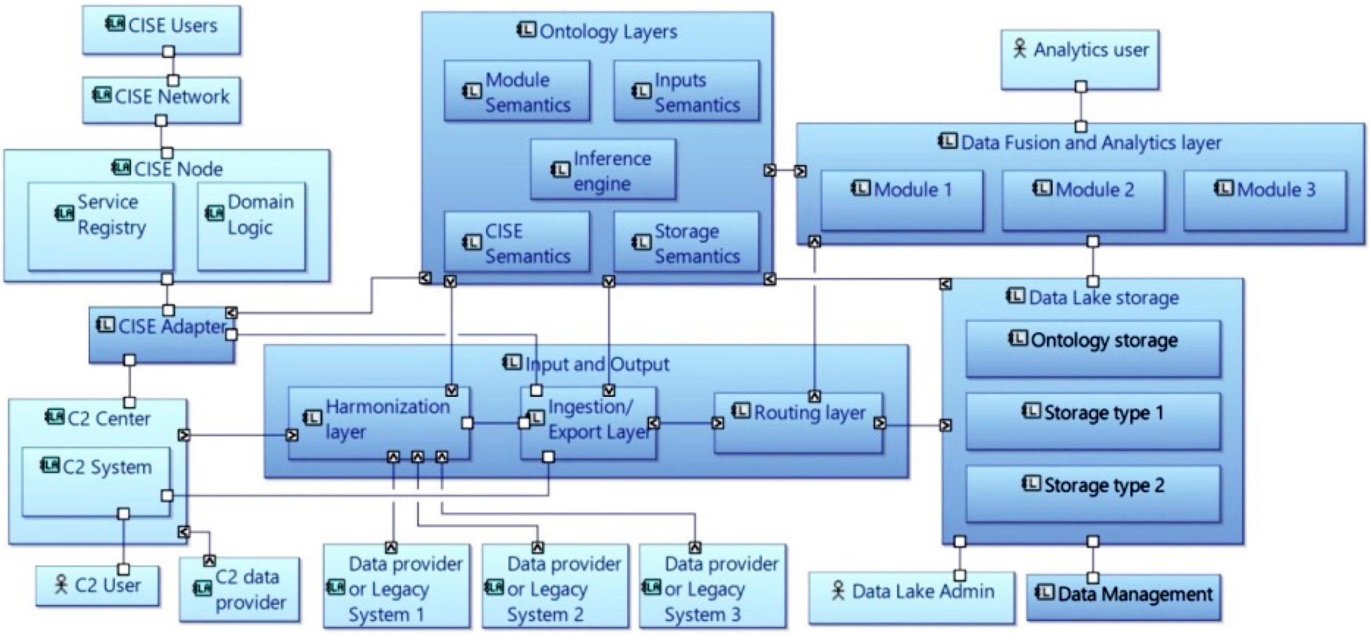


Figure 3 Modules Diagram

Source: [13]

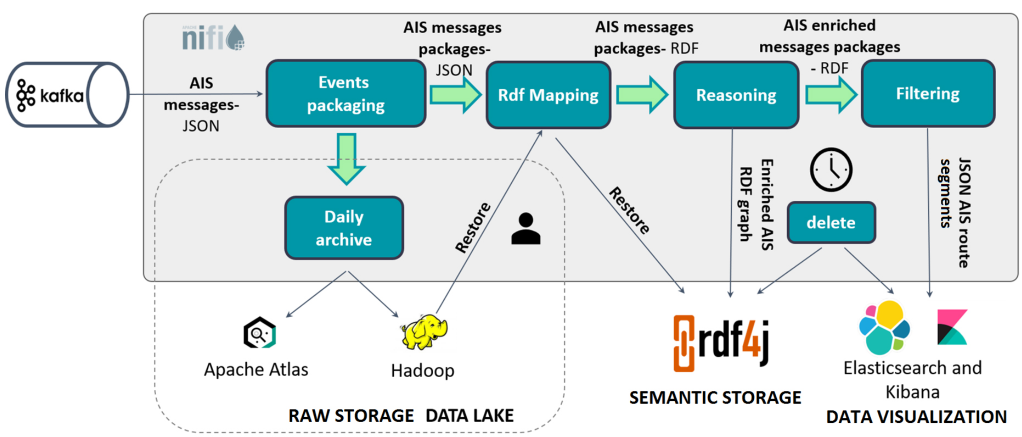


Figure 4 Data Lake implementation

Source: Authors

processed in the semantic layer, described in the next chapter.

Finally, the data after being enriched and further transformed by a reasoning process in the semantic layer, are sent to Elasticsearch, which allows indexing. Elasticsearch is also supported by a data visualization platform, Kibana. Kibana allows the creation of dashboards for data analysis and visualization. The visualization of the data through the dashboard provides a powerful mechanism able to interpret the information in a visual and direct way, which is a fundamental requirement to make the storage and reading mechanism of the implemented data, effective and useful.

Specifically, the Data Lake Architecture of the EFFECTOR Project played a key role in achieving the objectives of the project. It provided scalability, availability, and security. It is made up of a number of layers such as: Data Fusion Layer, Ontology Layer, Input and Output Layer, Storage Layer, and CISE Adaptation Layer.

The Data Fusion layer consists of a number of services whose purpose is to alert the operator about vessels with abnormal patterns. These alerts are propagated to the C2 Platforms and also stored in the Data Lake for future referencing. The Ontology Layer makes sense of the information flowing inside the Data Lake handling different data formats and inferring information not easily visible. The

Input and Output Layer is responsible for ingesting data from data sources, and processing and routing the information to other layers who are interested in receiving this information. The Storage Layer is responsible for managing the vast amounts of data that are generated by maritime systems. This ensures information is not lost and readily available scaling where necessary to cope with the volume. The CISE Adaptation Layer is responsible for handling incoming and outgoing CISE information received externally or sharing information found in the Data Lake.

3 Case Study: EFFECTOR Project

3.1 Project EFFECTOR overview and Multilayered Data Lake Introduction

One of the main aims of the ongoing EU Horizon2020 project EFFECTOR was to establish a model for big data collection from heterogeneous maritime data sources and legacy systems, which are to be stored in the form of national data lakes with decision support AI layers, incorporated in a CISE Network for maritime agencies. The purpose of this innovation action project is to develop the end-to-end interoperability framework based on novel data and surveillance technologies that will increase the resilience of the maritime safety and security domain and support the tactical and strategic operations at sea [14]. This is achieved by designing the complex architecture of data sources, legacy systems, external and internal data flows, decision support tools and AI modules, big data repositories, etc. The integration of all these components is designed to be compliant with the enhanced CISE Data and Services Model which is developed in the recently ended project ANDROMEDA (EU H2020) and managed by using several interoperable Command and Control Platforms (C2s), such as MUSCA, ENGAGE and SeaMIS. The developed Multilayered Data Lake, which represents the repository for collected and stored raw data and metadata, has the role of a structured and organized source from which the needed information could be quired when necessary for maritime authorities that participate as end-users in this project. Using the eCISE model classes and standard communication patterns the information from the national Data Lake could be retrieved and shared among participating authorities over the CISE Nodes, adaptors, and C2 they are operating. Therefore, the Data Lake, responsible for ingesting, cleansing, and aggregating heterogeneous data, consists of several plug-ins that transfer data from legacy systems and other data providers to storage units over adapters and ingestion layers, through aggregation, semantic and data fusion, and analytics layers that deliver the processed and timely information to CISE end-users on the national, regional and international level. These features are tested through several use cases including maritime Search and Rescue, illicit activities prevention and monitoring, weak signal detection, anomalies detection, situational picture provision, and early warnings for vessel collision or environmental threats [14].

In this paper, some of the core components of data lakes deployed for EFFECTOR Projects are shown including general remarks on Data Lake Architecture, SAT-AIS, CMEMS, and other external relevant satellite data for Data Lake, UAV layer, that enhanced overall collaboration between participating authorities in operational trials.

3.2 Semantic Layers of Data Lake and Recognized Maritime Picture

While the CISE works to improve interoperability in the operational context, to respond to the need to facilitate information search and sharing, the EFFECTOR exploits the semantic representation of information in Recognized Maritime Picture (RMP). The project was divided into three different pilots. For each pilot, a different system implementation was created, to be able to adapt to the specificities of each case. In this section we will illustrate the specific implementation of the Portuguese pilot. In the EFFECTOR semantic layer of the Portuguese pilot, the data relating to the situational picture are processed, organized, and structured according to the principles of Linked Data. Linked data are the means through which it is possible to build a network of connections, using data structured in triplets which are the basis of the semantic web.

Essentially the core of Linked Data and its semantic representation are a set of technologies, RDF and SPARQL.

The EFFECTOR semantic layer consists of a data flow, it is running on NiFi data flow manager, which acquires the data of the maritime situational picture and creates a semantic graph using an RDF data storage that also allows querying the system using SPARQL. Furthermore, the system is equipped with a reasoning system that uses a mini-batch architectural pattern to extract and infer further information. The inference mechanism is based on rules written in SPARQL that are applied within a time window of 10 minutes. The SPARQL role-based inference allows new connections to be added to the original information. The EFFECTOR semantic layer enables maritime users to request information and stored it on different nodes without the need to explicitly connect the databases, a SPARQL query can be performed on multiple repositories of different organizations, moreover thanks to the triplet structure (RDF) since there is no specific data schema, the information can be easily explored without the need to know its *a priori* structure.

The EFFECTOR semantic layer exploits a NiFi [15] data flow. The NiFi version used by the semantic layer exploits processors developed *ad-hoc* for EFFECTOR which add semantic potential to NiFi. The set of NiFi plus the newly developed processors create a framework for semantic data flow: NiFi for Semantic data. A section of the EFFECTOR semantic data flow is shown in Figure 5. The whole process starts with sending a collection of maritime data to the mapper processor that converts the data from JSON to RDF.

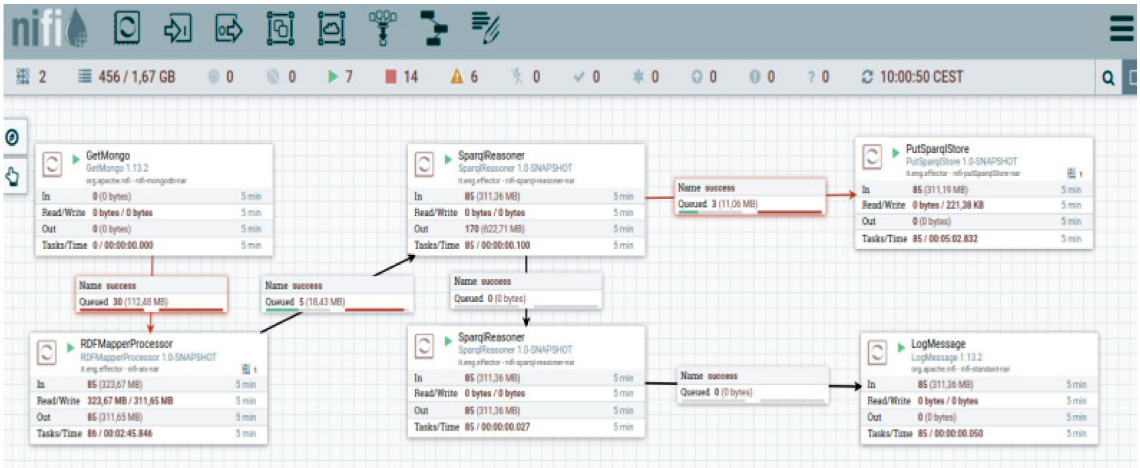


Figure 5 NiFi data flow

Source: Authors

```
<#VesselMapping> a rr:TriplesMap;
  rml:logicalSource [
    rml:source "nifi-processor";
    rml:referenceFormulation ql:JSONPath;
    rml:iterator "$.[*]"
  ];
rr:subjectMap [
  rr:template "http://effector.com/Vessel/{mmsi}";
  rr:class geo:Feature
];
rr:predicateObjectMap [
    rr:predicate effector:hasDraught;
      rr:objectMap [
         rml:reference "draught" ;
      ] ;
  ] ;
```

Figure 6 RMLMapper description

Source: Authors

The mapper processor is based on RMLMapper [16] an open-source project that uses a declarative language which is an extension of R2RML [17], a W3C standard for mapping data from relational DB to RDF. Below is shown a snippet of the mapping script (Figure 6) to convert JSON to RDF.

The mapping script starts from the LogicalSource which specifies the NiFi processor and the source type as JSON, it creates an RDF triple that describes the vessel, its draught object, and the property effector:hasDraught that connects subject and object. After the conversion of data in RDF, the flow is sent to the reasoner. The reasoner is an additional NiFi custom processor based on the RDF4J framework. The processor receives the data in RDF format and creates a temporal graph on which SPARQL queries are executed. These queries represent inference rules to obtain new information. The new information is inserted in the same graph and sent to the triple store. The triple store is the technology chosen to implement the storage of the semantic graph. Triple store allows saving data in a native mode in RDF, the framework used is RDF4J [18].

Below (Figure 7) the rules that create the "next" properties which connect two timepoint and the changeDir rule and (Figure 8) which infer if the vessel changes direction, are shown. The processed data are sent to the triple store. The data obtained and saved on the triple store are ready to be queried using SPARQL.

```
construct { ?g1 effector:next ?g2} where
{
  ?v geo:hasGeometry ?g2.
  ?g2 time:inXSDDateTimeStamp ?tm.
  {
    select ?v ?g1(min(?t2) as ?tm)
      where
      {
        ?v geo:hasGeometry ?g1.
        ?g1 time:inXSDDateTimeStamp ?t1 .
        ?g2 time:inXSDDateTimeStamp ?t2 .
        ?v geo:hasGeometry ?g2.
        filter(?g1!=?g2)
        filter(?t1<?t2)
      }
      GROUP BY ?v ?g1
  }
}
```

Figure 7 SPARQL Inference rules #1

Source: Authors

```
construct{?g2 a effector:changeDir. ?g2 effec-
tor:changeDirOf ?diff2 }
where{
  ?g1 effector:hasCog ?h1.
  ?g1 effector:next ?g2.
  ?g2 effector:hasCog ?h2.
  bind (if(?h1>?h2,?h1-?h2,?h2-?h1) as ?diff)
  bind (if(?diff>180,360-?diff,?diff) as ?diff2)
  filter(?diff2>45)
  ?g2 w3cGeo:long ?long.
  ?g2 w3cGeo:long ?lat
}
order by desc(?diff)
```

Figure 8 SPARQL Inference rules #2

Source: Authors

3.3 SAT-AIS, CMEMS and other relevant Satellite data for Data Lake

Earth observation satellite imagery was tasked, collected and processed by CLS through its VIGISAT ground receiving and processing network station [19]. Two kinds of imagery were collected and processed in near real-time for the detection of vessels in the area of interest [20]: High-resolution radar imagery composed of SAR images from the Sentinel-1 [21], TerraSAR-X [22] and Radarsat-2 [23] constellation, and high-resolution optical imagery from the Deimos constellation [24]. The SAR imagery allows all-weather day and night monitoring of the area of interest, while the optical imagery allows complementing the opportunity of revisit, however with a limitation related to cloud coverage. The non-cooperative vessel detection on earth observation imagery was complemented by collecting vessel information from AIS (Automatic Identification System), and from the D-AIS (Dynamic AIS) SPIRE system [25]. This AIS data was post-processed by CLS ensuring inter alia the validity of the messages with respect to their timestamps, IMO number, MMSI number, geographic coverage, vessel trajectory, location of the satellite, etc. Using AIS together with EO imagery aims to highlight vessels not reporting their positions or faking their positions. A better understanding of vessel behavior and detectability requires considering local weather and oceanic conditions (also referred to as Meto Oceanic / METOC data). For this purpose, CLS provided bridge access to Copernicus Marine Service data [26]. The following set of data was accessed: sea surface temperature and currents, wind, wave height and bathymetry.

3.4 UAV data provision to Data Lakes

3.4.1 The UAV system developed

Within the project EFFECTOR, objects detection algorithms based on deep learning have been trained and run on datasets gathered using UAVs ([27], [28], [29]). In our system, we chose to run the detection and tracking algorithm embedded in a processing unit carried on the UAV, instead of the ground control station where the video streams are sent. In this way the streamed video from the UAV already contains the bounding boxes drawn by the object detector and tracker, while a separate stream of data from the UAV transmits information on detected targets and their assigned unique IDs in JSON format. This was done in line with an edge processing approach to reduce the required bandwidth, improve response time, and avoid delays caused by video encoding and streaming, as well as inevitable interruptions and cuts in the video stream caused by connection issues, obstacles, and weather conditions ([30], [28]).

Since video streaming has high bandwidth requirements, any decision to reduce the streamed video resolution in case of a bad connection will not affect detection and tracking, as the detector directly receives the video feed from the UAV camera. Even if the video feed is interrupted, detection and tracking data are still transmitted via the JSON messages. If video streaming is interrupted the IDs of tracked targets are not lost. Another important aspect of this decision is the reduction of the load to the ground station, guaranteeing the solutions' expandability, as adding UAVs will not greatly affect the processing load at the ground station.

Table 1 Jetson AGX Xavier technical specifications

GPU	512-core Volta GPU with Tensor Cores
CPU	8-core ARM v8.2 64-bit CPU, 8MB L2 + 4MB L3
Memory	32GB 256-Bit LPDDR4x — 137GB/s
Storage	32GB eMMC 5.1
DL accelerator	(2x) NVDLA Engines
Vision accelerator	7-way VLIW Vision Processor
Encoder/Decoder	(2x) 4Kp60 — HEVC/(2x) 4Kp60 — 12-Bit Support
Size	105 mm x 105 mm x 65 mm
Deployment	Module (Jetson AGX Xavier)

The entire object detection and tracking stack that will be described runs on a Jetson AGX Xavier module (Figure 9). This embedded processing unit is lightweight and has low energy requirements. It will compress video using dedicated hardware and perform all necessary computation. The complete technical specifications of the module are presented in Table 1. The aim is for the computing module to be carried by a UAV and directly connect to its camera feed, making the system completely autonomous. We run all tests with the module attached to a specially built octocopter UAV (Figure 10). It is equipped with a pair of daylight and thermal cameras providing a video stream for the UAV's pilot, enabling Extended Visual Line-of-Sight (EVLOS) flight capability.

To provide extended capabilities of object detection and tracking, a 3-axis stabilized gimbal equipped with

Figure 9 The Jetson AGX Xavier module [29]

Figure 10 The octocopter UAV [29]

powerful RGB and thermal cameras is fitted under the UAV and connected to the processing unit for onboard processing. The results of the onboard processing are overlaid on the Full HD video stream and transmitted to the Intelligence Officer's workstation via a high bandwidth 2.4GHz radio ([31][32]).

3.4.2 Data Lakes and e-CISE compatibility

Information stored in the Data Lakes includes the vessels that have been detected and are being tracked, the status of the UAV as well as the missions assigned to the UAV, along with the outcomes or any modifications of the mission. The stack includes a set of adapters, written in the Python programming language that ensures compatibility with the CISE data model. When a vessel is detected or tracked, the relevant information is encoded to VESSEL entities within the CISE data model and can be transmitted directly. Directions sent to the UAV pilot by the command center of the operation, as well as acknowledgment or modification of the directions by the pilot will adhere to the TASK entity of CISE, while the UAV will transmit its status information in the AIRCRAFT entity of the eCISE (extended CISE) data model.

In this context information is shared with the Data Lakes through three data flows: the tasking flow, the status flow, and the detection/tracking flow. The tasking flow concerns the communication between the UAV ground station and operation control. Once a mission has been assigned to the UAV, operation control will transmit a TASK message. The UAV handler will receive the message at the ground station through the Data Lake and has the option to accept the mission, returning a positive response, deny the message (e.g. due to weather conditions), returning a negative response, or modify the mission (e.g. due to some obstacle in the flight path), returning the modified TASK entity (Figure 11). Adapters on both sides (operational control and UAV station) are responsible for the harmonization of messages to the CISE model. Data transmissions are through HTTPS. The status flow reports the status of the UAV, including its location and operational status through e-CISE AIRCRAFT entities. Information is saved in the Data Lake (Figure 12). The detection/tracking flow is responsible for sending information on detected and tracked vessels to the Data Lake. The UAV's onboard unit will detect vessels in the frames of the video feed and keep track of each individual vessel detected. It will also calculate the vessel's position, based on the location of the UAV, its altitude, the angle of the camera gimbal and the UAV's pitch and yaw. Information is transmitted to

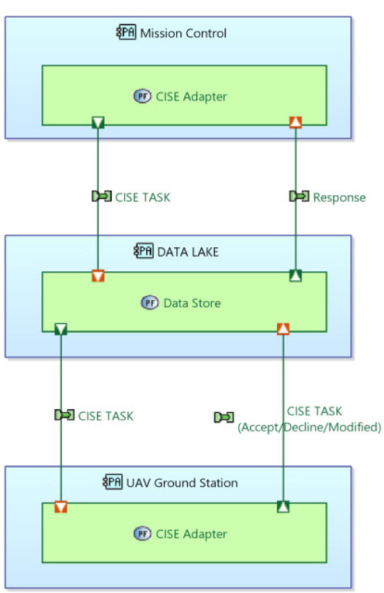


Figure 11 Tasking Data Flow

Source: Authors

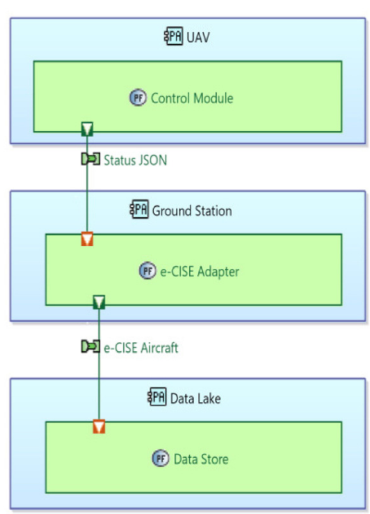


Figure 12 UAV Status Flow

Source: Authors

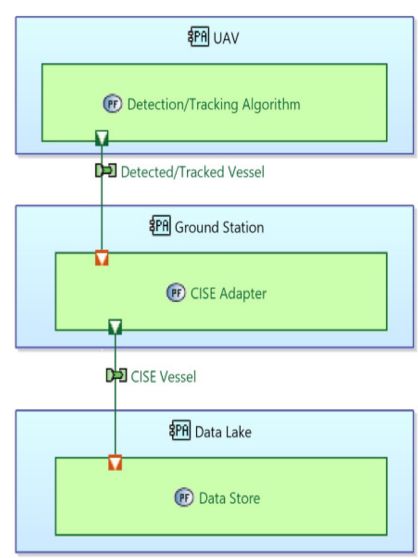


Figure 13 Vessel Detection and Tracking Flow

Source: Authors

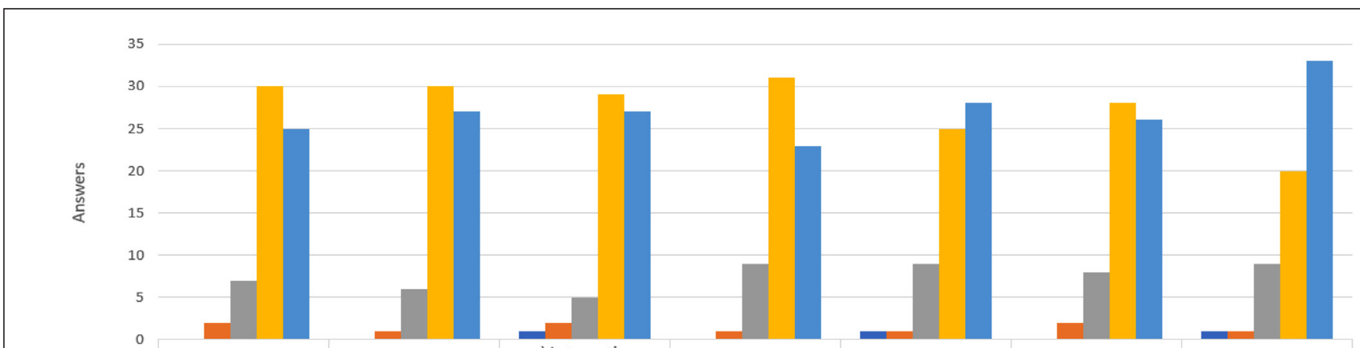
the ground station where the CISE adapter will create an appropriate VESSEL entity for each detection and tracking event (Figure 13).

4 Trials implementation, evaluation and projects results

The previously described components and software, integrated into a compact platform in the form of a multi-layered data lake with intelligent decision support plug-ins for faster and more efficient information retrieval and sharing, were tested and evaluated in the framework of project EFFECTOR Trials performed in France, Portugal, and Greece. Specific scenario parts of these trials and the involvement of end-users in performing the planned actions for testing and demonstrating the EFFECTOR Platform features are presented in [33] as a process of maritime shared data retrieval across the European CISE Network. Three validation e-surveys have been distributed for the French, Portuguese and Greek trial, respectively, and one general evaluation questionnaire summarizing the overall views on examined criteria. First of all, the most important indicators of project performance have been defined in order to test and confirm that the capabilities of developed software solutions have reached the expected level of technological readiness, efficiency, and interoperability. These criteria are classified in the following Key Performance Areas (KPA):

- Data Lake and Semantics capability,
- Information Exchange and Interoperability capability,
- Integration capability of various Data Sources and Novel Surveillance Systems,
- Data fusion and analytics capability,
- Command and Control System capability,
- Decision support capability,
- Legal and ethical compliance capability.

Each of these KPAs is divided into Key Performance Indicators (KPI) which were examined during the mentioned operational trials and scenarios. The participating project end-users with backgrounds in maritime safety, security, surveillance, joint operations, law enforcement, and other competencies in the maritime domain, tested the corresponding features of involved technologies, assets, C2s, legacy systems, and other services provided by the interoperable framework platforms. Specifically, in the Evaluation Questionnaire, which was distributed after all trial demonstrations, participated in addition to project consortium members, external end-users and practitioners from the same areas of expertise and with similar competencies, that were observing the performance of the trials. Of the totally collected 64 answers, 28 came from project end-users, 10 from technical partners, and 26 from the external end-user community. Rating the EFFECTOR Platform Capabilities quality, based on the abovementioned KPA, on the scale from “1 – Poor” to “5 – Excellent”, the respondents provided in general good marks with the detailed distribution of results depicted in Figure 14. Alongside this, the compliance of all examined technical components of EFFECTOR was evaluated at on a high level by project partners (Figure 15).



	a) Data Lake and Semantics capability:	b) Information Exchange and Interoperability capability:	c) Integration capability of various Data Sources and Novel Surveillance Systems:	d) Data fusion and analytics capability:	e) Command and Control System capability:	f) Decision support capability:	g) Legal and ethical compliance capability:
1 - Poor	0	0	1	0	1	0	1
2 - Fair	2	1	2	1	1	2	1
3 - Good	7	6	5	9	9	8	9
4 - Very good	30	30	29	31	25	28	20
5 - Excellent	25	27	27	23	28	26	33

Figure 14 General evaluation of EFFECTOR Platform Capabilities

Source: Authors

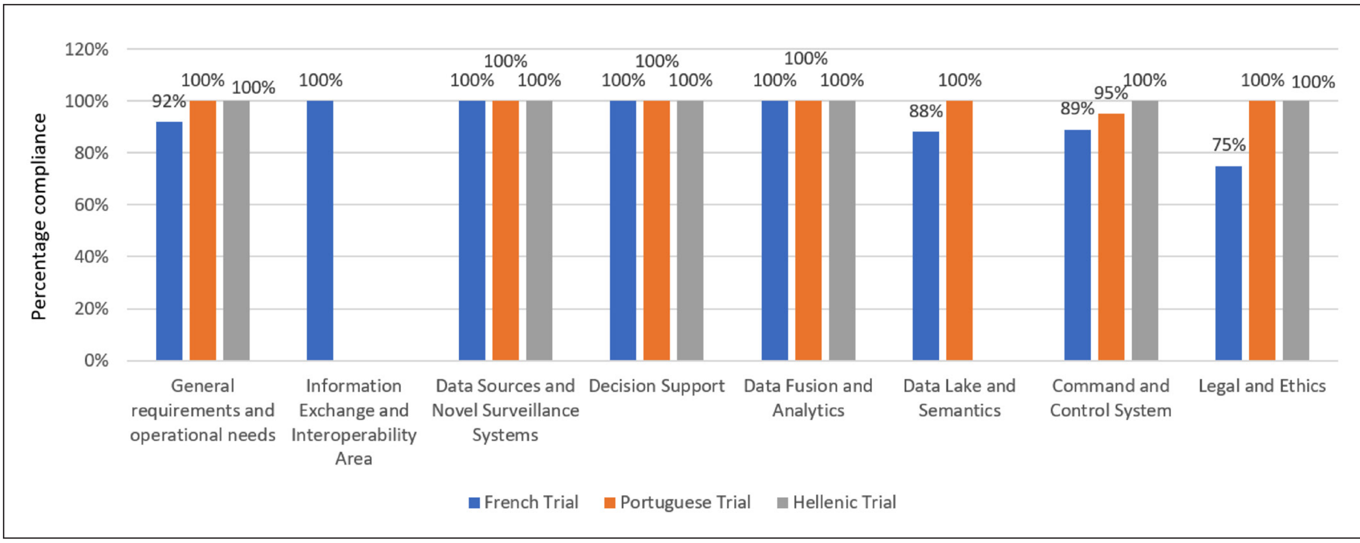


Figure 15 Compliance level (%) of all involved components of EFFECTOR Platform per trial

Source: Authors

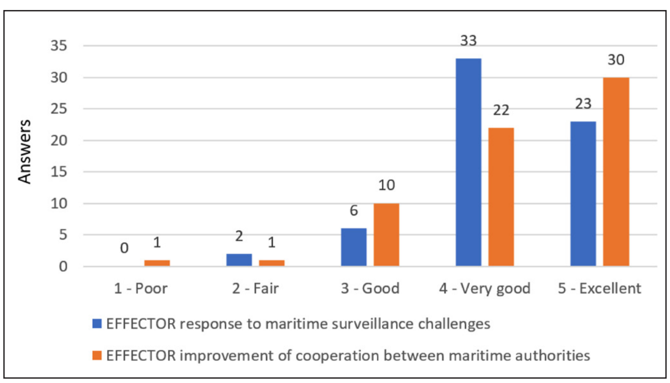


Figure 16 General evaluation of EFFECTOR Platform reponses on surveillance challenges and improvement of maritime authorities cooperation

Source: Authors

Participants were also asked to evaluate the response of the EFFECTOR Platform on maritime surveillance challenges and impacts on the improvement of cooperation between maritime authorities (Figure 16).

Using statistical analysis and description methods, all answers received by respondents participating in the three trials, have been compared and assessed. Due to many indicators monitored in surveys and size limitations, this paper shows some of the most important results and figures abstracted from Validation Surveys from Maritime Trials executed in France, Portugal and Greece. Regarding the first KPA, *General requirements,* and operational needs, the rate of improvement of the existing legacy systems on strategic and tactical levels have been examined, among the others. Considering the fact that EFFECTOR is a CISE-related project, aiming to foster the research and innova-

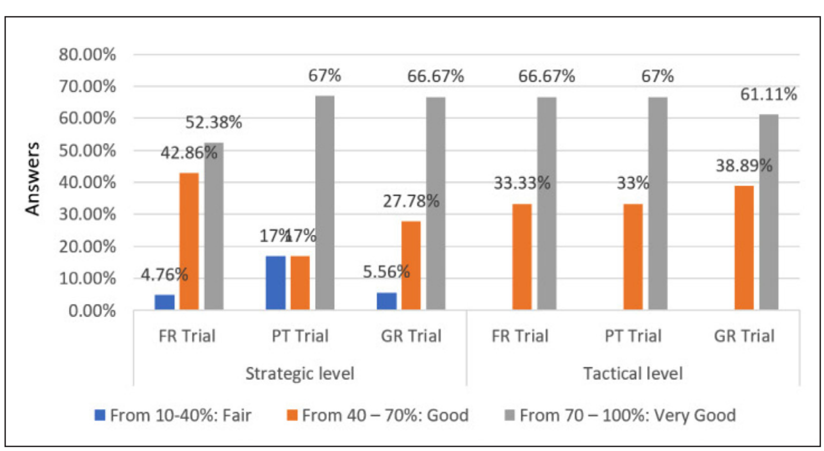


Figure 17 The improvement of the existing legacy systems on strategic and tactical level in all trials

Source: Authors

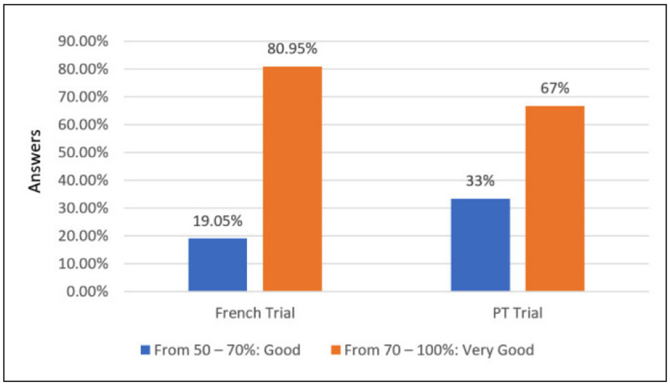


Figure 18 The degree of contribution to interoperability and CISE standardization in French and Portuguese Trials

Source: Authors

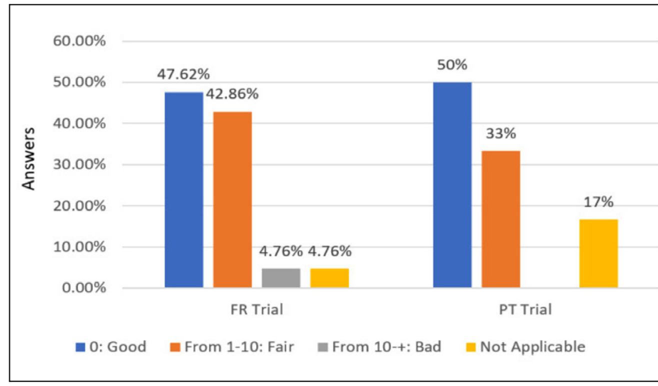


Figure 19 The number of C2 information unable to migrate to the Data lake (and should have migrated)

Source: Authors

tion actions on CISE network expansion, the degree of contribution of EFFECTOR Platform to interoperability and CISE standardization was examined, as well. The respondents have rated both criteria with percentual marks, where the majority stated that legacy systems and contribution to CISE standardization were improved from about 70 – 100% which is assessed as "Very Good". Figures 17 and 18 show the results of the survey obtained for these criteria.

The fourth KPA *Decision Support* was mostly focused on intelligent features of the EFFECTOR Platform, with a general assessment of plug-ins of C2 which provide support in the decision-making process of maritime operation centers. In surveys for all three trials, capabilities have been examined, such as showing the information in the report or inside a dashboard view, receiving and displaying the anomalies, integration with satellite (S-AIS, SAR), meteorological services, configuring of alarms for future alerts, creating and managing the missions and related tasks for available assets. Capabilities of showing information within the report and dashboard and integration with satellite and meteorological services, as well, showed maximum positive values assessed by respondents.

The sixth KPA *Data Lake and Semantics* examined the key product of the EFFECTOR Project, the multilayered Data Lake with national, regional, and international levels for collecting maritime information and their sharing using the eCISE Model. In the trials, the participating maritime authorities operated three C2s, of which the SeaMIS was used for the French Maritime Trial, MUSCA for the Portuguese Maritime Trial, and ENGAGE for Greek Maritime Trial. These C2s were connected to three national Data lakes respectively for each trial, having the capability to distribute and retrieve information to and from the specific data lake. Some of the considered questions in this KPA were the amount of C2 information unable to migrate to Data Lake and the number of Data Lake sources non-accessible from the trial C2. The positive results of the survey for these components related to French and Portuguese Trials are given in Figures 19 and 20.

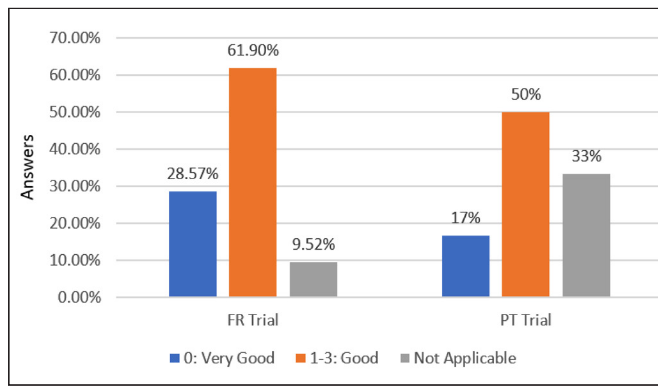


Figure 20 The number of Data Lake sources non-accessible from trial C2

Source: Authors

Furthermore, in this KPA the number of information systems used at the tactical and strategic levels for French and Portuguese trials were examined, as shown in Figure 21. The majority of answers indicated that 4-6 information systems (62% for French and 50% for Portuguese trials) interconnected with the data lakes, were used on the tactical level, which is marked as "Good", while some respondents recognized 7 or even more of these systems (33% for both trials). Regarding the information systems used on the strategic level, the majority of survey participants identified about 1-3 systems for the French Trial, while the rest identified 4-6 systems or 7 and more, depending on categorization and particular use of these information systems. The figures for the Portuguese trial are the same as for tactical-level information systems.

Regarding the other indicators tested and confirmed in trials, the respondents have mostly agreed on good interoperability of C2s and CISE and possibly EUROSUR with the capability of the majority of maritime authorities' partners to connect and exchange information via the CISE

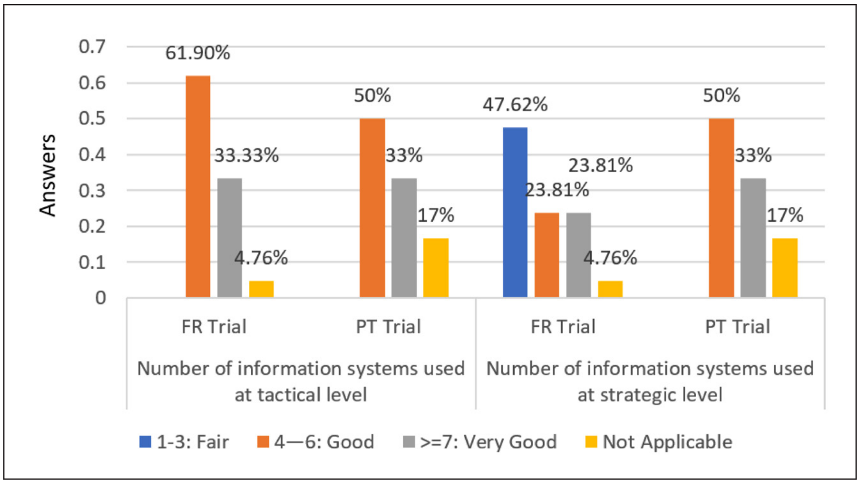


Figure 21 The number of information systems used at tactical and strategic level in French and Portuguese Maritime Trials

Source: Authors

Network. The majority of answers confirmed that 2 or more interoperable formats were used (Brokers + JSON format) with no compatibility issues throughout the data lake platform, and the number of open source technologies was higher than 1. Particularly, 52% of respondents identified that there were 1-20 eCISE classes interoperable with the EFFECTOR System, which is pretty fair, 38% of respondents spotted 20 and more of these classes, assessing them as very good, while 10% of respondents did not find applicable this question. Related to the local information system interoperability (wider than the C2 if relevant) evaluation, the majority of answers (48%) confirmed the level from 40-70%, which is assessed as "Good", while a significant number (43%) evaluated on the level from 70-100% assessing it as "Very good". Finally, all respondents examined and confirmed that more than 1 dataset (or data sources) were searchable in the system, as well as sources associated with an indexing process for easing searching. The same vast majority of answers confirmed more than one data source was enhanced using other data sources as well as data sources enhanced using semantics.

5 Conclusion

In this study, a holistic approach related to Maritime information sharing environment deployment using the advanced multilayered Data Lake capabilities was presented. The outcome of this study was also aligned with an EFFECTOR project case study. In a related context, an end-to-end interoperability framework for maritime situational awareness at strategic and tactical operations at sea was well defined and analyzed. Overall, by specifying all the needed features as inputs to a multilayered Data Lake, utilized through the EFFECTOR Project, we then provided a large-scale statistical analysis by validating and evaluating the e-surveys having been distributed to a significant number of stakeholders, including strategic and tactical operators and members from the external user community that participated to the EFFECTOR use case. Finally, the respective results are explicitly illustrated, based on an individual KPA definition.

This article is an extended version of our abstract published in Book of Abstracts of the International Conference on Sustainable Transport, Opatija, Croatia, 29 September – 1 October 2022, pp. 22–23.

Funding and 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.

Author Contributions: Conceptualisation: Zdravko Paladin. Research: all authors. Writing: all authors. Verification: all authors.

References

[1] Ackerman, H., King, J., "Operationalizing the Data Lake: Building and Extracting Value from Data Lakes with a Cloud-Native Data Platform, 1st Edition", O'Reilly Media, Inc., USA, 2019.

[2] Oreščanin, D., Hlupić, T. "Data Lakehouse – a Novel Step in Analytics Architecture", In 44th *MIPRO* Conference Proceedings, 2021.

[3] Paladin, Z., Kapidani, N., Lukšić, Ž., Mihailović, A., Scrima, P., Jacobé de Naurois, C., Laudy, C., Rizogiannis, C., Astyakopoulos, A., Blum, A., "Combined AI Capabilities for Enhancing Maritime Safety in a Common Information Sharing Environment", Proceedings of 35th Bled e-Conference, pp. 145-160, 2022.

[4] European Martime Safety Agency (EMSA): *Practical Guide for joining the CISE network*, v2.0, 13th January 2022.

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

[6] Berger, D., Hermida, J., Oliveri, F., Pace, G., "*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.

[7] https://www.andromeda-project.eu/ecise/index.html.

[8] Mihailović, A., Kapidani, N., Kočan, E., Merino, D., Rasanen, J., "Analysing the prospect of the maritime common information sharing environment's implementation and feasibility in Montenegro", In *Scientific Journal of Maritime Research*, *35(2)*, pp. 256-266, 2021.

[9] Paladin, Z., Mihailović, A., Kapidani, N., Merino, D., Grenner, J.-M., Vella, G., Moutzouris, M., Leuzzi, R., "Augmenting maritime Command and Control over a regional Common Information Sharing Environment implementation: Montenegro Case", *NMIOTC Journal,* 22(1), pp. 20-29, 2021.

[10] Hlupić, T., Oreščanin, D., Ružak, D., Baranović, M., "An Overview of Current Data Lake Architecture Models", In 45th *MIPRO* Conference Proceedings, 2022.

[11] Hlupić, T., Puniš, J. "An Overview of Current Trends in Data Ingestion and Integration", In 44th *MIPRO* Conference Proceedings, 2021.

[12] Paladin, Z., Kapidani, N., Lukšić, Ž., Nicoletti, T., Moutzouris, M., Blum, A., Astyakopoulos, A., "A Maritime Big Data Framework Integration in a Common Information Sharing Environment", Proceedings of 45th MIPRO Conference*,* pp. 1161-1166*,* 2022.

[13] Project EFFECTOR deliverable – D.3.1 Concept of Operation, System Design and Specification, 2021.

[14] 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/.

[15] https://nifi.apache.org/.

[16] https://rml.io/tools/.

[17] https://www.w3.org/TR/r2rml.

[18] https://rdf4j.org/.

[19] Hajduch, G., "VIGISAT ground receiving station and EMSA CleanSeaNet services", Remote Detection and Maritime Pollution: Chemical Spill Studies, pp. 77-86, 2021.

[20] Pelich, R., Longépé, N., Mercier, G., Hajduch, G., Garello, R. (2015, July). "Performance evaluation of Sentinel-1 data in SAR ship detection". In IEEE international geoscience and remote sensing symposium (IGARSS), pp. 2103-2106, July 2015.

[21] [Sentinel-1] https://sentinel.esa.int/fi/web/sentinel/user-guides/sentinel-1-sar, accessed July 2022.

[22] [TerraSAR-X] https://www.intelligence-airbusds.com/files/pmedia/public/r390_9_int_045_tsxpaz_en_low.pdf, accessed July 2022.

[23] [Radarsat-2] https://www.asc-csa.gc.ca/fra/satellites/radarsat2/, accessed July 2022.

[24] [Deimos-2] https://elecnor-deimos.com/project/deimos-2/, accessed July 2022.

[25] [Spire] https://spire.com/maritime/, access July 2022.

[26] [CMEMS] https://marine.copernicus.eu/fr, accessed July 2022.

[27] Mittal, P., Singh, R., Sharma, A., "Deep learning-based object detection in low-altitude UAV datasets: A survey. Image and Vision computing", 104, 104046, 2020.

[28] Bazi, Y., Melgani, F., "Convolutional svm networks for object detection in UAV imagery", IEEE transactions on geoscience and remote sensing 56(6), pp. 3107–3118, 2018.

[29] Vasilopoulos, E., Vosinakis, G., Krommyda, M., Karagiannidis, L., Ouzounoglou, E., Amditis, A., "Autonomous Object Detection Using a UAV Platform in the Maritime Environment", Proceedings of International Conference on Research Challenges in Information Science, pp. 567-579, Springer, Cham., 2022.

[30] Douklias, A., Karagiannidis, L., Misichroni, F., Amditis, A., "Design and Implementation of a UAV-Based Airborne Computing Platform for Computer Vision and Machine Learning Applications", *Sensors*, 22(5), 2049, 2022.

[31] Maltezos, E., Douklias, A., Dadoukis, A., Misichroni, F., Karagiannidis, L., Antonopoulos, M., Amditis, A., "The INUS platform: a modular solution for object detection and tracking from UAVs and terrestrial surveillance assets", *Computation*, 9(2), 12, 2021.

[32] Vasilopoulos, E., Vosinakis, G., Krommyda, M., Karagiannidis, L., Ouzounoglou, E., Amditis, A., "A Comparative Study of Autonomous Object Detection Algorithms in the Maritime Environment Using a UAV Platform", *Computation*, 10(3), 42, 2022.

[33] Mihailović, A., Kapidani, N., Lukšić, Ž., Tournier, R., Vella, G., Moutzouris, M., de Sousa, G.B., Blum,A., Paladin, Z., "Planning a Case for Shared Data Retrieval across the European Maritime Common Information Sharing Environment", The 26th International Conference on Information Technology (IT), IEEE, 2022. DOI: 10.1109/IT54280.2022.9743531.