Maritime Use Case:
Initial Requirements and Scenario Definitions
Work Package 3 Tasks 3.1-3.2 Deliverable 3.1

Authors

Konstantina Bereta (MarineTraffic), Konstantinos Chatzikokolakis (MarineTraffic), Dimitrios Zissis (MarineTraffic), Raffaele Grasso (CMRE), Gabriele Ferri (CMRE)
Elena Camossi (CMRE)

Distribution list:

Group:	Others:
WP Leader: MarineTraffic (MT) Task Leader: MarineTraffic (T3.1), CMRE (T3.2)	Internal Reviewer: RapidMiner (RM) INFORE Management Team INFORE Project Officer

Document history:

Revision	Date	Section	Page	Modification
0.2	22/02/19	1-3,8	1-15	First draft
0.3	28/02/19	3-8 Appendix A Appendix B	15-55	Integration of user design and user's questionnaire (roles, user stories, related use cases, datasets)
0.4	06/03/19	-	-	Internal Review Comments
0.5	12/03/19	All	All	Incorporation of reviewer comments
0.6	15/03/19	4.3, 5	21-26, 26-33	Finalisation of use cases. Created an additional section with useful data sources, to better identify the datasets needed for the use cases. Additions in description of datasets
0.7	18/03/19	5	26-33	Added more content on the datasets
0.8	19/03/19	All	All	Internal Feedback
1.0	26/03/19	-	-	Final Version

Approvals:

First Author	Konstantina Bereta (MarineTraffic)	Date: 28/02/2019
Internal Reviewer	David Arnu (RapidMiner)	Date: 06/03/2019
Coordinator:	Prof. Antonios Deligiannakis (Athena)	Date: 26/03/2019

Abbreviations

COP	Common Operational Picture (common operational picture)
MCCIS	Maritime Command and Control Information System (common operational picture on maritime level)
LINK	Tactical Data Link Network
CIC	Combat Information Centre
AIS	Automatic Identification System
SMART	Service oriented infrastructure for Maritime traffic Tracking
ESM	Electronic Support Measures
METOC	Meteorological Oceanographic
CMS	Combat Management System
ECDIS	Electronic Chart Display and Integrated System
POL	Patterns of Life
MSA	Maritime Situational Awareness
MDA	Maritime Domain Awareness

Table of contents:

1 Executive Summary

The present document describes the expert user requirements for the Maritime Situational Awareness (MSA) use case of the INFORE project. To achieve MSA, which is defined as the capability of understanding events, circumstances and activities within and impacting the maritime environment [2], the user has to combine different data sources and perform data analysis tasks in order to derive new information, e.g., activities and events at sea. Using large-scale data fusion techniques and real-time analytics, we can also predict such events, as well as possible risks and threats and enable stakeholders to take actions to prevent them.

Given these, this use case relies on the continuous/real-time fusion of heterogeneous data sources for specific areas of interest, combining global and local data, as shown in Figure 1. The term "global data" refers to Copernicus (Sentinel-1 and Sentinel-2 imagery) and Automatic Identification System (AIS) data, while the term "local data" refers to in-situ data sensed by autonomous maritime vessels (mobile platforms) equipped with appropriate sensor devices. Exploiting the continuous/real-time fusion of input data for predicting/forecasting critical events of interest, facilitates proactive decision making, and ultimately improves Maritime Intelligence, Surveillance and Reconnaissance. On the basis of the produced situational awareness, the mobile platforms will be able to autonomously adapt their navigation and sensor configurations to find a trade-off between area surveillance and the investigation of suspect targets. Sensor data fusion can support autonomous, real-time navigation decision making at the mobile platforms' level, while an effective coordination of a network of mobile platforms has the potential to highly increase the prediction and interesting event classification accuracy.

The use case will assess the robustness, performance, accuracy and usability of INFORE in real-world conditions. The aim of this document is to identify the kinds of conclusions that a user working in a maritime surveillance business scenario would like to draw from the INFORE prototype, and the results that must be provided from an analysis to aid drawing those conclusions. This document is structured as follows. Section 2 provides a brief introduction and Section 3 describes the pilot that will be employed in the scope of this use case in accordance with INFORE's workplan. Section 4 contains the compiled output from the questionnaires that were completed by expert end-users during the interviews that were conducted to collect user feedback and specify requirements, defining the use cases. These use cases comprise essentially the functional requirements for the implementation of the Maritime Situational Awareness use case of the INFORE project and they are followed by the description of non-functional requirements. The non-functional requirements refer to the performance, reliability and, availability of the services rather than the software functionalities (features) that are outlined by the use cases. Section 5 contains more information regarding some of the domain-specific data formats that are used in some of the datasets mentioned by the expert users. The questionnaire that was distributed to the users is provided in Appendix A and the respective answers are provided in Appendix B.

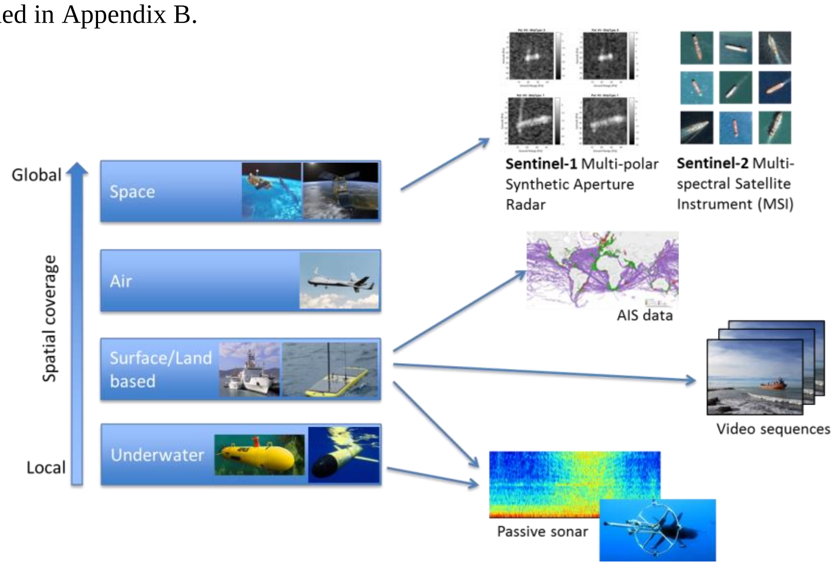

Figure 1: Integration of multiple global and local data sources

2 Introduction

Maritime Situational Awareness (sometimes also referred to as Maritime Domain Awareness) is the effective understanding of activities, events and threats in the maritime environment that could impact global safety, security, economic activity or the environment. The primary goals of MSA include “enhancing transparency in the maritime domain to detect, deter and defeat threats” and “enable accurate, dynamic, and confident decisions and responses to the full spectrum of maritime threats” [1]. As reported by the European Defence Agency (EDA), the full capability to act at sea depends primarily on the availability to establish a full and globalised picture essential for the preparation, the prevention and the potential required reaction to provide a sustainable stability [2]. EDA has recently included MSA in its revised Capability Development Plans, a list of capabilities needing further research, which is prepared in close cooperation with EU Member States and with the active contributions of the EU Military Committee (EUMC) and the European Union Military Staff (EUMS). This need results from an increasing number of dangerous or strange behaviour observed at sea, which, in turn, requires ensuring protection of sovereignty and infrastructure or counteracting terrorism and piracy from governments [3]. As such, there is now a growing number of commercial ships involved in illegal activities; such as i) trafficking/smuggling of all kinds (drugs, arms, oil etc.), ii) illegal fishing, iii) piracy, iv) terrorism and v) pollution (illegal discharges of oil and garbage, violations of ship emissions rules, etc.). Although in the past few years detecting these events in real time was very unlikely, today numerous monitoring systems produce constant streams of surveillance data at extreme rates and volumes.

These monitoring systems can be divided into two broad categories: cooperative and non-cooperative systems. Cooperative systems rely on the vessel’s crew to identify and report the vessel’s information, while non-cooperative systems are designed to detect and track vessels that do not voluntarily provide such an information. Cooperative systems include the Automatic Identification System (AIS), the Vessel Monitoring System (VMS), the Long Range Identification and Tracking (LRIT) system and others. Non-cooperative systems, on the other hand, include coastal and high-frequency (HF) radar, active and passive sonar, ground- or vessel-based cameras (e.g. thermal), satellite and airborne Earth Observation (EO) systems. EO systems can be divided into optical (generally visual and near-infrared) and Synthetic Aperture Radar (SAR) systems. Yet, despite the variety of tracking systems available and the massive data flows they produce, the awareness achievable based on these data remains largely partial.

The most commonly used collaborative tracking system is the Automatic Identification System (AIS), a self-reporting system that allows vessels to broadcast their identification information, characteristics, destination, along with other information originating from on-board devices and sensors, such as location, speed and heading. AIS messages are broadcasted periodically and they are received by other vessels equipped with AIS transceivers, as well as by on-ground stations. However, a common criticism regarding AIS-based surveillance is that it is not immediately useful for detecting the presence of non-cooperative vessels, the “non-shiners” or so-called “dark targets”. Ships involved in illegal activities attempt to hide their identification and positional information, either by counterfeiting their information (such as location coordinates, see Figure 2), masking it or not transmitting it at all, thus, attempting to remain undetected by law-enforcement bodies throughout the whole duration of the activity. On the other hand, non-cooperative systems are less sensitive to deception in non-military situations, but ship detection, identification, tracking, or speed and heading estimation are more complex. .This is because of their varying or limited coverage and of their reliability which can vary heavily depending on environmental conditions.

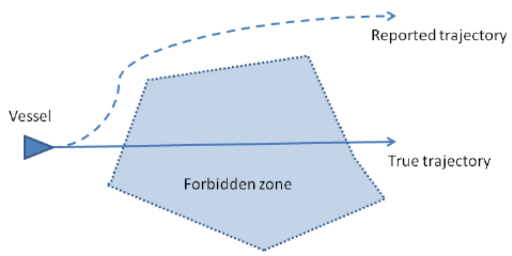

Figure 2: Patterns of life: Vessel route deviation (reported vs actual route). The light blue shaded area is a forbidden fishing zone As shown in the figure, reported locations forming the corresponding trajectory deviate from the actual ones, so as to conceal illegal fishing activities.

Often vessels involved in illegal activities try to hide their behaviour but follow a set of patterns: deviation from standard routes, unexpected port arrival, close approach, and forbidden zone entry [4]. One of the main pillars of maritime safety and security is Maritime Situational Awareness (MSA), or Maritime domain awareness (MDA), i.e., the capability of understanding events, circumstances and activities within and impacting the maritime environment [5]. MSA, or MDA involves developing the ability to identify patterns emerging within huge amounts of data, fused from various uncertain sources (information fusion) and generated from monitoring thousands of vessels a day, so as to act proactively to minimize the impact of possible threats.

Critical processes of MSA/MDA are thus knowledge discovery and data mining. The basic problem addressed by the knowledge discovery process is one of mapping low-level data into other forms that might be more compact (for example, a short report), more abstract (for example, a descriptive approximation or model of the process that generated the data), or more useful (for example, a predictive model for estimating the value of future cases) [7]. "Patterns of Life" (see Figure 2) are understood as observable human activities that can be described as patterns in the maritime domain related to a specific action (e.g. fishing) taking place at a specified time and place [8]. In the maritime domain, essentially, vessel based maritime activity can be described in space and time and it can also be classified into different kinds of activities at sea (fishing etc.). The spatial element describes recognised areas where maritime activity takes place; thus, including ports, fishing grounds, offshore energy infrastructure, dredging areas and others. The transit paths to and from these areas also describe the spatial element, e.g. commercial shipping and ferry routes etc., while the temporal element often holds additional information for categorising these activities. For instance, fishing period or time of the year plays an important role in categorising the motif behind the behaviour in the example of Figure 2. At the core of this process is data mining, an essential step in the process consisting of applying data analysis and discovery algorithms that, under acceptable computational efficiency limitations, identify patterns over the data. Hence, extracting patterns also means fitting a model to the data, finding structures, or in general any high-level description of a set of data. The fitted models play the role of inferred knowledge [7]. As such, anomaly detection can be understood as a method that supports situational assessment by indicating objects and situations that, in some sense, deviate from the expected, "normal" behaviour or patterns of life.

MSA is supported by the maritime situational picture: a combination of information from land, airborne, and satellite sensor systems, augmented with heterogeneous information from Geographical Information Systems (GIS) and vessel information repositories. The advantages in fusing data from multiple sensors and sources are that the final estimated vessel tracks are more accurate and with better confidence, extending to features that are impossible to perceive with individual sensors and sources, in less time, and at a lower cost. Also, it contributes towards better coverage and robustness to failure, thus improving the reliability and quality of the situational picture.

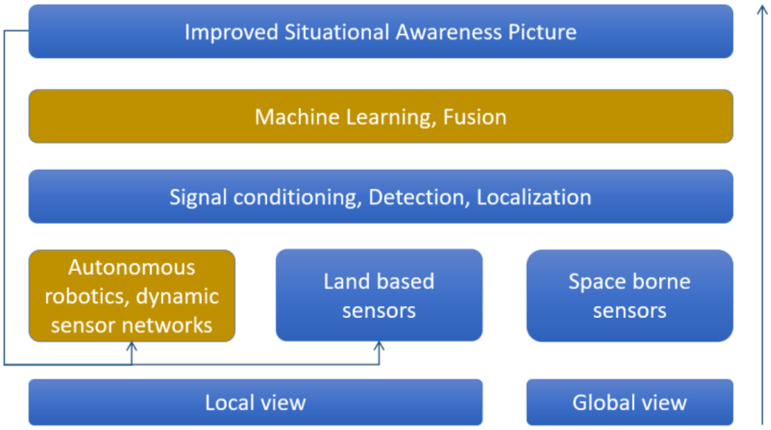

Figure 3: Pilot overview

Recent advances in marine robotics suggest that Maritime Unmanned Systems (MUS) can be used in the addressed surveillance scenario to increase the overall performance of the system [9,10,11]. Today's robots can provide persistent monitoring of areas of interest at lower costs than traditional approaches and can complement or substitute current solutions, typically based on fixed sensors and/or expensive ship-based operations (e.g. hydrophone arrays towed by vessels). Compared to traditional assets, these small, low-power and mobile units (mobile platforms) have usually limited processing and communication capabilities. However, when deployed in a spatially separated manner, they can be interconnected to form an intelligent network characterized by scalability, robustness, reliability and adaptability. Such sensor fusion supports autonomous, real-time decision making, while an effective network coordination of the mobile platforms has the potential to highly increase the classification performance. Increasing the effectiveness of the network is beneficial to reduce the workload of the network operator and to provide higher quality information to the final human decision-maker.

In real settings, data and information come at various degrees of accuracy, reliability and (im)precision, because information sources are of inherently different nature, sensors may have different precision, and various models may be more or less accurate. All these factors affect the quality of information and consequently the quality of the achieved situation awareness. The characterisation of information and source quality is thus critical for the proper consideration in the integration process (i.e., a source of lower quality can be discounted or discarded with respect to another of higher quality), but also for a coherent communication of information and source quality to the decision maker (e.g., highly conflicting or uncertain information may act as a warning and delay a decision). In this respect, formalising information and source quality is important towards improved awareness.

In the context discussed above, this use case covers typical border and harbour protection, and littoral areas surveillance, in which the interest is to classify single vessels of different size, their activities and possible interactions (e.g. a small boat approaching a larger vessel for illegal trafficking). Thus, the use case will allow to test and validate new approaches to:

1. Automatic detection, localisation and classification of targets of interest and their behaviour, exploiting data fusion and machine learning techniques for large quantities of heterogeneous data in real-time.
2. Autonomy of a network of heterogeneous mobile sensorised platforms. Based on the produced situational awareness, the mobile platforms will be able to autonomously adapt their navigation and sensor configurations to find a trade-off between area surveillance and the investigation of suspect targets. Towards this direction, forecasting events of interest to proactively adapt navigation is of the essence.
3. Semantic-driven description of heterogeneous information sources and their quality. The data collected during the experiments and the sensors used for the acquisition will be formally described and characterized with respect to their quality aspects, in support of an improved awareness and autonomy of mobile platforms.

3 Description of the Pilot

Overall the use case will assess the robustness, performance, accuracy and usability of INFORE in real-world conditions. According to INFORE's workplan and Work Package 3, the maritime use case will investigate and validate a multi-platform network for detecting and classifying possible threatening maritime activity. During the prototype evaluation and validation/trials, devices will be installed on land and will exploit data fusion between networked heterogeneous sensors, fixed and mobile. Fixed optical/thermal sensors will be installed on land to monitor the surveillance region. The acquired "local" optical/thermal data will be fused with "global views" of the area, e.g. AIS data, to improve the awareness on "dark" targets and to provide a first level of warning by detecting and classifying possible suspicious behaviour. The network will integrate a team of Autonomous Surface Vehicles, Wavegliders (http://liquidr.com/), which exploit the wave energy to move, equipped with acoustic passive sensors, and a more reactive and rapid vehicle equipped with an optical/thermal camera. The Wavegliders will autonomously survey the area of interest and react to data from fixed sensors with the aim of adapting their navigation and increasing their detection, localisation and classification of the suspect vessels. Furthermore, the reactive rapid vehicle will be triggered by the fixed sensors and/or the Wavegliders to reach and inspect the vessels of interest, thus helping to reach a final classification. A scheme of the system operating in the marine use case is visible in Figure 4.

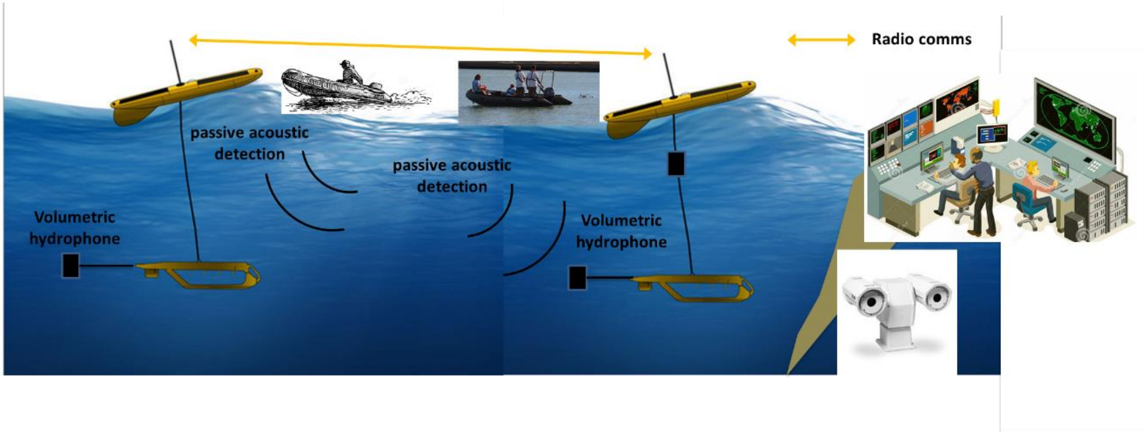

Figure 4: Maritime system in operation

In Figure 5 a high level description of the sequence of activities involved in the marine use case is reported. Fixed sensors (e.g. the thermal camera) inspects an area of interest together with the two Waveglider autonomous vehicles. When a tentative target is detected, the Wavegliders cooperate with the fixed sensors adapting their navigation to estimate the alarm position. Data fusion will be the basis for target localisation and classification. Finally, if the alarm is confirmed, the fast vehicle (FV) can be tasked for a final inspection of the confirmed target.
Different configurations will be evaluated with fixed sensors raising the alarm to the mobile robots and vice-versa.
Possible targets of interest are:

- Rubber boats – class 1: slow, heavy, with many people onboard. Typical of vessels transporting immigrants.
- Rubber boats – class 2: fast, light, with weapons onboard. Typical of pirates.
- Rubber boats – class 3: fast, with pilot and few people onboard (or without pilot). Typical of explosive manned or autonomous boat.
- Big fishing boat – class 4: larger and slow fishing boats.
- Oher vessels of opportunity – class 5: tankers, cargoes, etc.

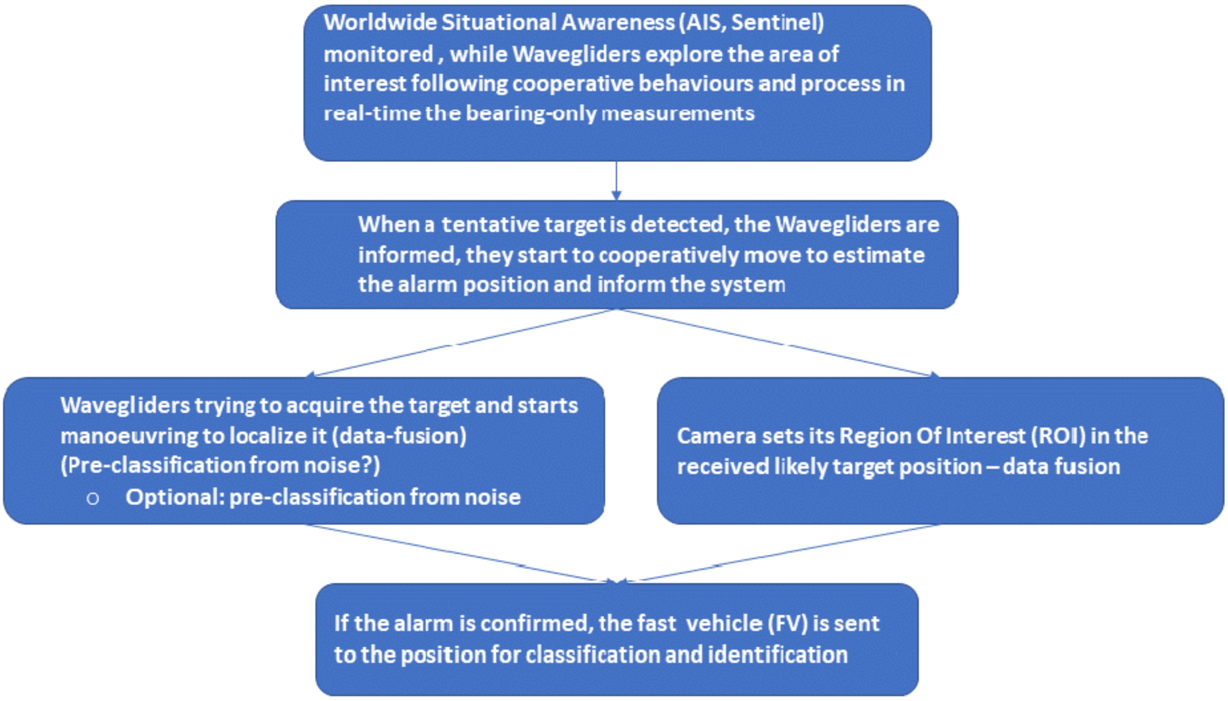

Figure 5: Maritime Use case Workflow

3.1 Scenario Description

The aim of the sea trials is the evaluation of the INFORE platform in the real-time detection of illegal vessel activities (such as piracy/terrorism, trafficking/smuggling, and un-economic-pollutive vessel behaviour) while assessing it in terms of robustness, performance, accuracy and usability in real-world conditions.

The generic workflow of the use case will be conducted as follows

- Data streams originating from the plethora of available sensors (AIS, Radar, Sonar, Optical and Satellite) are fed into the INFORE platform to be combined into a common operational picture
- Vessels are detected, identified and localised throughout these data streams, which are fused into a common operational picture for further investigation by an operator.
- Vessel tracks are further investigated for the detection of potentially illegal behaviour
- If potentially illegal behaviour is detected assets (Unmanned Underwater Vehicle (UUV) and/or fast-moving vessel) are deployed to the last known position or forecasted position of the event for further exploration

An overview of this workflow is depicted in Figure 5.

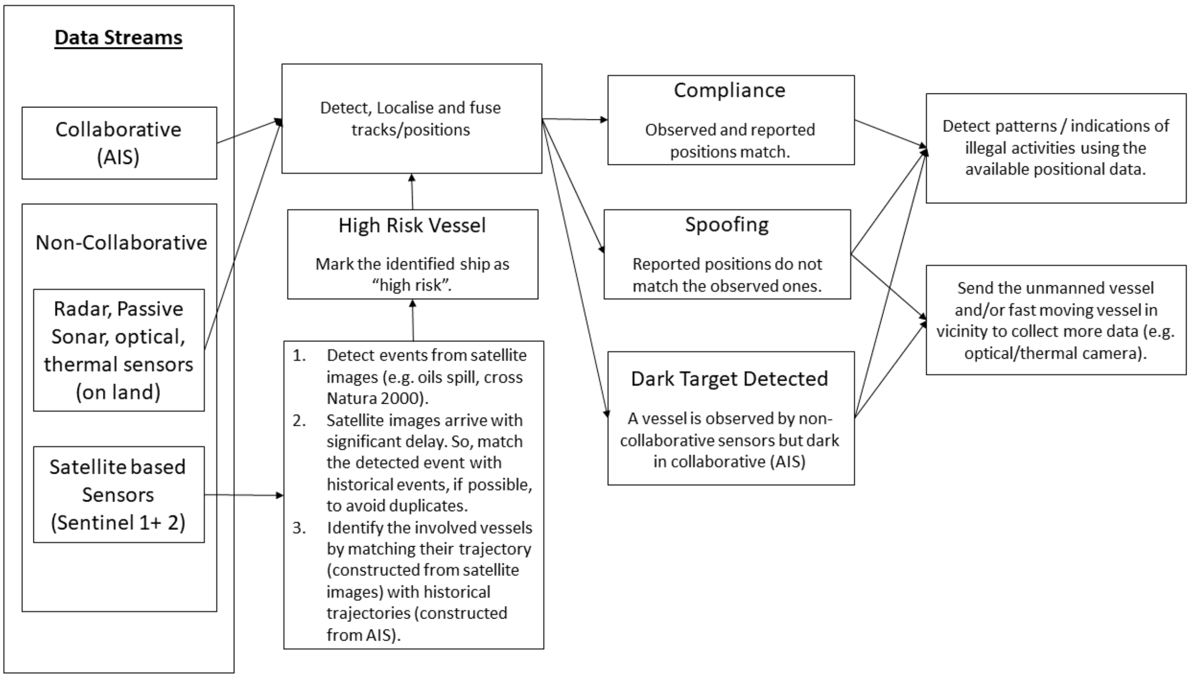

Figure 6: Experiment scenario overview

We define a separate scenario for each illegal activity to be detected. The first part of each scenario is the detection of events based on stream data analytics. Since different event detection algorithms and different combinations of data sources will be used in each case, we describe below the scenario that will be used in the respective cases. These scenarios are also shown graphically in Figure 6.

- **Illegal fishing:** Illegal, Unreported and Unregulated (IUU) fishing activities have become one of the ocean's biggest enemies. Illegal fishing is defined as fishing activities that take place without authorisation and violate conservation and management measures as well as national laws and international obligations. Marine protected areas are zones[1] where fishing activities are restricted, and they are believed to be essential for preserving marine resources. These areas can include marine reserves, closed areas, no-take zones or gear-restriction zones. In this case the experiment vessel will enter an MPA that will in fact be a part of the surveillance area. The vessel will attempt to hide its identity and type of activity. The INFORE platform should be able to detect other vessels violating the MPA using positional data from the fused, different sources and infer its activity. The expected behavior of the fishing vessel, initially outside the fishing area, is to enter the area with travelling speed and then move within the area with reduced speed. Additionally, the INFORE system should have the ability to detect the possible transshipment of illegal catches.
- **Piracy/terrorism:** Past activity has shown than in areas of high piracy (e.g. off the Coast of Somali), small crafts approach large merchant vessels at high speeds. Pirates mount their attacks from very small crafts, even where they are supported by larger vessels or 'mother ships', which tends to limit their operations to moderate sea states. Commonly, two or more small high speed (up to 25 knots) open boats or 'skiffs' are used in attacks, often approaching from either quarter or the stern[2]. For the purposes of the INFORE experiment, we will simulate the behaviour of a small vessel moving at high speed close to a larger merchant vessel. The INFORE system should accurately detect such incidents in real time and issue alerts.
- **Smuggling/trafficking:** Smuggling prohibited substances, or breaking embargoes is a common illegal activity at sea. Collaborative (e.g. AIS) and non-collaborative (e.g. Radar) data sources will be combined to obtain information about the location of vessels that are located in the surveillance area. Some of the

[1] http://ec.europa.eu/environment/integration/research/newsalert/pdf/58na2_en.pdf
[2] https://www.yen.gr/documents/20182/139132/bmpv3-lowres.pdf/8c3b235c-d25d-4bff-8fa1-117cd4a3b648

vessels used in the experiment will simulate the behaviour of smuggling/trafficking activities by following patterns such as going “dark” (e.g., switching-off their AIS transponder), performing ship-to-ship transfer activities and spoofing (i.e., reported position is different from the actual position, as in Figure 2) that the INFORE platform should be able to detect and issue alerts.

- **Pollution:** Illegal discharge of bilge waters is a significant source of oil and other environmental pollutants. Oil discharges can be observed directly using surveillance aircraft or using satellite imagery from Synthetic Aperture Radar (SAR) systems such (as that available thought the Copernicus Sentinel programme). While Sentinel Imagery is available at a different temporal resolution than the other sources, when a spill is detected, combining sources can generate a list of potential vessels implicated in the illegal discharge. For this scenario, the simulation vessels will act as vessels that have been previously reported as vessels involved in oil spills detected through satellite images. INFORE platform will be required to detect such illegal discharges, associate vessels implicated and issue alerts.

Following this, the patterns that correspond to the cases mentioned above are detected, the Unmanned Vessels (UVs) will be notified and they will closely monitor the simulation vessels.

3.2 Timeline, Surveillance Area and Setup

The use case will investigate and validate the INFORE platform for detecting and classifying possible threatening maritime activity. The platform will be evaluated in two separate phases

- During Phase 1 of the Maritime Use Case Experiment (to be reported in D3.2 on M18), devices will be installed on land and will exploit the fusion of networked heterogeneous sensors, fixed and mobile. Fixed optical/thermal sensors will be installed on land to monitor the surveillance region. The acquired “local” optical/thermal data will be fused with “global views” of the area, e.g. AIS data, to improve the awareness on “dark” targets and to provide a first level of warning by detecting and classifying possible suspicious behavior. Passing by vessels will be continuously monitored and illegal behaviour patterns will be simulated so as to test the performance of the system.

- During Phase 2 (to be reported in D3.3 on M36) the network will integrate a team of Autonomous Surface Vehicles, Wavegliders and a more reactive and rapid vehicle equipped with an optical/thermal camera. The Wavegliders will autonomously survey the area of interest and react to data from fixed sensors with the aim of adapting their navigation and increasing their detection, localisation and classification of the suspect vessels. Furthermore, the reactive rapid vehicle will be triggered by the fixed sensors and/or the Wavegliders to reach and inspect the vessels of interest, thus helping to reach a final classification. The fast moving vessels will perform a number of illegal patterns thus testing the ability of the system to detect these.

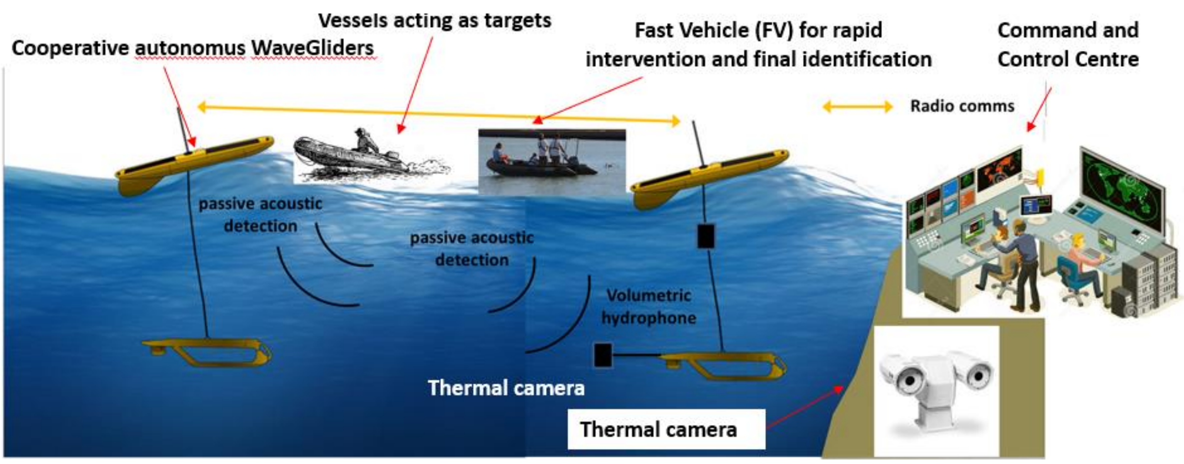

Figure 7: Sea experiment set up

The selected location of the experiment is the sea area of the Palmaria Island situated in the Ligurian Sea (Italy), close to CMRE facilities. This area is capable to host and deploy every type of sensor required (including the Autonomous Underwater Vessel (AUV)).

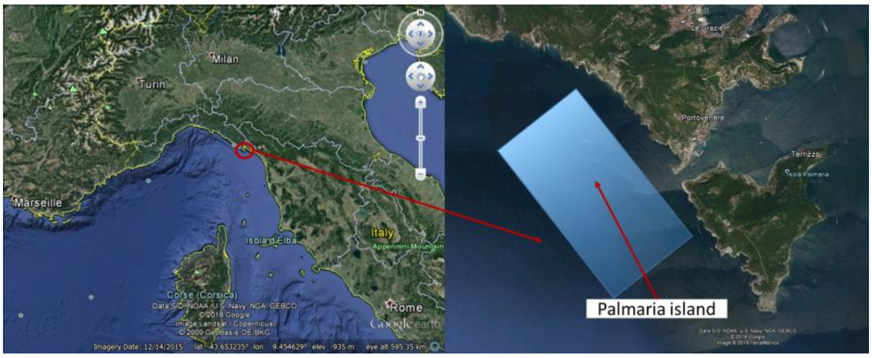

Figure 8: Palmaria island

Additionally, it is an area in the proximity of popular shipping routes (close to La Spezia port), as apparent in the density map below.

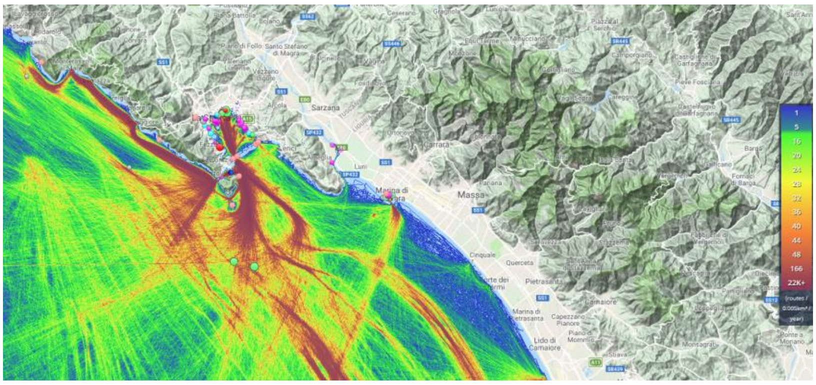

Figure 9: Density map showing shipping routes

4 User Requirements Questionnaires

Expert end users of the anomaly detection service involve national and international intelligence and security agencies. Usually, surveillance operators have to search and predict emerging conflict situations, e.g., risk for collision, anomalous vessels or suspicious activities from a large number of vessels within vast sea areas. Exploring and monitoring the data through surveillance systems may become a demanding activity for operators, not only due to the complexity and heterogeneous nature of the data, but also due to other factors like uncertainty, fatigue, cognitive overload or other time constraints [6]. It is common for agencies to use automated software which will assist in the fusion, detection and classification of surveillance data and related anomalous behaviour. However, these tools are either targeting specific stakeholders (e.g., Navielektro) or use proprietary data (e.g., Windward).

Windward offers risk management and marine insurance insights products aiming to increase maritime situational awareness and "*help intelligence, security and law enforcement agencies protect their national interests by revealing potential threats*".

Navielektro has developed the Maritime Anomaly Detection System (MADIS), a toolset that gives a unified view of maritime security incidents to operators, relieving them from conducting tedious and time-consuming tasks, such as identifying unexpected vessel behaviour and assessing the situation picture. This toolset is also integrated in Maritime Awareness Tactical Information System (MATIS), a similar commercial product offered by Navielektro that enables integration and correlation of data streams of multiple sensors (e.g., AIS receivers, maritime surveillance radars, etc.) to detect possible threats at sea. Both MADIS and MATIS, however, are specialised for military purposes.

Finally, Maero Space offers its TimeCaster™ service, which uses customers' satellite AIS, coastal AIS and other data sources to estimate vessels' current and future positions. The estimated positions are updated once every 10 minutes, giving a rough identification of potential anomalies such as vessels' proximity events.

Table 1 shows the current state-of-the-art regarding Marine Situational Awareness practices regarding anomalous vessel behaviour taking into consideration three stakeholders. It can be observed that none of the adopted approaches are able to perform forecasting, while most of the approaches deliver results in near real time (not near time). In the context of INFORE, we plan to go considerably beyond the state-of-the-art by offering services that employ forecasting mechanisms over streaming data for real-time dark target detection.

Table 1: Summary of MSA-related current practices for three specific stakeholders

Stakeholder	Real time	Data driven Vessel risk profiling	Warnings/Alerts	Complex Event Detection and Classification	Data driven definition of Patterns of Life	Adaptive (not rule based)
EMSA AUTOMATED BEHAVIOUR MONITORING[3]	Near real time	The user can identify the type of behaviour to be monitored (e.g. #drifting, #not reporting, #sudden change of	Operators are automatically alerted in real time via email or through the graphical user interface	Patterns, such as entering an area of interest, encounters at sea, approach to shore, and deviation from the usual route,	Not available	Rule based system

[3] http://www.emsa.europa.eu/news-a-press-centre/external-news/item/3206-automatic-detection-and-alert-triggering-of-ship-behaviour.html

Stakeholder	Real time	Data driven Vessel risk profiling	Warnings/Alerts	Complex Event Detection and Classification	Data driven definition of Patterns of Life	Adaptive (not rule based)
		speed)		are detected		
SeaVision[4]	Near real time and historical contexts	Updates provided on maritime security incidents including piracy alerts Vessel and port risk assessments provided	Yes, customisable alerts such as AIS Field Change	Not available	Not available	User Defined Rules
Smart Eyes on the SEas SEonSE (e-GEOS)[5]	Delivers Near Real Time services within 30 minutes		Automatic alert workflow based on customizable rules	Services for specific requirements (e.g. Oil Spill, Illegal fishing, anti-piracy, intelligence and security) based on customizable rules	Provides statistical traffic information	User Defined Rules

In this section, we provide the outcomes of the interviews that we conducted with the end users. We interviewed 3 expert users from MarineTraffic's network and 4 expert users from CMRE's network. In the rest of this section, we concisely aggregate their answers to our questions. In the next section, we provide this feedback in the form of user stories which are then translated into use cases in Section 5.

4.1 User Roles

The roles of expert users responding to INFORE's questionnaires are:

Table 2: Roles of Interviewed Expert Users

Role	Description	Tasks
MDA Researcher	Researcher in the Maritime Awareness Domain	Development of big data workflows to detect anomalous behaviour of vessels
MDA data scientist	Data Scientist in the Maritime Awareness Domain	Development of workflows that involve machine learning techniques using AIS data

[4] https://info.seavision.volpe.dot.gov/userguide

[5] http://www.egeos-services.com/SEnSE/SEonSE%20Brochure.pdf

Role	Description	Tasks
MDA data scientist	Data Scientist in the Maritime Awareness Domain	Development of workflows that involve machine learning techniques using AIS data
MDA data scientist engineer	Data Scientist in the Maritime Awareness Domain	Development of workflows that involve machine learning techniques using AIS data and satellite
Navy officer, Commander	Navy Officer on Commander level	Planning and conduct of Maritime Operations Procurement process in support of these operations
Navy officer, Captain	Research vessel captain and ship manager	Planning and execution of safe science at sea Ship management operation
Navy officer	Warfare officer	Support to Maritime Operations

4.2 User Stories

User stories have been reported during the user requirements interviews. Interviewed users describe the desired functionalities using their terminology, according to their position and domain of expertise. Once user stories are reported, they can then be compiled into use cases, where we formalise the scenarios that will have to be implemented and evaluated in the context of the project.

Table 3: User Stories of Interviewed Expert Users

ID	Title	Description	Actors Involved	Data/Systems involved
US_01	Combining heterogeneous sources of data	Combining heterogeneous sources of data using a common schema.	Data scientist	AIS, Satellite data, NATURA, Exclusive Economic Zones
US_02	Compliance with existing standards	Processing results should be standard compliant (e.g., OGC standards for geospatial data). This ensures compatibility with existing data processing and visualisation tools (e.g., QGiS)	Researcher	AIS data (CSV), Satellite (GeoTIFF), Other geospatial data like NATURA 2000 and EEZ (Shapefiles)
US_03	User Interface	Ability to choose from a pool of algorithms or data for a specific task in a user-friendly way.	Data scientist	AIS
US_04	User-friendly fine-tuning	Depending on the input data, I need to change the tuning parameters of my algorithms several times and for this I have to change the source code. It would be great if this could be done via a UI.	Data scientist	AIS

ID	Title	Description	Actors Involved	Data/Systems involved
US_05	Optimisation	I don't want to spend too much time fine tuning my algorithms in order to find out the best possible configuration for each use case	Researcher	AIS, Satellite, NATURA, EEZ
US_06	Modularity	Ideally, I want my workflows to be partitioned into building blocks as much as possible, as I reuse the same algorithms to build different workflows	Researcher	AIS, Satellite, NATURA, EEZ
US_07	Real-time big and historical and streaming data processing	I have workflows that combine big historical data and real-time streaming data in order to identify anomalies in vessel tracks and produce real-time results	Data scientist	AIS
US_08	Data fusion	I would like to be able to combine AIS data with satellite data so that I can discover AIS-off vessels that appear in satellite images.	Data scientist	AIS, Satellite, NATURA, EEZ
US_09	Vessel detection	I use satellite images to identify vessels, but I do this manually using QGIS.	Data scientist	AIS, Satellite
US_10	Data fusion	I would like to be able to incorporate the use of a neural network in my workflow in order to provide annotated images (e.g, Satellite images that contain vessels) and predict the existence of vessels in other, test satellite images	Data scientist	AIS, Satellite data
US_11	Cross validation	I would like to combine different data sources that contain the same information in order to perform cross-validation tasks.	Data scientist	AIS, Satellite
US_12	Seamless integration of different datasets	I would like to be able to import datasets of different formats (e.g., GeoTIFF images, vector file formats like Shapefiles, jdbc connections to relational databases) in a uniform way, without the need to handle each kind of data source differently.	Data scientist	AIS, Satellite data (GeoTIFF images, Shapefiles)
US_13	Integration and fusion of real time and historical heterogeneous data sources and systems	Depending on the task, command level has to find the appropriate, reliable and supporting databases and to merge them into the common picture. Day-to-day activity is to assess the situation, filtering, prioritizing and merging the data to get more specific and connected ones, depending on the task, on the context, in turn on the operational level. Current tools are not really supporting / linking / connecting and instead simply acting as a database (e.g. to find referenced	Commander, Captain,	AIS, Radar, LINK, MCCIS, CIC, SMART, ESM, CMS, Intelligence reports, METOC forecast and observations, VHF radio communication, Visual aids, acoustic sensors, ECDIS

ID	Title	Description	Actors Involved	Data/Systems involved
		main documents or POC's) Normally they are not flexible, access is limited or to adapted to specific task (e.g. adding applications or merging of information not feasible) The work happens not online with or supported by the system – instead normally offline without any connectivity by e.g. simply using or programming Excel / word / common tools System should be flexible, adaptable in cooperation with other tools / systems / databases. If any other [source] is identified acquiring data would be vital to fill the possible capability gap		
US_14	Quality of information sources	Reliability of AIS Information/need to increase reliance/verify information through automated multiple source data fusion (Radar/ESM/AIS/Visual ID/POL…) [Approximation is acceptable] knowing the level of accuracy / Having an estimated error in mind	Commander, Captain, Operational officer	AIS, Radar, LINK, MCCIS, CIC, SMART, ESM, CMS, Intelligence reports, METOC forecast and observations, VHF radio communication, Visual aids, acoustic sensors, ECDIS
US_15	Planning of maritime operations	Assessment / defining Courses of Action: Fulfil the Mission by calculating risks / defining decisive points / comparing advantages – disadvantages / assessing need for manpower – material – education of personnel for the requirements / developing Courses of Action (plan)	Commander	AIS, Radar, LINK, MCCIS, CIC, SMART, ESM, CMS, Intelligence reports, METOC forecast and observations, VHF radio communication, Visual aids, acoustic sensors, ECDIS
US_16	Interactive Systems/Algorithms customization/tuning, workflow adaptation	Current systems cannot be adapted to the requirements, are overwhelming the user with (mostly) unneeded information or are too complicate to handle, a lot of information is not linked or even more not suggested (e.g. intelligent network) Existing tools for presenting information are not flexible and tailored to specific tasks.	Commander, Captain	AIS, Radar, LINK, MCCIS, CIC, SMART, ESM, CMS, Intelligence reports, METOC forecast and observations, VHF radio

ID	Title	Description	Actors Involved	Data/Systems involved
		Filters are necessary to enrich or decrease the information flow in the system.		communication, Visual aids, acoustic sensors, ECDIS
US_17	Target detection and classification	Detect objects of interest (ships, aircrafts), classify them, assess what they are doing, and decide whether any type of follow-on action is required. Fusion of heterogeneous sources and availability of information (e.g. Small craft identification in coastal areas) are two typical challenges […] allow the officer responsible to determine whether a vessel was to be classified as “Friend” or “Foe”.	Commander, Captain	AIS, Radar, LINK, MCCIS, CIC, SMART, ESM, CMS, Intelligence reports, METOC forecast and observations, VHF radio communication, Visual aids, acoustic sensors, ECDIS
US_18	Anomaly detection, event recognition and forecasting	It is important to detect on the spot what is relevant for the conduct of an operation. In current systems, important information is not suggested. Maritime Operations: Maritime traffic, environmental conditions and impact on operations, detecting anomaly behavior, forecasted sensor ranges to assess own capabilities and impact on tasking (area coverage) [Evaluate] coherence with normal POL in the area, counter smuggling or illicit activity [Detect] potential anomalies for the maritime surveillance, increase of safety of navigation Visual overlay of information is supported by e.g., ECDIS, which automatically displays information on ship’ sensors (AIS, RADARs, ship’s meteo and acoustic sensors) to give command an updated and live representation of the surface picture. This is not a surveillance tool but a navigational aid	Commander, Captain	AIS, Radar, LINK, MCCIS, CIC, SMART, ESM, CMS, Intelligence reports, METOC forecast and observations, VHF radio communication, Visual aids, acoustic sensors, ECDIS
US_19	Timeliness	Achieve and maintain a timely, reliable, accurate Recognized Maritime Picture functional to combat threats such as human smuggling, terrorism at sea, piracy, as well as arms and drug trafficking. The daily challenge is always to keep the overview to the whole process without	Commander, Captain, Operational officer	AIS, Radar, LINK, MCCIS, CIC, SMART, ESM, CMS, Intelligence reports, METOC forecast and observations,

ID	Title	Description	Actors Involved	Data/Systems involved
		going too far to details. Situation might change quickly and mostly operators ashore or at sea are limited by time. Users need to have a reliable, real time answers to their problems from the BIG DATA picture; real time help to maritime and harbours surveillance		VHF radio communication, Visual aids, acoustic sensors, ECDIS
US_20	Workflow automation and optimization	Processes are not automated, almost entirely manual Human factor always predominant	Commander, Captain, Operational officer	AIS, Radar, LINK, MCCIS, CIC, SMART, ESM, CMS, Intelligence reports, METOC forecast and observations, VHF radio communication, Visual aids, acoustic sensors, ECDIS
US_21	User friendly system user interface	System should be simple to handle. User interface should be really simple and intuitive allowing new operator to learn advanced skills in short time.	Commander, Captain,	AIS, Radar, LINK, MCCIS, CIC, SMART, ESM, CMS, Intelligence reports, METOC forecast and observations, VHF radio communication, Visual aids, acoustic sensors, ECDIS

4.3 Use Cases

This section describes the use cases that have been derived from of the user stories that are provided in the previous section. Use cases are compiled user stories and they describe in more detail a step-by-step process that is followed in the implementation of each workflow, the input data and the expected output.

Table 4: Use cases derived from user stories

Use case ID	**MT_01**
Short description	Computation of anomaly detection indicators, such as estimation of the estimated arrival time (ETA) at a port, which can deviate considerably from the reported ETA.
Actors	Data scientist Commander, Captain, Operational officer (stakeholders)

Initial Conditions	When a vessel departs for a new voyage, the Chief Engineer is responsible to register the estimated arrival time to the destination as part of the AIS data that is transmitted using the transponder of the vessel. Initially, this is the only information that we have about the vessel's estimated arrival time to the destination.
Final results	Accurate estimation of the actual arrival time of a vessel to its destination.
Main Process	Import historic voyage data for a period of a year for a specific vessel market segment Based on these data, we build a machine-learning regression model based on historic voyage data The model learns to estimate vessel speed at a specific point along the vessel's chosen route based on a number of weather characteristics at this point as well as market specific features Then, the model is used to point speed estimates along a predefined number of waypoints. This in turn translates into estimations of voyage duration between consecutive waypoints and, thus, into indirectly estimating the time of arrival at the destination port.
Processed data	Historic AIS data Weather data Market-specific data
Generated data	Predictions about the speed of the vessel at a future point and the voyage duration between two consecutive waypoints Estimation about the arrival time of a vessel at a port
Related user stories	US_01, US_02, US_04, US_05, US_6, US_7, US_8, US_11, US_12, US_18

Use case ID	**MT_02**
Short description	Execute a clustering algorithm on vessels positions
Actors	Data scientist
Initial Conditions	The dataset that contains vessel positions is stored in a relational database.
Final results	Clusters of the locations of vessels
Main Process	Extraction of AIS data from the relational database as CSV files Import the CSV files into another, spatially-enabled PostGIS database Execute some data processing tasks in PostGIS Perform DBSCAN clustering on this data, which is written as a python program
Processed data	AIS data
Generated data	Clusters of vessels positions
Related user stories	US_01, US_2, US_4, US_5, US_7, US_12

Use case ID	**MT_03**
Short description	Use satellite images to identify AIS-off vessels and mark them as possible dark targets (i.e., possibly involved in illegal activities such as smuggling,

	trafficking, illegal fishing. etc.)
Actors	Data scientist, Researcher, Commander, Captain, Operations department officer (stakeholders)
Initial Conditions	The dataset that contains vessel positions is stored in a relational database and we have Sentinel 1 and Sentinel 2 images available.
Final results	Identification of vessels in Satellite images.
Main Process	Extraction of AIS data from the relational database as CSV files and import them in QGIS Import Satellite image in QGIS Correlate Satellite and AIS data to annotate image tiles about whether or not they contain vessels (and also maybe vessel characteristics like the type of vessels, etc.) Train a neural network using the training set produced from above so that it is able to identify vessels in satellite image tiles.
Processed data	AIS data, Satellite imagery
Generated data	Information about vessels contained in a satellite image.
Related user stories	US_01, US_02, US_03, US_04, US_05, US_6, US_7, US_8, US_09, US_10, US_11, US_12

Use case ID	**MT_04**
Short description	Detection of vessels with uneconomic, unenvironmental behaviour
Actors	Researcher Commander, Captain
Initial Conditions	We have information about the location of vessels in real-time but we do not know whether it enters a protected site or an emission controlled zone.
Final results	Real-time identification of vessels that enter emission controlled zones (ECA) or protected zones (NATURA 2000 sites)
Main Process	Extraction of AIS data, as well as ECA data from the relational database as CSV files and import them in QGIS Import NATURA 2000 dataset that is available in Shapefile format into QGIS. Identify which vessels enter protected sites or ECA zones.
Processed data	AIS data, ECA zones, NATURA 2000 sites.
Generated data	Given the current position of the vessel, real-time information about whether or not it has entered an ECA zone or has violated a protected site.
Related user stories	US_01, US_02, US_03, US_04, US_05, US_6, US_7, US_8, US_09, US_10, US_11, US_12

Use case ID	**MT_05**
Short description	Discover possible dark targets for illegal fishing

Actors	Data scientist, researcher Commander, Captain, Operations department officer (stakeholders)
Initial Conditions	We know the positions of vessels as well as their type (e.g., fishing vessels) from the AIS data but we don't know if they perform illegal fishing
Final results	Indicators that suggest that it is possible that a group of vessels perform illegal fishing.
Main Process	Extraction of AIS data from the relational database as CSV files Import datasets about areas where fishing is illegal (exclusive economic zones, protected sites such as NATURA 200) Correlate the two datasets and discover fishing vessels that are stopped at sea (possibly fishing) in areas where fishing is illegal
Processed data	AIS data, EEZ data, NATURA 2000 sites
Generated data	Added information about possible dark targets involved in illegal fishing.
Related user stories	US_01, US_02, US_03, US_04, US_05, US_6, US_7, US_8, US_09, US_10, US_11, US_12

Use case ID	**MT_06**
Short description	Detection and classification of suspicious vessels with anomalous behaviour in proximity of harbour areas: - Anomalous manoeuvring - Entrance in interdicted areas - Suspect rendez-vous - Dark targets (not transmitting AIS)
Actors	Data scientist Operational officer (system user), Commander (decision maker)
Initial Conditions	We have interdicted areas in proximity of the harbour to be monitored. A network of sensors (fixed and mobile) have been deployed to survey the area and to investigate the behaviour of suspicious targets.
Final results	Real-time detection and classification of targets as: • Rubber boats – class 1: slow, heavy, with many people onboard. Typical of vessels transporting immigrants. • Rubber boats – class 2: fast, light, with weapons onboard. Typical of pirates. • Rubber boats – class 3: fast, with pilot and few people onboard (or without pilot). Typical of explosive manned or autonomous boat. • Big fishing boat – class 4: larger and slow fishing boats. • Oher vessels of opportunity – class 5: tankers, cargoes, ecc. and identify anomalous behaviours.

Main Process	• Fixed sensors (e.g. the thermal camera) inspect an area of interest together with the two Waveglider autonomous vehicles. • When a tentative target is detected, the Wavegliders cooperate with the fixed sensors adapting their navigation to estimate the alarm position. • Data fusion (between the acoustic sensors of the robots and between cameras and robot mounted sensors) will be used for target localization and classification. • Finally, if the alarm is confirmed, the fast vehicle (FV) can be tasked for a final inspection of the confirmed target.
Processed data	AIS, Acoustic data, Thermal/optical data
Generated data	Recognized environmental picture, alarm (with class, position, speed, classification accuracy)
Related user stories	US_13, US_14, US_17, US_19, US_20

Use case ID	**MT_07**
Short description	Semantic driven integration of information sources and quality related information in support to maritime situational awareness, assessment, analytics
Actors	Researcher, Data scientists, Operations Dept. officer (system user), Commander (decision maker)
Initial Conditions	Information sources including sensors and data processing software (e.g. detection and classification algorithm) generate non-integrated flows of information
Final results	Information from sensors and software is integrated according to the maritime situational awareness data model (ISQ ontology and derived models) and stored in the knowledge base. Information and information sources are quality evaluated, relying on quality axioms specified in the data model. Software and algorithms might benefit of the quality evaluation of sources and information and adapt the processing parameters. The final knowledge base is an integrated view of the maritime situation, enriched with quality evaluation of information and information sources.
Main Process	AIS data, acoustic sensors data, thermal/optical videos, satellite images, data processing output, are ingested in the ontology and automatically enriched with quality evaluations, and corresponding sources are automatically enriched with quality information. The knowledge base (TBox) is available to users and detection and classification software, improving situational awareness
Processed data	AIS, Acoustic data, Thermal/optical data, Satellite images
Generated data	Populated ontology (TBox) or semantic repository
Related user stories	US_01, US_13, US_14, US_17, US_19, US_20

5 Datasets

In this section we provide an aggregated view of the available datasets that will be used in the use cases, among those identified by the user experts.

Table 5: Aggregated view of use case datasets

Kind of data sources	Format	Volume (approx.)	Purpose (task involved)	Tools used to process the data*	Automatic/Manual/Semi-automatic	Historical/ real-time	Update frequency
AIS data	Structured data	Approx. 1TB	Generate density maps, Create routes MSA/anti collision/ monitoring	SQL server, PostGIS, Python, Spark Human Operator/Combat Management System	Semi-automatic	Historical/ Real-time	Daily (Historical)/ continuous (Real-time)
Vessel statistics	Structured Data	300GB	Vessel static correction	SQL server, Python,	Semi-automatic	Historical	Daily
Kafka streams	Semi-structured	25GB	Generation of vessel metrics	Python	Automatic	Real-time	Sub-second
Kafka streams	Semi-structured	5GB	Generation of vessel metrics	Akka (https://akka.io) + Scala	Automatic	Real-time	Sub-second
Patterns of Life	Database Table	5 GB (10,000,000 polygons)	Anomaly detection	Akka (https://akka.io) + Scala	Automatic	Historical	Annually
Satellite image data (Sentinel-1 SAR, Sentinel-2 MSI)/ labelled ship target training data sets	GeoTiff, JP2, Tiff for image data/ xml, csv, json for metadata	Image data from 600MB to 2GB per image/Training data sets from 50MB to	Target detection and classification	Matlab, R, Python	Automatic	H/RT	Daily

Kind of data sources	Format	Volume (approx.)	Purpose (task involved)	Tools used to process the data*	Automatic/Manual/Semi-automatic	Historical/ real-time	Update frequency
		500MB depending on the number of targets Total size estimation: 350 GB/day					
Acoustic data from hydrophones towed by autonomous vehicle (detected target bearing, bearing uncertainty) data also include asset angles for the acoustic sensor; vehicle speed and position/heading; acoustic sensor position/heading (depth); information relative to target classification	Binary files storing acoustic data from hydrophones; XML files storing target bearing, associated uncertainty and vehicle and acoustic sensor positions/heading	Binary files: 10 MB/s XML files: 100 MB every hour of operation Total Size: ~400 GB	Target detection, localization and classification	Matlab, R, Python, C++	Automatic	H/RT	continuous
Vehicle status: position, heading, speed, battery level, next waypoint, etc	XML file containing the information relative to the vehicle status required to control its operation	100 MB every hour of operation	Vehicle control and supervision	Matlab, R, Python, C++	Automatic	H/RT	continuous

Kind of data sources	Format	Volume (approx.)	Purpose (task involved)	Tools used to process the data*	Automatic/Manual/Semi-automatic	Historical/ real-time	Update frequency
Visible/Thermal camera	PAL stream/MPEG4 or AVI digital stream	108 GB/day	Target detection and classification	Matlab, R, Python	Automatic	H/RT	continuous

5.1 Notes on Datasets for the Use Case

In this section we provide additional information on the data sources used in the maritime use case. In particular, the satellite, the acoustic and the video/thermal camera data formats used for vessel target detection localization and classification will be briefly described.

5.1.1 Satellite Data

Satellite data sets for the maritime use case are provided by the European Space Agency (ESA) Copernicus data Hub[6] and include:

- Sentinel-1 Synthetic Aperture Radar (SAR) images
- Sentinel-2 Multi-Spectral Instrument (MSI) images

In addition, derived data sets will be considered, including labeled ship target data sets used to train machine learning algorithms for target classification.

The Sentinel-1 satellite carries on board a C band multi-polarized SAR sensor operating at four different modes: strip map, interferometric wide swath, extra wide swath and wave mode. Each operational mode is characterized by its own viewing geometry resolution and polarization channels (named HH, VV, VH and HV channels). The Ground Range Detected (GRD) Level 1 data product is one of the possible format that will be utilized in the maritime use case. The product is provided for the channels VH and VV. For each channel the data are stored in a GeoTIFF format file on a grid with a pixel spacing of 10x10m. The size of the grid is variable depending on the acquisition and of the order of ten thousands of pixels in each direction. Metadata files in XML format are associated to each grid to provide, for example, the image geo-localization parameters. Figure 10 shows an example of SAR image in false colours.

6 https://scihub.copernicus.eu/

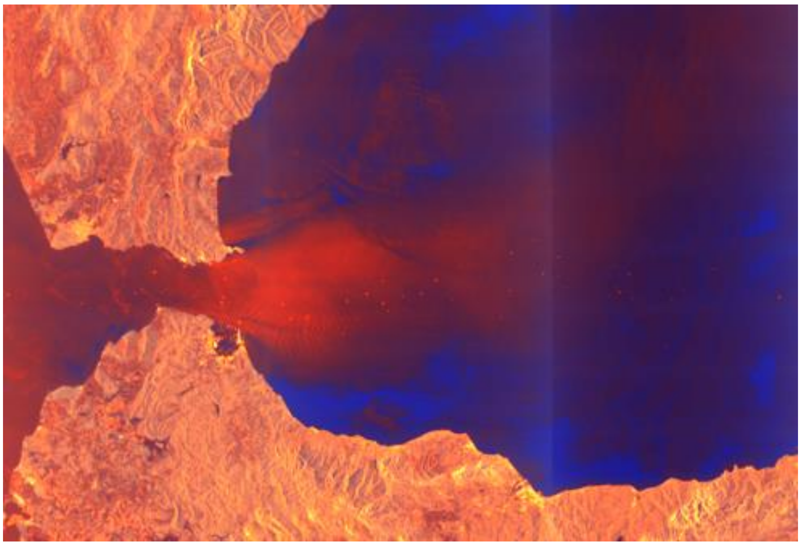

Figure 10: Example of ESA Sentinel-1 SAR image in false colours

The Sentinel-2 MSI acquires the solar radiation from the Earth surface and the atmosphere in 13 spectral bands (or channels): four bands at 10m, six bands at 20m and three bands at 60m spatial resolution. For the L1C product level [12], the image data of each band is provided as a geo-located regular grid (in UTM coordinates) in JP2000 uncompressed format on 16 bits, with auxiliary XML metadata files which include information such as the geo-localization grid parameters, the satellite ephemerides, the acquisition time, applied post-processing procedures if any, calibration coefficients and quality check. Figure 11 displays an example of an RGB composite of an MSI image on the Gibraltar-Strait. The image grid size depends on the channel resolution; at 10m of resolution, for instance, the image size is 10800x10800 samples.

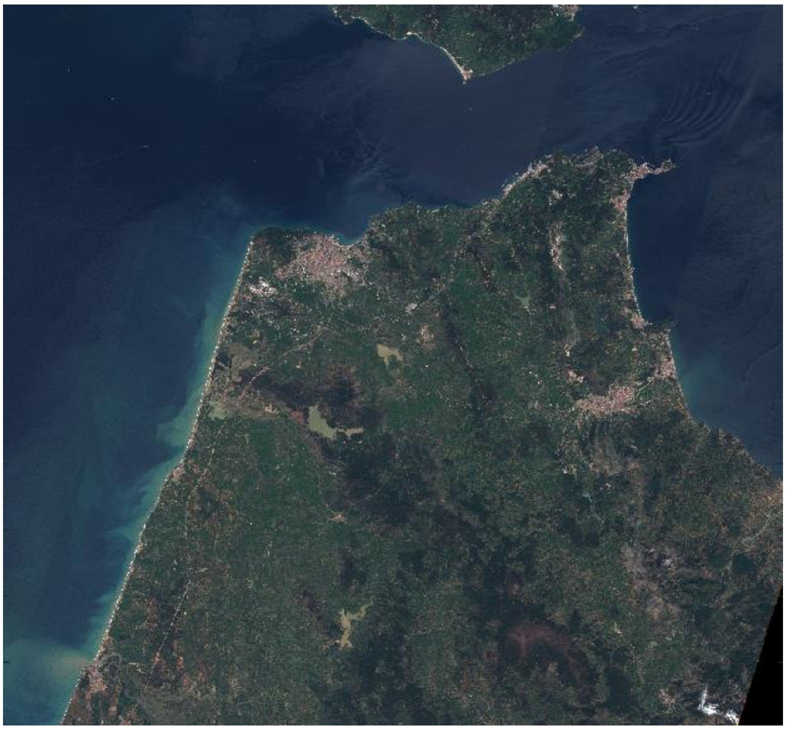

Figure 11: Example of Sentinel-2 MSI Red-Green-Blue (RGB) composite image.

Sentinel-1 SAR and Sentinel-2 MSI images from historical data sets are processed by automatic vessel target detection algorithms (see Figure 12, Figure 13). The detected targets are then associated to AIS tracks using time and positional information to produce labeled target data sets that are used to train vessel classification machine learning algorithms (e.g. using convolutional neural networks). Figure 14 displays an example of a Sentinel-1 SAR image with detected targets and associated AIS tracks. While Figure 15 shows an example of detected and labeled targets for both Sentinel-1 and Sentinel-2 sensors.

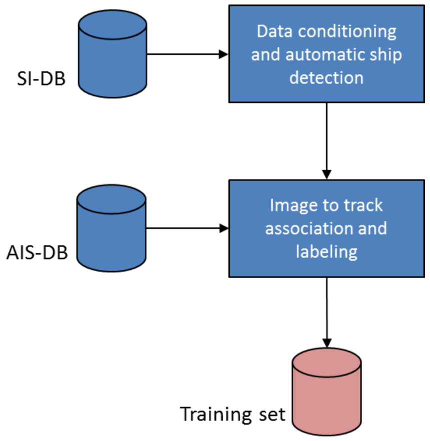

Figure 12: Satellite images and AIS data from historical databases (SI-DB, AIS-DB, respectively) are processed to build a labelled data set of ship targets to train machine learning classification algorithms

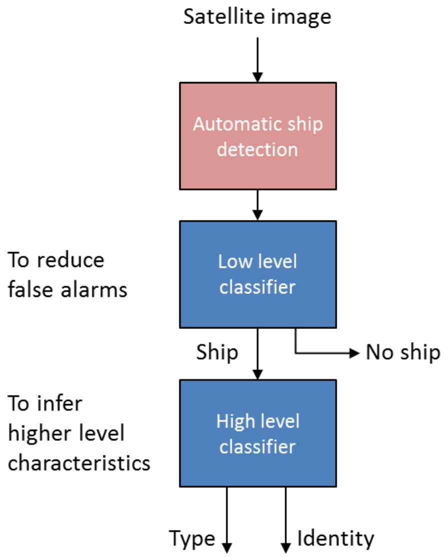

Figure 13: Classification algorithms are used to process new images in order to reduce false alarms and extract information about the identity of a vessel

The format of the training data set for the Sentinel-2 MSI sensor is as follows. The data are in a single folder. In particular, a .csv file stores a table of global metadata. Each row of the table is a target associated to an AIS track. The table includes information about the vessel and a field that point to a .json file storing local metadata associated to the target. The local metadata .json file includes parameters about the sensor, the position of the target and the number of spectral channels associated to the target. For each target there are K images in .tiff format where K is the number of channels considered in the processing (from a minimum of 1 to a maximum of 13 spectral channels). The name of each file is stored in the .json local metadata file.

It is worthwhile to mention that this format for labelled data sets can be a reference scheme for other kind of sensors used for target detection and classification, including SAR, acoustic and video ones.

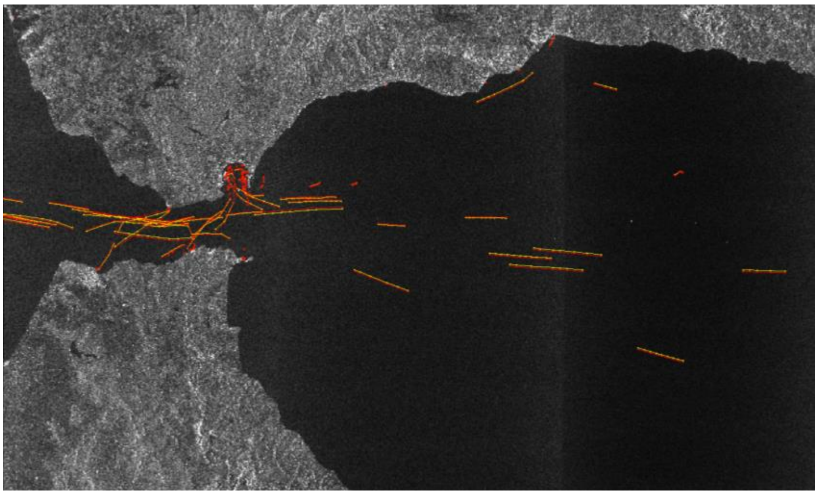

Figure 14: Examples of Sentinel-1 SAR image and detected targets with associated AIS tracks

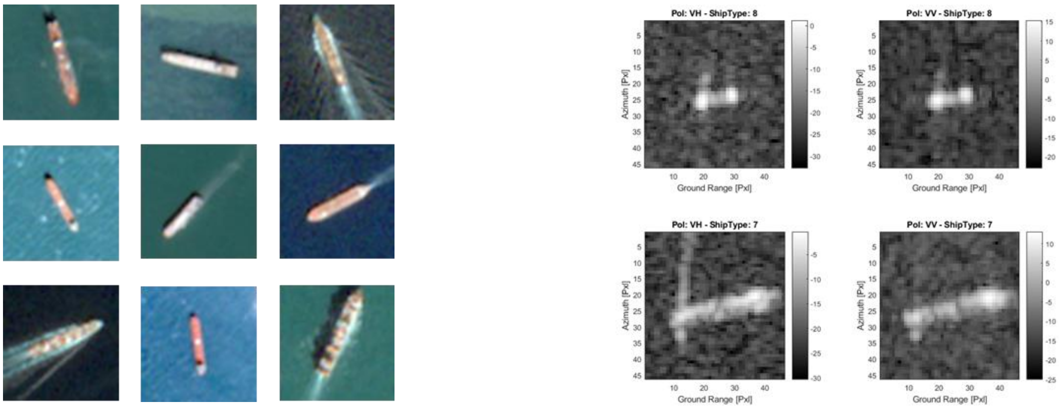

Figure 15: Examples of Sentinel-2 MSI target image chips (RGB composite for visualization purposes) and Sentinel-1 multi-polar SAR target chips.

5.1.2 Acoustic Data

The acoustic data are acquired by arrays of passive hydrophones towed by autonomous surface vehicles to detect, localize and classify targets of interest. The acoustic data are processed on board the vehicles to extract target bearing and associated uncertainty data. The bearing and the bearing uncertainty for each detected target, together with the position of the vehicles, are continuously exchanged between vehicles and a fusion centre for further processing. The vehicles also exchange some results from a first target classification based on the acoustic data. The produced bearing measurements are the input on one side to cooperative target localization behaviours running on the vehicles, on the other are used at a fusion centre to compute the probabilities of the presence of targets in the monitored area. This processed information can provide hints to the thermal camera data on specific areas of interest to be further imaged. Furthermore, this information can be fused with the output of the thermal camera data for target confirmation and classification.

5.1.3 Thermal Camera Data

Thermal video camera data will be acquired for target detection and classification in a spectral range between 3.6 and 4.9 µm using a 640x480 pixel Indium Antimonide (InSb) Focal Plane Array (FPA) detector. The camera is able to provide a continuous composite video PAL or NTSC stream which can be converted in a digital stream (e.g. MPEG4 or AVI) for further processing including automatic target detection and classification.

6 Expected Benefits from INFORE

Table 6: Summary of expected benefits from INFORE as stated by interviewed expert users

Role	Expected benefit from INFORE	Functional/ non-functional
MDA data scientist	Accurate calculation of metrics in the context of real-time detection of anomalies in a vessel's trajectory, such as the deviations in the arrival time of a vessel at a port (ETA).	Functional
MDA data scientist, researcher	Data fusion coming from various sources, such as satellite data, market data (e.g., freight rates, cargo carried).	Functional
MDA data scientist	Ability to design, implement and fine-tune workflows that involve data processing tasks and deliver results in real-time. Automatic fine-tuning of machine learning algorithms used in the workflows would be a desirable feature.	Functional
MDA data scientist	Ability to interactively explore the data in order to detect patterns/features of interest	Functional
MDA data scientist	Ability to automatically optimise data analysis tasks over different data processing platforms	Functional
MDA data scientist	Ability to receive quick approximate answers instead of 100% accurate, but long running queries (depending on the reduction in accuracy)	Non- Functional
MDA data scientist	Ability to use INFORE platform as a unified data processing environment instead of using multiple systems and programming languages for a single workflow	Functional
MDA data scientist	Ability to use a collaborative platform to share results among colleagues	Functional
MDA data scientist	Support for open-source custom solutions	Functional
MDA data scientist	Fast prototyping	Non-functional
MDA data scientist	Ability to design, implement and fine tuning workflows using a graphical interface, but not in the expense of extensibility, i.e., ability implement their own data processing operators and integrate them into custom workflows.	Functional
MDA data scientist	Ability to process large volumes of vessel positions, execute analytics workflows and deliver results real-time.	Non-functional
Commander	Ability to get more time and useful data for an assessment and decisions how to proceed (courses of action) by reducing the preparation work	Functional
Commander, Captain, Operational officer	Ability to have a real time help to maritime and harbours surveillance, and to allow users to have reliable, real time answers to their problems from the BIG DATA picture	Functional
MDA researcher	Real-time detection of dark targets involved in illegal activities, such as smuggling, trafficking, illegal fishing via fusion of information from multiple sources	Functional

7 Conclusions

In this deliverable, we described the user requirements for the maritime situational awareness use case of the INFORE project, as derived by the replies to the questionnaires that were completed by 3 expert users from MarineTraffic's network and 4 expert users from CMRE's network. This work is aligned with the work described in the WT 3.1 of the project. We also described the definition and the requirements of the sea experiment, in the context of WT 3.2 of the project.

The next version of this deliverable in M18, will contain updated information on the user requirements, which will take into account the evaluation results of the developed prototype in the context of the work described in WT3.3.

8 References

[1] "US National Strategy for Maritime Security: National Maritime Domain Awareness Plan," 2017.

[2] European Defence Agency, "The EU Capability Development Priorities. 2018 Revision," Brussels-Belgium, 2018.

[3] "Anomaly Detection in the Maritime Domain: Comparison of Traditional and Big Data Approach," in *Big Data and Artificial Intelligence for Military Decision Making.*

[4] Enrica d'Afflisio, Paolo Braca, Leonardo M. Millefiori, Peter Willett: Maritime Anomaly Detection Based on Mean-Reverting Stochastic Processes Applied to a Real-World Scenario. FUSION 2018: 1171-1177

[5] Fabio Mazzarella, Michele Vespe, Alfredo Alessandrini, Dario Tarchi, Giuseppe Aulicino, Antonio Vollero: A novel anomaly detection approach to identify intentional AIS on-off switching. Expert Syst. Appl. 78: 110-123 (2017)

[6] Maria Riveiro, Giuliana Pallotta, Michele Vespe: Maritime anomaly detection: A review. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 8(5) (2018)

[7] A. Margara, G. Cugola, and G. Tamburrelli. Learning from the past: automated rule generation for complex event processing. In The 8th ACM International Conference on Distributed Event-Based Systems (DEBS). Mumbai, India, May 26-29, 2014, 2014, pp. 47–58.

[8] J. Gama and M. M. Gaber, Learning from Data Streams: Processing Techniques in Sensor Networks, 1st ed. Springer Publishing Company, Incorporated, 2010.

[9] Eriksen, C.C., Osse, T.J., Light, R.D., Wen, T., Lehman, T., Sabin, P.L., Ballard, J.W., & Chiodi, A.M. (2001). Seaglider : A Long-Range Autonomous Underwater Vehicle for Oceanographic Research. IEEE Journal of Oceanic Engineering, vol. 26, no. 4, pp. 424–436, October 2001.

[10] T. Fossum. Intelligent autonomous underwater vehicles. Norwegian University of Science and Technology. Technical Report, 2016.

[11] G. Ferri et al. Cooperative robotic networks for underwater surveillance: an overview. IET Radar Sonar And Navigation, vol. 11, no. 12, pp. 1740–1761, Jul. 2017.

[12] Sentinel 2 ESA mission website, https://sentinel.esa.int/web/sentinel/missions/sentinel-2. Accessed: 2018-12-07.

Appendix A. INFORE- User Requirements Questionnaire

1. User background information

- Company/organisation
- Professional position [and years of experience]
- Domain of expertise
- Background studies (university degree major, etc.)
- What are your main job tasks?
- To whom are you responsible for performing these tasks?

2. Existing workflow

- Please describe the different kinds of data sources that you use in your day-to-day tasks and the tools that you use to process them

Kind of data sources	Format	Volume (approx.)	Purpose (task involved)	Tools used to process the data*	Automatic/Manual/Semi-automatic	Historical/ real-time	Update frequency

*If custom programs are used to process the data, please mention the programming language.

- Which is the aim of the analysis you perform (what kind of insights do you try to find)?
- What data processing challenges do you experience in your day-to-day tasks (e.g., fusion of heterogeneous sources, performance, analytics)?
- Provide examples of use/ case studies
- What problems do you run into in your day-to-day work when performing your data analysis? Is there a standard way of solving each of them, or do you have a workaround?
    - Why is this a problem?
    - How do you currently solve the problem?
    - How would you ideally like to solve the problem?
- Is any of the tools, mentioned in the table above, a must (one that no alternative execution on other tools/platforms would be allowed) for the case studies that you describe?
- Are you able to program/set up a new/custom data processing workflow?
- How long does it take to program/set up a new data processing workflow?
- Are you capable of optimising your data analysis operations?
- On what kind of infrastructure do you usually run the analysis (e.g. personal laptop, high spec workstation, server, cluster, HPC etc)

3. Expected benefits from using INFORE

- A. Please mention more data sources that would you like to use and why.
- B. Which new information would you like to extract from these (old and new) data sources?
- C. Are there specific events that you would like to forecast in real-time, which you currently cannot forecast?
- D. Would you find it an acceptable trade-off to significantly speed up your data analysis tasks, if the provided output was a fairly accurate approximation of the correct result?
- E1. Rate the following objectives of INFORE, based on how useful they may be at YOUR data analysis (1: Not useful, 2: Little Use, 3: Average Use, 4: Quite useful, 5: Very useful)
  - Ability to design data processing workflows with no code required
  - Ability to change algorithm parameters graphically
  - Ability to receive quick approximate answers instead of 100% accurate, but long running queries
  - Ability to interactively explore the data in order to detect patterns/features of interest
  - Ability to accurately forecast events of interest
  - Ability to automatically optimise your data analysis task over different data processing platforms (HPC, Big Data Platforms, etc).
- E2. Rate the following objectives of INFORE, based on how useful they may be at the data analysis of OTHER data analysts in your organization (1: Not useful, 2: Little Use, 3: Average Use, 4: Quite useful, 5: Very useful)
  - Ability to design data processing workflows with no code required
  - Ability to change algorithm parameters graphically
  - Ability to receive quick approximate answers instead of 100% accurate, but long running queries
  - Ability to interactively explore the data in order to detect patterns/features of interest
  - Ability to accurately forecast (defined or currently unknown) events of interest
  - Ability to automatically optimise your data analysis task over different data processing platforms (HPC, Big Data Platforms, etc).
- F. What are your expectations regarding the system usability?
- G. What is the expected added value from INFORE for you and your corporation?

Appendix B.1 User 1 - Answers to Questionnaire

1. User background information

- Professional position
  Navy Officer on Commander level
- Domain of expertise
  Maritime Operations / Operational Officer
- What are your main job tasks?
  1. Planning and conduct of Maritime Operations
  2. Procurement process in support of these operations

2. Existing workflow

- Please describe the different kinds of data sources that you use in your day-to-day tasks and the tools that you use to process them

Kind of data sources	Format	Volume (approx.)	Purpose (task involved)	Tools used to process the data*	Automatic/Manual/Semi-automatic	Historical/ real-time	Update frequency
Personal Computer	MS		Daily workflow	Excel / word / Lotus Notes			
Web	Mil or civilian		Collect Information	Browser			

Rem.: These are the main tools on all levels for the daily workload – see also below "examples of use" are describing more specific tools / databases for different purpose on different levels

*If custom programs are used to process the data, please mention the programming language

- Which is the aim of the analysis you perform (what kind of insights do you try to find)?

  (The answer to this question was hidden from the public document for ethics reasons as it could possibly reveal the identity of the user)

- What data processing challenges do you experience in your day-to-day tasks (e.g., fusion of heterogeneous sources, performance, and analytics)?

  The daily challenge is always to keep the overview to the whole process without going too far to details (Situation might change quick and mostly operators ashore or at sea are limited by time within the battle-rhythm).
  1. Within the maritime domain, always depending from the command level to find the appropriate, reliable and supporting data bases (typical examples: e.g. web - email - structured or formatted message forms, MCCIS, COP, LINK, AIS, Radar, Operational Orders, etc.) and to merge them into the common picture

(on different levels). The daily problem is not simply to add new data but to assess / filter / prioritize / merge etc. and to get more specific / linked data e.g. on the spot which is relevant for the conduct of operation (merged layers on different levels). Current systems cannot be adapted to the requirements, are overwhelming the user with (mostly) unneeded information or are too complicate to handle, a lot of information is not linked or even more not suggested (e.g. intelligent network)
2. Within the procurement process again the above mentioned need for situational awareness / keeping the overview in order not be chased by the process or fast development but instead planning the next steps with sufficient time.
The procurement process is less flexible – normally there are typical benchmarks for different phases on a timeline. The need is here more on managing that time line, budget, big data bases for material, documents like specifications – requirements – suggested or possible solutions which have to be handled – compared – assessed for a final decision.

- Provide examples of use/ case studies

COP	*common operational picture*
MCCIS	*common operational picture on maritime level*
LINK	*tactical data link on tactical level for different systems / purposes*
CIC	*Combat Information Centre (on different levels)*
AIS	*Civilian Information System (ships traffic)*

- What problems do you run into in your day-to-day work when performing your data analysis? Is there a standard way of solving each of them, or do you have a workaround?
  The latter - current tools are not really supporting / linking / connecting and instead simply acting as a database (e.g. to find referenced main documents or POC's)
  Normally they are not flexible, access is limited or to adapted to specific task (e.g. adding applications or merging of information not feasible)
  - Why is this a problem?
    The work happens not online with or supported by the system – instead normally offline without any connectivity by e.g. simply using or programming Excel / word / common tools

  - How do you currently solve the problem?
    Working in a time-consuming and insufficient way by using simple commercial tools e.g. to collect or to display information, creating long lists in order to merge information from different data bases
  - How would you ideally like to solve the problem?
    A System which works on different layers (for Command levels) which supports the daily work by managing data, timelines etc. by e.g. collecting, merging, linking but also filtering information. System should be flexible, adaptable in cooperation with other tools / systems / databases.
- Is any of the tools, mentioned in the table above, a must (one that no alternative execution on other tools/platforms would be allowed) for the case studies that you describe?
  No
- Are you able to program/set up a new/custom data processing workflow?
  Not really
- How long does it take to program/set up a new data processing workflow?
  Unknown
- Are you capable of optimizing your data analysis operations?
  No
- On what kind of infrastructure do you usually run the analysis (e.g. personal laptop, high spec workstation, server, cluster, HPC etc)

Personal laptop (restricted / unrestricted networks)

3. Expected benefits from using INFORE

- A. Please mention more data sources that would you like to use and why.
  Implementation, merging of existing data bases for further exploitation
- B. Which new information would you like to extract from these (old and new) data sources?
  Comparison of datasets, warnings, proposals, suggestions, conflicts
- C. Are there specific events that you would like to forecast in real-time, which you currently cannot forecast?
  Maritime Operations: Maritime traffic, environmental conditions and impact on operations, detecting anomaly behavior, forecasted sensor ranges to assess own capabilities and impact on tasking (area coverage)
- D. Would you find it an acceptable trade-off to significantly speed up your data analysis tasks, if the provided output was a fairly accurate approximation of the correct result?
  Yes – having an estimated error in mind. Calculating the most probable solution would support if there are simply two ways to go.
- E1. Rate the following objectives of INFORE, based on how useful they may be at YOUR data analysis
  (1: Not useful, 2: Little Use, 3: Average Use, 4: Quite useful, 5: Very useful)
  - 4 Ability to design data processing workflows with no code required
  - 5 Ability to change algorithm parameters graphically
  - 5 Ability to receive quick approximate answers instead of 100% accurate, but long running queries
  - 5 Ability to interactively explore the data in order to detect patterns/features of interest
  - 2 Ability to accurately forecast events of interest
  - 5 Ability to automatically optimize your data analysis task over different data processing platforms (HPC, Big Data Platforms, etc.).

  Rem.: (Personal point of view depending from task level)
- E2. Rate the following objectives of INFORE, based on how useful they may be at the data analysis of OTHER data analysts in your organization
  (1: Not useful, 2: Little Use, 3: Average Use, 4: Quite useful, 5: Very useful)
  Rem.: No assessment made. Depends very much from the task level
  - Ability to design data processing workflows with no code required
  - Ability to change algorithm parameters graphically
  - Ability to receive quick approximate answers instead of 100% accurate, but long running queries
  - Ability to interactively explore the data in order to detect patterns/features of interest
  - Ability to accurately forecast (defined or currently unknown) events of interest
  - Ability to automatically optimize your data analysis task over different data processing platforms (HPC, Big Data Platforms, etc.).
- F. What are your expectations regarding the system usability?
  Simple to handle – flexible – adaptable – merging – warning - proposing
- G. What is the expected added value from INFORE for you and your corporation?
  To get more time and useful data for an assessment and decisions how to proceed (courses of action) by reducing the preparation work.

Appendix B.2 User 2 - Answers to Questionnaire

1. User background information

- Professional position: RESEARCH VESSELS CAPTAIN/ SHIP MANAGER
- Domain of expertise: MARITIME (MILITARY AND COMMERCIAL)
- What are your main job tasks? PLANNING AND EXECUTION OF SAFE SCIENCE AT SEA WITH OUR VESSELS – SHIP MANAGEMENT, OPERATION

2. Existing workflow

- Please describe the different kinds of data sources that you use in your day-to-day tasks and the tools that you use to process them. *N/A*

Kind of data sources	Format	Volume (approx.)	Purpose (task involved)	Tools used to process the data*	Automatic/Manual/Semi-automatic	Historical/ real-time	Update frequency

*If custom programs are used to process the data, please mention the programming language.

- Which is the aim of the analysis you perform (what kind of insights do you try to find)?
  Potential anomalies for the maritime surveillance, increase of safety of navigation
- What data processing challenges do you experience in your day-to-day tasks (e.g., fusion of heterogeneous sources, performance, analytics)? *N/A*
- Provide examples of use/ case studies:
  During some particular activities in the Navy service, such as "embargos" or "Choke Point Maritime Surveillance" we used to have an integrated system where all information and sensor information were mixed together in order to have the best possible surface scenario to allow the officer responsible to determine whether a vessel was to be classified as "Friend" or "Foe". This was during the 90's where most of the Big Data were not 100% digital but half manual processed.
  During the time spent as Captain of Research Vessels, doing science at sea, we used to have an ECDIS (Electronic Chart Display and Integrated System) on the bridge, where all digital information from AIS, RADARs, ship's meteo and acoustic sensors were displayed automatically on the screen in order to have an updated and live representation of the surface picture. Filter to enrich or decrease the information flow in the system were set by the operator. This was not a surveillance tool but a navigational aid, only.
- What problems do you run into in your day-to-day work when performing your data analysis? Is there a standard way of solving each of them, or do you have a workaround? *N/A*
  - Why is this a problem?
  - How do you currently solve the problem?
  - How would you ideally like to solve the problem?
- Is any of the tools, mentioned in the table above, a must (one that no alternative execution on other tools/platforms would be allowed) for the case studies that you describe? *N/A*

- Are you able to program/set up a new/custom data processing workflow? *No, but I may help with my at sea background and experience*
- How long does it take to program/set up a new data processing workflow? *N/A*
- Are you capable of optimizing your data analysis operations?
- On what kind of infrastructure do you usually run the analysis (e.g. personal laptop, high spec workstation, server, cluster, HPC etc) *With reference to my experience, they were high spec workstation but with no artificial intelligence (not capable to analyze and customize the result for the operator)*

3. Expected benefits from using INFORE

- A. Please mention more data sources that would you like to use and why. *N/A*
- B. Which new information would you like to extract from these (old and new) data sources? *N/A*
- C. Are there specific events that you would like to forecast in real-time, which you currently cannot forecast? *Meteorological events which could jeopardize the surface picture*
- D. Would you find it an acceptable trade-off to significantly speed up your data analysis tasks, if the provided output was a fairly accurate approximation of the correct result? *YES*
- E1. Rate the following objectives of INFORE, based on how useful they may be at YOUR data analysis (1: Not useful, 2: Little Use, 3: Average Use, 4*: Quite useful*, 5: Very useful)
  - Ability to design data processing workflows with no code required *3*
  - Ability to change algorithm parameters graphically *4*
  - Ability to receive quick approximate answers instead of 100% accurate, but long running queries *4*
  - Ability to interactively explore the data in order to detect patterns/features of interest *4*
  - Ability to accurately forecast events of interest *4*
  - Ability to automatically optimise your data analysis task over different data processing platforms (HPC, Big Data Platforms, etc). *4*
- E2. Rate the following objectives of INFORE, based on how useful they may be at the data analysis of OTHER data analysts in your organization (1: Not useful, 2: Little Use, 3: Average Use, 4: Quite useful, 5: Very useful)
  - Ability to design data processing workflows with no code required *3*
  - Ability to change algorithm parameters graphically *4*
  - Ability to receive quick approximate answers instead of 100% accurate, but long running queries *4*
  - Ability to interactively explore the data in order to detect patterns/features of interest *4*
  - Ability to accurately forecast (defined or currently unknown) events of interest *4*
  - Ability to automatically optimise your data analysis task over different data processing platforms (HPC, Big Data Platforms, etc). *4*
- F. What are your expectations regarding the system usability? *To allow users to have reliable, real time answers to their problems from the BIG DATA picture.*
- G. What is the expected added value from INFORE for you and your corporation? *To have a real time help to maritime and harbors surveillance*

Appendix B.3 User 3 - Answers to Questionnaire

1. User background information

- Professional position [and years of experience]:
  - *Officer*
- Domain of expertise:
  - *Maritime, Navigation*
- What are your main job tasks?
  - *Operational officer, Navigation, Security.*

2. Existing workflow

- Please describe the different kinds of data sources that you use in your day-to-day tasks and the tools that you use to process them

Kind of data sources	Format	Volume (approx.)	Purpose (task involved)	Tools used to process the data*	Automatic/Manual/Semi-automatic	Historical/ real-time	Update frequency
AIS Receiver	*AIS data*	*Pending Merchant traffic in the area ~ 50 CTCs/h*	*MSA/anti collision/ monitoring*	*Human Operator/Combat Management System*	*Manual*	*Real-time*	*continuous*
Radar	*RDR CTC*	*Pending traffic in the area ~ 75 CTCs/h*	*MSA/anti collision/ monitoring*	*Human Operator/Combat Management System*	*Manual*	*Real-time*	*continuous*
ESM	*ESM*	*Pending traffic in the area ~ 50 CTCs/h*	*MSA/anti collision/ monitoring*	*Human Operator/Combat Management System*	*Manual*	*Real-time*	*continuous*
Intel Report	*Text*	*Pending CTCs of interest not valuable*	*MSA/monitoring*	*Human Operator*	*Manual*	*Historical*	*~ Every 4 hrs*
Web based AIS data collection (e.g. MARINE TRAFFIC)	*AIS*		*MSA/monitoring*	*Human Operator*	*Manual*	*Historical/ real-time*	*continuous*
Radio VHF Hailing	*Voice*	*~ 5 hailing/h*	*Monitoring /checking*	*Human Operator*	*Manual*	*real-time*	

Visual	*Visual*	*Pending on traffic ~ 10/h*	*MSA/anti collision/ monitoring*	*Human Operator*	*Manual*	*real-time*	*continuous*

*If custom programs are used to process the data, please mention the programming language.

- Which is the aim of the analysis you perform (what kind of insights do you try to find)?
  - *Coherence with normal POL in the area, counter smuggling or illicit activity*
- What data processing challenges do you experience in your day-to-day tasks (e.g., fusion of heterogeneous sources, performance, analytics)?
  - *Human factor always predominant*
- Provide examples of use/ case studies
  - *Maritime Surveillance Operations*
- What problems do you run into in your day-to-day work when performing your data analysis? Is there a standard way of solving each of them, or do you have a workaround? *Having enough human resources no problem can be identified.*
  - Why is this a problem?
  - How do you currently solve the problem?
  - How would you ideally like to solve the problem?
- Is any of the tools, mentioned in the table above, a must (one that no alternative execution on other tools/platforms would be allowed) for the case studies that you describe?
  - *Combat Management System is crucial and mission critical.*
- Are you able to program/set up a new/custom data processing workflow?
  - *Human factor: yes.*
  - *On CMS (Automatic): no.*
- How long does it take to program/set up a new data processing workflow?
  - *Few hours*
- Are you capable of optimising your data analysis operations?
  - *Human factor: yes.*
  - *On CMS (Automatic): no.*
- On what kind of infrastructure do you usually run the analysis (e.g. personal laptop, high spec workstation, server, cluster, HPC etc):
  - *High spec workstation, server, Combat Management System.*

3. Expected benefits from using INFORE

- A. Please mention more data sources that would you like to use and why.
  - *At the moment and from my knowledge no more data sources would be available. If any other is identified acquiring data would be vital to fill the possible capability gap with other navy's.*
- B. Which new information would you like to extract from these (old and new) data sources?
  - *See above*
- C. Are there specific events that you would like to forecast in real-time, which you currently cannot forecast?
- D. Would you find it an acceptable trade-off to significantly speed up your data analysis tasks, if the provided output was a fairly accurate approximation of the correct result?
  - *Knowing the level of accuracy and being capable to choose it: Yes.*
- E1. Rate the following objectives of INFORE, based on how useful they may be at YOUR data analysis (1: Not useful, 2: Little Use, 3: Average Use, 4: Quite useful, 5: Very useful)
  - Ability to design data processing workflows with no code required 5

- ○ Ability to change algorithm parameters graphically 4
- ○ Ability to receive quick approximate answers instead of 100% accurate, but long running queries 5, *but always knowing the level of accuracy.*
- ○ Ability to interactively explore the data in order to detect patterns/features of interest 5
- ○ Ability to accurately forecast events of interest 5
- ○ Ability to automatically optimise your data analysis task over different data processing platforms (HPC, Big Data Platforms, etc). 4

- E2. Rate the following objectives of INFORE, based on how useful they may be at the data analysis of OTHER data analysts in your organization (1: Not useful, 2: Little Use, 3: Average Use, 4: Quite useful, 5: Very useful)
  - ○ Ability to design data processing workflows with no code required 5
  - ○ Ability to change algorithm parameters graphically 4
  - ○ Ability to receive quick approximate answers instead of 100% accurate, but long running queries 5
  - ○ Ability to interactively explore the data in order to detect patterns/features of interest 5
  - ○ Ability to accurately forecast (defined or currently unknown) events of interest 5
  - ○ Ability to automatically optimise your data analysis task over different data processing platforms (HPC, Big Data Platforms, etc). 4
- F. What are your expectations regarding the system usability?
  - *User interface should be really simple and intuitive allowing new operator to learn advanced skills in short time.*
- G. What is the expected added value from INFORE for you and your corporation?

Appendix B.4 User 4 - Answers to Questionnaire

1. User background information

- Professional position [and years of experience]:
  - *Officer*
- Domain of expertise:
  - *Naval Operations, Above Water Warfare*.
- What are your main job tasks? Relevant to the questionnaire
  - *Principal Warfare Officer, Commanding Officer.*

2. Existing workflow

- Please describe the different kinds of data sources that you use in your day-to-day tasks and the tools that you use to process them

Kind of data sources	Format	Volume (approx.)	Purpose (task involved)	Tools used to process the data*	Automatic/Manual/Semi-automatic	Historical/ real-time	Update frequency
AIS Receiver	AIS data	Highly variable dependent on marine traffic in the area ~ 50 CTCs/h	MSA/anti-collision	Combat Management System/operator	Automatic	Real-time	continuous
Radar (including organic assests like Helos)	RDR CTC	Dependent on traffic in the area ~ 70 CTCs/h	RMP/anti-collision	Combat Management System/operator	automatic	Real-time	continuous
Tactical data link networks	CTCs	Strongly dependent on Networks 50ctc/h	RMP	Combat Management System/operator	Semiauto	Real- time	continuous
Maritime Command and Control Information System **(MCCIS)**	CTC	Dependent on scenario 50 ctc/h	MSA/RMP	Operator	Manual	NRT	continuous
SMART (Service oriented infrastructur	CTC/anomaly detection	Dependent on scenario 50 ctc/h	MSA	Operator	Manual	NRT	continuous

e for Maritime traffic Tracking)							
ESM	ESM	Dependent on traffic in the area ~ 60 CTCs/h	RMP/anti collision/ monitoring	Human Operator	Manual	Real-time	continuous
Intel Report	Text	Dependent on CTCs of interest.	MSA/moni toring	Human Operator	Manual	Historical	daily
Web based AIS data collection (e.g. MARINE TRAFFIC)	AIS		MSA	Human Operator	Manual	Historical/ real-time	continuous
Metoc Support	Reports		Decision support	Human Operator	manual	NRT	Daily
Radio VHF Hailing	Voice	~ 5 hailing/h	Surveillanc e-verify informatio n	Human Operator	Manual	real-time	
Visual	Visual	Dependent on traffic ~ 10/h	RMP/anti collision	Human Operator	Manual	real-time	continuous

*If custom programs are used to process the data, please mention the programming language.

- Which is the aim of the analysis you perform (what kind of insights do you try to find)?
    - *Achieve and maintain a timely, reliable, accurate Recognized Maritime Picture functional to combat threats such as human smuggling, terrorism at sea, piracy, as well as arms and drug trafficking*.
- What data processing challenges do you experience in your day-to-day tasks (e.g., fusion of heterogeneous sources, performance, analytics)?
    - *Achieve RMP means: detect objects of interest (ships, aircrafts), classify them, assess what they are doing, and decide whether any type of follow-on action is required. Fusion of heterogeneous sources and availability of information (e.g., Small craft identification in coastal areas) are two typical challenges);*
- Provide examples of use/ case studies
    - *Maritime Interdiction Operations or Naval Blockade Operations such as Sharp Guard; Maritime Security Operations;*
- What problems do you run into in your day-to-day work when performing your data analysis? Is there a standard way of solving each of them, or do you have a workaround? *Reliability of AIS Information/need to increase reliance/verify information through automated multiple source data fusion (Radar/ESM/AIS/Visual ID/POL…)*
    - Why is this a problem? *Process not automated*

    - How do you currently solve the problem? *Almost entirely Manually*
    - How would you ideally like to solve the problem? *New tools*
- Is any of the tools, mentioned in the table above, a must (one that no alternative execution on other tools/platforms would be allowed) for the case studies that you describe? *Radars, Intel reports.*
- Are you able to program/set up a new/custom data processing workflow? N.A.
- How long does it take to program/set up a new data processing workflow? N.A.
- Are you capable of optimising your data analysis operations? N.A.
- On what kind of infrastructure do you usually run the analysis (e.g. personal laptop, high spec workstation, server, cluster, HPC etc) *Ship's Combat Management System*

3. Expected benefits from using INFORE

- A. Please mention more data sources that would you like to use and why.
- B. Which new information would you like to extract from these (old and new) data sources?
- C. Are there specific events that you would like to forecast in real-time, which you currently cannot forecast?
- D. Would you find it an acceptable trade-off to significantly speed up your data analysis tasks, if the provided output was a fairly accurate approximation of the correct result?
- E1. Rate the following objectives of INFORE, based on how useful they may be at YOUR data analysis (1: Not useful, 2: Little Use, 3: Average Use, 4: Quite useful, 5: Very useful)
    - Ability to design data processing workflows with no code required
    - Ability to change algorithm parameters graphically
    - Ability to receive quick approximate answers instead of 100% accurate, but long running queries
    - Ability to interactively explore the data in order to detect patterns/features of interest
    - Ability to accurately forecast events of interest
    - Ability to automatically optimise your data analysis task over different data processing platforms (HPC, Big Data Platforms, etc).
- E2. Rate the following objectives of INFORE, based on how useful they may be at the data analysis of OTHER data analysts in your organization (1: Not useful, 2: Little Use, 3: Average Use, 4: Quite useful, 5: Very useful)
    - Ability to design data processing workflows with no code required
    - Ability to change algorithm parameters graphically
    - Ability to receive quick approximate answers instead of 100% accurate, but long running queries
    - Ability to interactively explore the data in order to detect patterns/features of interest
    - Ability to accurately forecast (defined or currently unknown) events of interest
    - Ability to automatically optimise your data analysis task over different data processing platforms (HPC, Big Data Platforms, etc).
- F. What are your expectations regarding the system usability?
- G. What is the expected added value from INFORE for you and your corporation?

Appendix B.5 User 5 - Answers to Questionnaire

1. User background information

- Professional position [and years of experience]: Data Scientist (3 years)
- Domain of expertise: Maritime Analytics
- Background studies (university degree major, etc.): MEng in Mechanical Engineering, MBA, MSc in Business Analytics
- What are your main job tasks?: Build products based on maritime data (primarily AIS) that offer insights and intelligence to the maritime industry

2. Existing workflow

- Please describe the different kinds of data sources that you use in your day-to-day tasks and the tools that you use to process them

Kind of data sources	Format	Volume (approx.)	Purpose (task involved)	Tools used to process the data*	Automatic/Manual/Semi-automatic	Historical/ real-time	Update frequency
AIS data	Tabular	~3.5B records per semester (~300GB DB storage space)	Build Maritime Intelligence products	SQL, Python		Historical	1 per day
AIS data	Stream		Build consumers that implement the business logic behind in the product	Python		Real-Time	

*If custom programs are used to process the data, please mention the programming language.

- Which is the aim of the analysis you perform (what kind of insights do you try to find)?: *Provide the industry with tools that can be used to assess supply and demand of vessels based on monitoring historic fluctuations, provide situational awareness based on (near) real-time AIS data, and predict future events based on historic AIS data.*
- What data processing challenges do you experience in your day-to-day tasks (e.g., fusion of heterogeneous sources, performance, analytics)?: *Assessment of data sources that go beyond AIS, like market and cargo data. Also, need to automate or minimize hurdle in setting up and using cloud infrastructure for demanding ML projects.*
- Provide examples of use/ case studies: *AIS (acronym for Automatic Identification System) is a globally adopted protocol of communication between commercial vessels as well as between a vessel and a transponder installed ashore. Among the*

various figures contained in AIS signals is the vessel's next port of destination as well an estimate of the time of arrival there (ETA), as these are being declared by the vessel's Chief Engineer when departing for a new voyage. For a number of reasons reported ETA is far from accurate in practice and several attempts have been made in the past to algorithmically improve its accuracy.

In one of those attempts, we attempted to build a machine-learning regression model based on historic voyage data over a period of one year for a specific vessel market segment. The model learned to estimate vessel speed at a specific point along the vessel's chosen route based on a number of weather characteristics at this point as well as market specific features and was then used to provide point speed estimates along a predefined number of waypoints. This in turn translated into estimations of voyage duration between consecutive waypoints and, thus, into indirectly estimating the time of arrival at the destination port.

Model performance was not satisfactory, mainly due to lack of market data that we believe could greatly improve predictive power. Also, deploying into production would have resulted in significant prediction-time cost given that the model should run on batches of thousands of vessels/voyages, estimating speeds at 25-30 waypoints along each voyage. This would be worked around by using cloud infrastructure and having prediction results being fetched through a service, with listeners and schedulers handling communication between our production server and the cloud.

- What problems do you run into in your day-to-day work when performing your data analysis? Is there a standard way of solving each of them, or do you have a workaround?
  lack of a unified and user-friendly framework for data analysis
  - *Why is this a problem? we use numerous different tools and libraries, which makes it hard to share the results of an analysis with another colleague*
  - *How do you currently solve the problem? dockers*
  - *How would you ideally like to solve the problem? have a unified framework that would cover all our needs and facilitate easy collaboration*
- Is any of the tools, mentioned in the table above, a must (one that no alternative execution on other tools/platforms would be allowed) for the case studies that you describe?
  SQL, due to its ease of use compared to alternatives and its widespread adoption by the rest of the team.
- Are you able to program/set up a new/custom data processing workflow?
  Yes, but we haven't yet implemented it. In specific, we will be switching over to stream processing soon, breaking down the business logic into independent consumers.
- On what kind of infrastructure do you usually run the analysis (e.g. personal laptop, high spec workstation, server, cluster, HPC etc):
  personal laptop, occasionally Spark clusters for big-data projects such as density maps.

3. Expected benefits from using INFORE

- A. Please mention more data sources that would you like to use and why. If you use a GIS system (e.g., QGIS, ArcGIS) or an application that projects data on the map (e.g., marinetraffic.com), please mention the layers that you use and/or data that you typically import. : *Ideally we would need access to commercial maritime data such as agreed Laycans, freight rates, bunker prices, cargo lists and type of charter contracts (bareboat, voyage etc.).*
  Regarding usage of GIS systems, we currently utilize QGIS mainly for visualization of datasets (vessel positional data, geometry tables/layers reflecting maritime assets such as ports and berths, layers containing linestrings that represent commercial voyage routes). We also use PostGIS extension of PostgreSQL server for efficient querying and analysis of geospatial data.
- B. Which new information would you like to extract from these (old and new) data sources?: *please see my answer in A.*
- C. Are there specific events that you would like to forecast in real-time, which you currently cannot forecast?:

estimation of time of departure, estimation of main engine and/or hull maintenance needs based on real-time operational and hydrographic data, differentiation between various proximity events (eg. bunkering, supplying, etc.) ideally also with drilldown capability (eg. Supplies: lubes, provisions, stores, etc.). The end goal being to create a replenishment heatmap with quantities and fluctuation of prices of goods being bought, that could provide future intelligence into Procurement decisions.

- D. Would you find it an acceptable trade-off to significantly speed up your data analysis tasks, if the provided output was a fairly accurate approximation of the correct result? :
  Yes.
- E1. Rate the following objectives of INFORE, based on how useful they may be at YOUR data analysis (1: Not useful, 2: Little Use, 3: Average Use, 4: Quite useful, 5: Very useful)
  - Ability to design data processing workflows with no code required (3)
  - Ability to change algorithm parameters graphically (2)
  - Ability to receive quick approximate answers instead of 100% accurate, but long running queries (4)
  - Ability to interactively explore the data in order to detect patterns/features of interest (4)
  - Ability to accurately forecast events of interest (5)
  - Ability to automatically optimise your data analysis task over different data processing platforms (HPC, Big Data Platforms, etc). (3)
- E2. Rate the following objectives of INFORE, based on how useful they may be at the data analysis of OTHER data analysts in your organization (1: Not useful, 2: Little Use, 3: Average Use, 4: Quite useful, 5: Very useful)
  - Ability to design data processing workflows with no code required (3)
  - Ability to change algorithm parameters graphically (2)
  - Ability to receive quick approximate answers instead of 100% accurate, but long running queries (3)
  - Ability to interactively explore the data in order to detect patterns/features of interest (4)
  - Ability to accurately forecast (defined or currently unknown) events of interest (5)
  - Ability to automatically optimise your data analysis task over different data (3) processing platforms (HPC, Big Data Platforms, etc). (3)
- F. What are your expectations regarding the system usability?
  flexible scaling according to the website's incoming requests, fast prototyping, support for open-source custom solutions
- G. What is the expected added value from INFORE for you and your corporation?
  ability to fuse data from different sources and provide a unified framework for data extraction & manipulation.

Appendix B.6 User 6 - Answers to Questionnaire

1. User background information

- Professional position [and years of experience]: *Data Scientist: 3 years*
- Domain of expertise: *Machine Learning and Optimisation*
- What are your main job tasks? *Develop machine learning products*

2. Existing workflow

- Please describe the different kinds of data sources that you use in your day-to-day tasks and the tools that you use to process them

Kind of data sources	Format	Volume (approx.)	Purpose (task involved)	Tools used to process the data*	Automatic/Manual/Semi-automatic	Historical/ real-time	Update frequency
AIS data	Structured data	Approx. 1TB	Generate density maps, Create routes	SQL server, PostGIS, Python, Spark	Semi-automatic	Historical	Daily
Vessel statics	Structured Data		Vessel static correction	SQL server, Python,	Semi-automatic	Historical	Daily
Kafka streams	Semi-structured		Generation of vessel metrics	Python	Automatic	Real-time	daily

*If custom programs are used to process the data, please mention the programming language.

- Which is the aim of the analysis you perform (what kind of insights do you try to find)?
  - *Generate intelligence from our data to create products*
- What data processing challenges do you experience in your day-to-day tasks (e.g., fusion of heterogeneous sources, performance, analytics)?
  - *Performance related issues*
- Provide examples of use/ case studies
  - *Data is extracted from the SQL database and is then uploaded to a server dedicated to the Data team. The dbscan algorithm is run on python 3 and its output is then entered in Postgis where some cleanup procedures have been written to filter and finalise the results.*
- What problems do you run into in your day-to-day work when performing your data analysis? Is there a standard way of solving each of them, or do you have a workaround?
  - Why is this a problem?
  - How do you currently solve the problem?
  - How would you ideally like to solve the problem?

The volume of AIS data places a big overhead of our computing infrastructure and this is usually dealt with by using down sampled data which create information loss. In an ideal world we would like to work with all the data without having the need to downsample

- Is any of the tools, mentioned in the table above, a must (one that no alternative execution on other tools/platforms would be allowed) for the case studies that you describe?
  SQL Server is, at the moment, irreplaceable as it is the infrastructure where all of our data is stored by default
- Are you able to program/set up a new/custom data processing workflow? *Yes*
- How long does it take to program/set up a new data processing workflow?
  Depends on requirements
- Are you capable of optimising your data analysis operations? *Yes*
- On what kind of infrastructure do you usually run the analysis (e.g. personal laptop, high spec workstation, server, cluster, HPC etc)
  Personal computer and a server with more RAM than my PC

3. Expected benefits from using INFORE

- A. Please mention more data sources that would you like to use and why.
  Vessel information collected from sensors, market data like freight rates and cargo carried
- B. Which new information would you like to extract from these (old and new) data sources?
  Fuel efficiency models, route optimisation, supporting vessel scheduling
- C. Are there specific events that you would like to forecast in real-time, which you currently cannot forecast?
  Estimated time of arrival/departure
- D. Would you find it an acceptable trade-off to significantly speed up your data analysis tasks, if the provided output was a fairly accurate approximation of the correct result?
  Yes but it depends on the reduction in accuracy
- E1. Rate the following objectives of INFORE, based on how useful they may be at YOUR data analysis (1: Not useful, 2: Little Use, 3: Average Use, 4: Quite useful, 5: Very useful)
  - Ability to design data processing workflows with no code required 2
  - Ability to change algorithm parameters graphically 1
  - Ability to receive quick approximate answers instead of 100% accurate, but long running queries 4
  - Ability to interactively explore the data in order to detect patterns/features of interest 5
  - Ability to accurately forecast events of interest 5
  - Ability to automatically optimise your data analysis task over different data processing platforms (HPC, Big Data Platforms, etc). 5
- E2. Rate the following objectives of INFORE, based on how useful they may be at the data analysis of OTHER data analysts in your organization (1: Not useful, 2: Little Use, 3: Average Use, 4: Quite useful, 5: Very useful)
  - Ability to design data processing workflows with no code required 3
  - Ability to change algorithm parameters graphically 4
  - Ability to receive quick approximate answers instead of 100% accurate, but long running queries 4
  - Ability to interactively explore the data in order to detect patterns/features of interest
  - Ability to accurately forecast (defined or currently unknown) events of interest 4
  - Ability to automatically optimise your data analysis task over different data processing platforms (HPC, Big Data Platforms, etc). 5
- F. What are your expectations regarding the system usability?
  - *Increased computational speed, some form of UI/UX*
- G. What is the expected added value from INFORE for you and your corporation?
  - *Faster development of products, ability to analyse higher data volume, easier horizontal scalability in terms of computing power*

Appendix B.7 User 7 - Answers to Questionnaire

1. User background information

- Professional position [and years of experience]: *Researcher*
- Domain of expertise: *Maritime domain awareness*
- What are your main job tasks?: *Anomaly detection*

2. Existing workflow

- Please describe the different kinds of data sources that you use in your day-to-day tasks and the tools that you use to process them

Kind of data sources	Format	Volume (approx.)	Purpose (task involved)	Tools used to process the data*	Automatic/Manual/Semi-automatic	Historical/ real-time	Update frequency
AIS stream	Kafka	25 GB (or 520,000,000 vessel positions) per day	Anomaly detection	Akka (https://akka.io) + Scala	Automatic	Real-time	Sub-second
Patterns of Life	Database Table	5 GB (10,000,000 polygons)	Anomaly detection	Akka (https://akka.io) + Scala	Automatic	Historical	Annually

*If custom programs are used to process the data, please mention the programming language.

- Which is the aim of the analysis you perform (what kind of insights do you try to find)?
  Early detection and forecasting of anomalous behavior of vessels.
- What data processing challenges do you experience in your day-to-day tasks (e.g., fusion of heterogeneous sources, performance, analytics)?
  The data processing challenges are:
  - *Complicated processing flow of data*
  - *Real-time requirements (not just streaming)*
  - *Continuously querying/updating in-memory data structures/indices*
- Provide examples of use/ case studies
  Examples of anomalies in vessel behavior are:
  - *Detecting anomalous events*
    - *Proximity between two vessels, outside port*
    - *AIS transmitter switch off (waiting for data timed out whilst within coverage area)*
    - *Route deviation*
    - *Etc.*
  - *Forecasting events, e.g. on collision course*

- What problems do you run into in your day-to-day work when performing your data analysis? Is there a standard way of solving each of them, or do you have a workaround?
    - Why is this a problem?
    - How do you currently solve the problem?
    - How would you ideally like to solve the problem?

  Some vessels do not comply with AIS, i.e. they do not transmit or transmit inaccurate data. There are some processes in place to reduce noise in the available data, but currently there is no way to find data for ships not transmitting (dark targets).
- Is any of the tools, mentioned in the table above, a must (one that no alternative execution on other tools/platforms would be allowed) for the case studies that you describe?

➔ *Akka has a very simple, intuitive, and generic data abstraction model that can scale transparently. This abstraction is the "actor" which is an object that has a state, reacts to messages of different type, and can communicate with any other actor. So, it allows us to create complex data flows and structures very fast and easily.*
There is a plethora of other platforms and frameworks for big data processing (both streaming and batch). Each of them provides a predefined API on top of which common data processing techniques can be implemented. The drawback of these frameworks is that when a custom processing flow (one that requires querying in-memory data structures for example) is necessary the predefined APIs pose limitations. Such processes could for example be: (a) execution of an action when no data is received from a vessel for a certain length of time (timeout) (b) finding the vessels that are close to a vessel every time a new position is received for it.

- Are you able to program/set up a new/custom data processing workflow?
  Yes, the abstractions of the framework allow for highly custom data processing workflows.
- How long does it take to program/set up a new data processing workflow?
  It depends on the complexity and the requirements of the data processing flow.
- Are you capable of optimising your data analysis operations?
  Yes.
- On what kind of infrastructure do you usually run the analysis (e.g. personal laptop, high spec workstation, server, cluster, HPC etc)
  Rack server.

3. Expected benefits from using INFORE

- A. Please mention more data sources that you would like to use and why.
  To gather data for vessels that do not transmit (dark targets) other sources (besides AIS) are needed, e.g. satellite, radar etc.
- B. Which new information would you like to extract from these (old and new) data sources?
  Positional data.
- C. Are there specific events that you would like to forecast in real-time, which you currently cannot forecast?
    - *Spoofing, i.e. transmission of inaccurate/false data (integrate with other sources such as satellite data)*
    - *Smuggling (e.g., real-time proximity estimation and ship-to-ship transfers)*
    - *Illegal fishing (e.g., vessel entering exclusive economic zone of another country and following some patterns that suggest that it could be a possible dark target)*
    - *Piracy*
    - *Pollution (for example, vessel in emission control zone or protected sites such as NATURA areas)*

  D. Would you find it an acceptable trade-off to significantly speed up your data analysis tasks, if the provided output was a fairly accurate approximation of the correct result?

Certainly.

- E1. Rate the following objectives of INFORE, based on how useful they may be at YOUR data analysis (1: Not useful, 2: Little Use, 3: Average Use, 4: Quite useful, 5: Very useful)
  - Ability to design data processing workflows with no code required 1
  - Ability to change algorithm parameters graphically 1
  - Ability to receive quick approximate answers instead of 100% accurate, but long running queries 4
  - Ability to interactively explore the data in order to detect patterns/features of interest 4
  - Ability to accurately forecast events of interest 5
  - Ability to automatically optimise your data analysis task over different data processing platforms (HPC, Big Data Platforms, etc). 4
- E2. Rate the following objectives of INFORE, based on how useful they may be at the data analysis of OTHER data analysts in your organization (1: Not useful, 2: Little Use, 3: Average Use, 4: Quite useful, 5: Very useful)
  - Ability to design data processing workflows with no code required 2
  - Ability to change algorithm parameters graphically 1
  - Ability to receive quick approximate answers instead of 100% accurate, but long running queries 4
  - Ability to interactively explore the data in order to detect patterns/features of interest 5
  - Ability to accurately forecast (defined or currently unknown) events of interest 4
  - Ability to automatically optimise your data analysis task over different data processing platforms (HPC, Big Data Platforms, etc). 4
- F. What are your expectations regarding the system usability?
  The system should be stable and efficient enough to handle high volumes of incoming data whilst performing complex operations needed in order to detect and forecast anomalous events.
- G. What is the expected added value from INFORE for you and your corporation?
  Fusion of data from surveillance sources (radar, satellite etc.).