Grant Agreement No: 687591

30/06/2016

Big Data Analytics for Time Critical Mobility Forecasting

datAcron

Maritime Data Preparation and Curation (interim)

Deliverable Form	
Project Reference No.	H2020-ICT-2015 687591
Deliverable No.	5.2.1
Relevant Work Package:	WP 5
Nature:	R
Dissemination Level:	Public
Document version:	1.2
Due Date:	30/06/2016
Date of latest revision:	30/06/2016
Completion Date:	30/06/2016
Lead partner:	NARI
Authors:	Cyril Ray, Elena Camossi, Anne-Laure Jousselme, Melita Hadzagic, Christophe Claramunt, Ernie Batty
Reviewers:	
Document description:	This deliverable (interim report) provides a description of available maritime data for the project.
Document location:	Documents/datAcron/WP5/Deliverables/Final

HISTORY OF CHANGES

Version	Date	Changes	Author	Remarks
1.0	18/05/2016		IMISG - CMRE - NARI	Initial version
1.1	17/06/2016		IMISG - CMRE - NARI	Draft version released for comments
1.2	30/06/2016		IMISG - CMRE - NARI	Final version

EXECUTIVE SUMMARY

The second task (WP5.2) of the maritime use case work package aims to deliver a definition of the datasets to be used in the research and for the evaluation and validation purposes. Specifically, it describes multiple, heterogeneous datasets that may be used for maritime scenarios and use case validation.

This datAcron deliverable entitled *Maritime Data Preparation and Curation* is an interim report for the task WP5.2 dedicated to the identification of available sources. The objective of this report is to list and describe available data sources relevant to the maritime scenario and project objectives, as well as to provide the initial details about their nature and accessibility. It briefly introduces how they support the different scenarios, while the choice of their effective use is left to the datAcron partners.

After an introduction, the report summarises the maritime use case as described in Deliverable D5.1. Section 3 briefly describes data challenges regarding velocity, volume, veracity and variety aspects. Section 4 is dedicated to the description of identified datasets for the maritime use case. This section first describes AIS data sources, with a specific focus on data provided by the datAcron partners. In a second part, contextual data sources are described. Three main contextual data types have been identified in support to the maritime use case: data supporting navigation, data defining official registers and data about weather and ocean conditions. The section 4 then specifies the relationship between data and scenarios defined in Deliverable D5.1. Finally the definition of synthetic and pseudo-synthetic datasets in introduced.

TABLE OF CONTENTS

Contents

1 INTRODUCTION **1**

2 MARITIME USE CASE **2**
2.1 Use Case Design . . . 2
2.2 Monitoring Fishing Activities . . . 3

3 DATA CHALLENGES **6**
3.1 Variety . . . 6
3.2 Veracity . . . 7
3.3 Volume and Velocity . . . 7

4 DATA IN SUPPORT OF THE MARITIME USE CASE **8**
4.1 Automatic Identification System (AIS) datasets . . . 9
4.1.1 Spatial coverage . . . 10
4.1.2 IMISG AIS dataset . . . 10
4.1.3 NARI AIS dataset . . . 16
4.2 Additional Sources of Geo-Spatial Information . . . 17
4.3 Contextual data sources . . . 17
4.3.1 Ports and Navigation rules, Vessel information and Fishing facts . . . 19
4.3.2 Weather and ocean data . . . 23
4.4 Relation Between Data and Scenarios . . . 25
4.5 Synthetic and pseudo-synthetic datasets . . . 25

LIST OF FIGURES

1 Conceptual diagram of the datAcron logical flow for the Maritime Use Case. . . . 4
2 Monthly counts of unique vessels from AIS messages in the CMRE historical database [1]. . . . 7
3 datAcron retained spatial area . . . 10
4 Fishing vessels in Brest Bay . . . 11
5 Geographical coverage of the maritime use case datasets (IMISG) . . . 12
6 IMISG architecture . . . 12
7 Single Line Messages . . . 14
8 Multiple Line Messages . . . 15
9 Coverage of NARI dataset . . . 16
10 Dynamic datasets . . . 18
11 Contextual datasets: Ports and Navigation Rules . . . 20
12 Contextual datasets: Vessel registers and lists . . . 22
13 Contextual datasets: wheather and ocean datasets . . . 26
14 Illicit Activities . . . 27
15 Secured Fishing . . . 28
16 Sustainable Development . . . 29
17 Sketch of datAcron evaluation . . . 30
18 Example of pseudo-synthetic anomalies . . . 30

LIST OF TABLES

1 Scenarios' objectives, user's role and actions, MSI 4

1 INTRODUCTION

Crucial to Maritime Situation Awareness (MSA) is the compilation of the Recognized Maritime Picture from multiple and possibly heterogeneous sources and its continuous monitoring and assessment against contextual information (e.g. maritime routes or loitering areas inferred from the analysis of historical data). This requires not only detecting, tracking and classifying vessels but also detecting, classifying and predicting their behaviour.

Sensor networks mixing cooperative self-identification systems (e.g., Automatic Identification System - AIS) and non-cooperative systems (e.g., coastal radars or satellite imagery) provide the necessary complementarity and redundancy of information to help overcome signals deception (e.g., GPS manipulation and spoofing are frequent for AIS [5]) in order to increase the clarity and accuracy of the maritime picture. In many cases, intelligence information can also be helpful in refining and guiding the search in the huge amount of data to be processed, filtered and analysed, as well as representing the contextual information for decision support systems in MSA applications [2].

Facing the huge *volume* of *various* information with high *velocity* which often lacks *veracity*, a system to automatically process both historical and timely information would greatly support the operator in monitoring and analyss tasks. From these real-world practical issues stem the research aims of the *Big Data Analytics for Time Critical Mobility Forecasting* project datAcron [1]. The goal of datAcron is to develop novel methods for real-time detection and prediction of trajectories and important events related to moving entities, together with advanced visual analytics methods, over multiple heterogeneous, voluminous, fluctuating, and noisy data streams from the moving entities, correlating them with archived data expressing, among others, entities' characteristics, geographical information, mobility patterns, regulations and intentional data (*e.g.* planned routes), in a timely manner. The ultimate goal is to increase the safety, efficiency and economy of operations concerning moving entities in the maritime domain, while the general goal of the project underlines five main research objectives to be addressed within the development of highly scalable methods for advancing:

Obj.1 Spatio-temporal data integration and management solutions;

Obj.2 Real-time detection and forecasting accuracy of moving entities' trajectories;

Obj.3 Real-time recognition and prediction of important events concerning these entities;

Obj.4 General visual analytics infrastructure supporting all steps of the analysis through appropriate interactive visualisations;

Obj.5 Producing streaming data synopses at a high-rate of compression.

datAcron addresses two critical domains: maritime and aerial traffic, which will guide the research and development and will drive the assessment of the datAcron approach.

This document is Deliverable 5.2 of the Work Package 5 (WP5) of the datAcron project. It concerns the maritime use case and describes maritime data sources in support to the datAcron research objectives.

[1] datAcron project website: http://www.datacron-project.eu/.

2 MARITIME USE CASE

This section summarizes the maritime use case of datAcron, which describes possible operational uses of datAcron for *Fishing Activity Monitoring* while focusing on relevant practical challenges, such as the data characteristics, and operational questions. It emphasises a human-centric automatic processing of data, stressing the role of the user (or decision maker) in his/her interaction with the system. Details are given in deliverable D5.1.

2.1 Use Case Design

The methodology used to develop the use case described herein relies on previous experiences, where use cases were designed to support collaborative research on context-based reasoning in high-level information fusion. It and adopts the definition of the use case given by McBreen et al. in [3], where a *use case* describes the interaction of a user with a system to be designed, to achieve a specific goal or accomplish a specific task. The system requirements can then be derived so as to enable the user to achieve his/her objectives in different scenarios. The *scenarios* illustrate different usages of the system, and eventually define success (if the goal is achieved) or failure (if the goal is not achieved).

The resulting use case provides a tool to address different aspects of a large research problem, describing users' needs, operational problems and underlying challenges. Illustrating research findings on a common use case, sharing the same datasets, and utilising outputs from other teams are all benefits of having an integrated picture of the general research problem.

As such, the datAcron maritime use case has to satisfy the following requirements:

Req.1 Address challenging problems deemed of interest for the maritime operational community in general;

Req.2 Be aligned with the European Union maritime policies and needs in particular;

Req.3 Be aligned with datAcron research objectives and expected outcomes such that the use case challenges the datAcron's technical solutions to be developed, while accommodating the research interests of the different partners;

Req.4 Describe the problem in a simple way as a kind of "skeleton", flexible enough to allow further evolution and developments as possibly requested by partners' interests;

Req.5 Provide the necessary information to understand the user's goal, from which the corresponding sub-goals, associated levels of granularity required, the information needs and the desired output quality can be deduced;

Req.6 Act as an "integrator" for the different aspects to be pursued so that teams can illustrate their findings within a common story;

Req.7 Provide a background and support for close interactions between the different work packages and teams involved with the team in charge of the maritime use case;

Req.8 Rely on the available datasets (unclassified, shareable) among the teams and others of interest in the research community (e.g., AIS data, radar datasets, databases of past events, intelligence reports, etc).

The design and the development of decision support systems for improving MSA need to take into account Big Data challenges, that is, they need to be able to process in real-time *voluminous* and high *velocity* information of different nature (numerical, natural language statements, objective or subjective assessments, ...), originating from a *variety* of sources (sensors and humans - hard and soft), which often lacks *veracity* (data are either uncertain, or imprecise, vague, ambiguous, incomplete, conflicting, incorrect).

The datAcron Maritime Use Case comprises multiple scenarios that describe how actors in the use case perform a set of operations in order to achieve a specific goal. Scenarios describe the current operations that will serve as a basis for understanding and validating the datAcron technology, while demonstrating how it can be effectively used in the maritime domain.

The collaboration with the operational partners ensures that the use case is operationally relevant. In particular, the use case describes the general context of use of datAcron algorithms. The operational information needs are captured by relevant Maritime Situational Indicators (MSI), which formalise events of interest for the operator and the information required to detect them (cf. Table 1 in the next Section).

Operational performance criteria are also considered to specify user expectations and to drive the assessment of the datAcron prototype, closely tying the experimental plan to the use case development. The use case requirements (Req.1-Req.8) may also be used as qualitative system performance metrics, while at the implementation level, they may act as result of validation measures.

2.2 Monitoring Fishing Activities

The datAcron Maritime Use Case focuses on fishing activity monitoring, which is a complex maritime surveillance mission that encompasses several maritime risks and environmental issues such as environmental destruction and degradation but also maritime accidents, Illegal, Unreported, and Unregulated (IUU) fishing and trafficking problems, which will be addressed in different scenarios.

Ensuring *security* and *control* of fishing activities is one of the most important aspect of the European Union Maritime Security Strategy (EUMSS) - Action Plan[2], which defines several strategic interests for the European Union and the Member States in terms of maritime security. Europe, as the world's biggest market for seafood wants to promote better international governance across the world's seas and oceans to keep them clean, safe and secure. Since fishing is an activity that exploits common natural resources, it needs to be regulated to safeguard fair access, sustainability and profitability for all.

In particular, the IUU fishing is a global threat to the marine environment and honest fishermen alike, whose global cost is estimated in about 10 Billion Euros per year. The European Union, in collaboration with International organisations, is committed to fighting IUU fishing worldwide.

Besides the detection of IUU fishing activities, another core issue of the EUMSS is safety. Fishing, in peace situation, is known as one of most dangerous activity. An issue here is that fishing vessels may repeatedly be switching off their AIS devices while fishing. Therefore, ensuring fishing safety requires processing and predicting fishing trajectories in real-time, detecting fishing events, movement patterns and fishing areas when AIS is off, computing collision prediction information with all surrounding ships (regardless of their type) on a time scale of typically 5 to 15 minutes.

datAcron will support the European Union's fishing control and fishing regulations enforcement strategy, while providing the necessary scientific support for processing, analysis and visualisation of fishing vessels at the European scale, together with the capability of predicting the movement of maritime objects and the identification of patterns of movement and navigational events that shall improve existing solutions to monitor the compliance to the European common fisheries policy.

In order to support datAcron's challenges within the fishing monitoring use case, six scenarios have been considered. All scenarios highlight the need for continuous (real-time) tracking of fishing vessels and surrounding traffic, as well as for the contextually enhanced offline data

[2]EUMSS Action Plan: http://ec.europa.eu/maritimeaffairs/policy/maritime-security/doc/20141216-action-plan_en.pdf, published in December 2014

Table 1: Scenarios' objectives, user's role and actions, MSI

Scenario			Objective	Actions	MSI examples
Secure fishing	SC11	Collision prevention	Protect fishing vessels from collision with large vessels (cargos, tankers, ferries)	Warn fishing vessels at risk, warn vessels heading to fishing areas	Vessel is: in proximity of other vessels; drifting
	SC12	Vessel in distress / MOB (SAR)	Provide early assistance to a vessel in distress	Warn the closest vessels for early assistance, provide precise location of the vessel for the SAR team	Vessel is drifting; AIS emission has interrupted
Sustainable development	SC21	Protection of ecological areas	Protect specific areas from illegal fishing activities	Send control patrol boat to suspicious vessels location	Vessel's course is not compatible with expected destination; AIS emission has interrupted
	SC22	Fishing pressure	Estimate and predict fishing pressure, identify areas at risk	No direct action but influence European regulation policies in the foreseeable future.	Vessel is: engaged in fishing; within a given areas
Maritime security	SC31	Migrants and human trafficking	Detect possible human trafficking involving fishing vessels (or the like)	Communicate to security boarder control authorities, provide possible assistance (see SC12)	Vessel's course is not compatible with expected destination; AIS emission has interrupted
	SC32	Illicit activities	Detect suspicious activities involving fishing vessels	Send control boats for further checking	Vessel is on a maritime route; AIS emission has interrupted

analytics. They have been elaborated in order to stress datAcron's algorithms in terms of *velocity*, *veracity*, *variety* and *volume*. They should provide a complete support for trajectory and event detection, prediction and visualisation. For each scenario, the user information needs are expressed through a corresponding list of MSIs. In Table 1, scenarios are summarised with corresponding objectives, possible actions, and example MSIs.

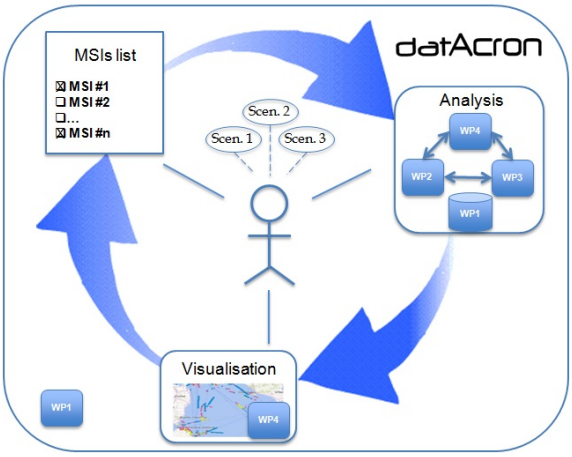


Figure 1: Conceptual diagram of the datAcron logical flow for the Maritime Use Case.

The MSIs defined for fishing monitoring in datAcron formalise the events of interest for the use case, capture the required information while formalising the goals of datAcron algorithms and drive the analysis. A list of the MSIs is detailed in Deliverable D5.1. Specifically, the MSIs have been filtered out to comprise only (1) the MSIs that datAcron can provide, (2) the MSIs that are relevant to the fishing monitoring scenarios.

The conceptual diagram in Figure 1 illustrates the operational flow and the interaction of the datAcron software in the fishing monitoring use case. Depending on the scenario, the user may accomplish different tasks (i.e., monitoring, detecting or preventing the events described by the scenario), and may express his/her information needs through a list of MSIs of interest at a given

time. He/she selects the appropriate algorithms and parametrises them accordingly to run the analysis. He/she is able to observe results of the selected algorithms using the visualisation tools and additional visual analytics, which furhter allow the refinement of the analysis by varying the parameters of the MSIs (e.g. change the areas of interest, speed thresholds).

3 DATA CHALLENGES

The maritime use case will be supported by the appropriate datasets that will be described in the following. An initial list of datasets was provided in the deliverable D5.1 and is updated in this document. We will ensure that the data are usable and aligned in time and space. Additional datasets may be identified in the future upon specific identified needs. The performance criteria specify what the user expects They also help to closely tie the experimental plan (in development) to the use case, resulting in an integrated product.

Reaching appropriate Maritime Situation Awareness (MSA) for the decision maker requires processing in real-time a high *volume* of information of different nature (numerical, natural language statements, objective or subjective assessments, ...), originating from a *variety* of sources (sensors and humans - hard and soft), with a lack of *veracity* (uncertain, imprecise, vague, ambiguous, incomplete, conflicting, incorrect, etc), and coming with high *velocity*. These challenges underlying the datAcron objectives are reflected in the description of the datasets to be used by datAcron (see table of Figure 15, Figure 16 and Figure 14 in Section 4.4) and reminded here.

3.1 Variety

Different types of data are available, which only if properly combined and integrated these data can provide useful knowledge. However, since they are obtained by various types of sensor technologies, they also need to be need to be cleaned up from inconsistencies, standardised in format and summarised.

Maritime surveillance can rely of a variety of heterogeneous sources such as:

1. *Physical sensors* such as Automatic Identification Systems, coastal or on-board radars as traditionally used for tracking objects, Synthetic Aperture Imagery, cameras, ESM on board military ships;

2. *Automated processors* such as trackers, Automatic Target Recognition algorithms or classifiers in general;

3. *Human sources* including operators or analysts themselves possibly manipulating lower level data (*e.g.* videos, radar images) to reflect the chain of information processing, from automation to possible subjective assessments, intelligence reports;

4. *Databases* as records of past events (*e.g.* piracy, accidents, illegal fishing activities), records of vessels such as the Lloyds database.

The complementarity of the information provided by these sources is certainly an asset to an improved MSA, while it nevertheless raise several challenges of:

- managing inconsistencies or conflicting information,
- aligning the data formats or models,
- aligning the semantics,
- aligning the mathematical representation.

The suggested datasets to be used by datAcron project contain the characteristics which will permit to address some of the stated challenges associated with data.

3.2 Veracity

Data measurements have an intrinsic uncertainty, which may be addressed by proper fusion algorithms and clustering in the preparation/preprocessing phase (by assessing the quality of data themselves) and by combining measurements from complementary sources.

The sources themselves lack the quality; they may be unreliable, incompetent, badly intentioned, imprecise, uncertain, etc., the resulting in providing the information that suffers from equivalent drawbacks, and it can thus be conflicting. For instance, AIS data are incomplete, intermittent, with errors, and the signal can be spoofed. The assessment of data quality is very challenging task and an example is provided in Section 4.1.2 for the AIS dataset, where flags have been included. In general, the quality of data is not accessible and can only be estimated based on available statistics (for AIS data for instance), or on some provider of devices.

In Section 4.4, example of lack of veracity of the different datasets are provided, while actually the list cannot be exhaustive and should be updated upon the use of the datasets. A record of quality issues on practical uses of datasets would be interesting.

3.3 Volume and Velocity

The growing number of sensors (in coastal and satellite networks) makes the sea one of the most challenging environments to be effectively monitored; the need for methods for the data processing of vessel motion data at sea, which are scalable in time and space, is highly critical for maritime security and safety.

In particular, the analysis of streaming data from multiple sensors is essential to detect critical events as soon as they occur at sea. This poses the emphasis on incremental clustering that is able to include new data into the data-at-rest already processed as well as on sequential methods able to detect critical events by continuously processing data.

For instance, about 800,000,000 AIS messages are recorded per month by CMRE MSA database, produced by up to 120.000 unique vessels. About 12,000 ships/day are tracked in EU waters and about 100,000,000 AIS positions are recorded per month in EU waters (EMSA, 2012).

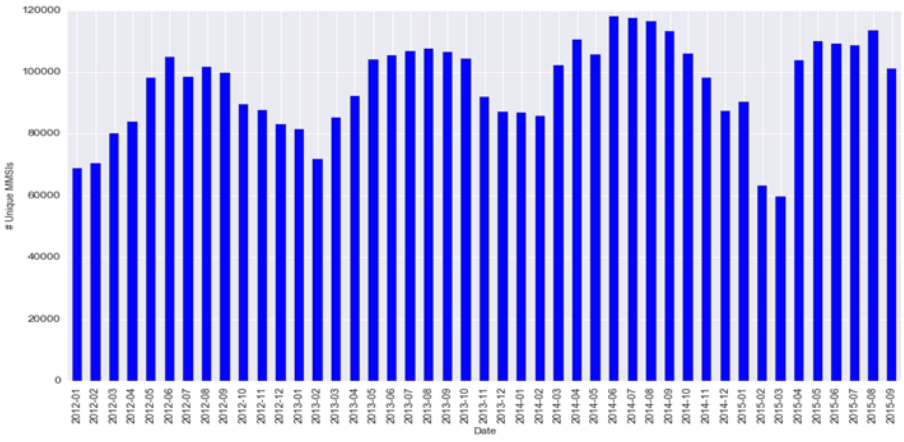


Figure 2: Monthly counts of unique vessels from AIS messages in the CMRE historical database [1].

4 DATA IN SUPPORT OF THE MARITIME USE CASE

Ships' positions obtained from AIS are essentials for datAcron's algorithms, however they might not be always sufficient. Only if properly combined and integrated with other data acquired from other data/information sources (not only AIS), they can provide useful information and knowledge for achieving the maritime situational awareness in support to the datAcron maritime use case and related fishing scenarios.

Therefore, the data to be used in datAcron comprise real and quasi-real data streams as well as archival (or historical) European datasets. These usually need to be cleaned up from inconsistencies, converted into standard formats, harmonized and summarised. The datasets to be used may also be synthetised.

The following list briefly summarizes typical datasets that are relevant to maritime surveillance.

- *Automatic Identification System (AIS)* [3] messages broadcasted by ships for collision avoidance;
- *Long-range identification and tracking (LRIT)* [4] data for global identification and tracking of ships;
- *Vessel Monitoring Systems* (VMS) data used for used in commercial fishing;
- *Marine and coastal Radar* data, used for supporting navigation and collision avoidance;
- *Synthetic aperture radar (SAR) or Inverse SAR (ISAR)* images;
- *Maritime regulations*, that specify the legislation and the rules for navigation and fishing;
- *Marine protected/closed areas*, where fishing and sea traffic may be (temporarily) forbidden;
- *Traffic separation schemes* and *Nautical charts*, useful to define vessel routes;
- *Vessel routes* and *Fishing areas* estimated from historical traffic data;
- *Registry* data on vessels and ports;
- *Records* of past events, such as incidents and illegal activities reports;
- *Meteorological and oceanographic (METOC)* data on atmospheric and sea state conditions and currents.

For the maritime use case, AIS data have been identified as the main data source for generating the maritime moving objects trajectories (synopses) that will be analysed by the algorithms developed by WP2, WP3 and WP4 and visualised by WP4. At this stage, *two internal sources from project partners* have been selected and secured for the project. Their detailed description is provided in 4.1. Additional AIS sources via Internet providers could also be used. Other data can be simulated or derived from the real ones for particular applications, *e.g.*, to include simulated alarms or events of interest for algorithms testing.

To define the context of the use case, information such as marine protected areas and maritime regulations can be used (e.g., to define the area to monitor or the time range of interest in order to prevent potential illegal activities). Historical and METOC data may be useful to correctly set up the analysis parameters and to validate the results (e.g., to filter out the false alarms). Registry data and records of past events may be used to define ground truth information useful to refine the analysis algorithms. A large part of these data is accessible through Internet sources.

[3] AIS: www.navcen.uscg.gov/?pageName=AISmain
[4] LRIT: www.imo.org/en/OurWork/Safety/Navigation/Pages/LRIT.aspx

Other types of information, in particular RADAR, Long-range identification and tracking (LRIT) and Vessel Monitoring Systems (VMS) could be considered in a later stage to enlarge available ground truth data conditioned on their accessibility within the project, however, not all these datasets are necessary to the use case.

In the remainder of the document, potential datasets that can be used in support of the datAcron's Maritime Use Case for testing and validating the datAcron developments of WP1, WP2, WP3 and WP4 are presented, and their availability and how they may support the use case are described.

These datasets will be further consolidated in final task 5.2 report (Deliverable D5.2.2) and Deliverable D5.3, which will discuss the experimental evaluation.

4.1 Automatic Identification System (AIS) datasets

According to the European Commission, several types of ships are obliged to broadcast AIS messages, including: ships of 300 gross tonnage and upwards in international voyages; 500 and upwards for cargoes not in international waters and passenger vessels; and, more recently, smaller fishing vessels. Raw AIS messages comply with ITU-R.M 1371-5 and NMEA 4.0 standards, and are differentiated in 27 type of messages. Two main classes of messages are identified as useful for datAcron:

kinematic messages from which two-dimensional (2D) vessel routes can be derived, and which include information on position (latitude and longitude), speed over ground (SOG), heading, course over ground (COG), Rate of Turn (ROT);

static messages providing ship meta-information such as ship identifiers (MMSI and IMO number), name, type, and dimension of vessel, and the voyage-related information, such as destination (Port of Call), danger, Estimated Time of Arrival (ETA), draught.

AIS messages can be collected by coastal and satellite networks of receivers. Terrestrial AIS (T-AIS) messages from coastal receivers are characterized by high persistence but limited coverage, while Satellite AIS (S-AIS) messages collected by satellite receivers can pick up messages in the open sea, far away from the coastline, and have a larger coverage than T-AIS. AIS messages from different data providers might have complementary spatial and temporal coverage. It is therefore advisable, depending on the area or on the period of interest for the use case, to have a harmonized dataset including AIS messages from different providers and from different types of receivers.

There exist many AIS data providers. Some of them are listed in Figure 10. Among them, open network of AIS receivers freely exchange AIS sentences within communities of volunteers, usually in a merge stream using the NMEA format, such as:

AISHub[5], whose data can be freely distributed.

MarineTraffic [6], sharing real-time AIS data, made available in a *quasi* real-time stream, together with vessels and ports information.

Vesseltracker[7]

The *Maritime Safety and Security Information System (MSSIS)*[8] is an institutional network created by the US government for the sharing of terrestrial and Satellite AIS sentences.

[5] AISHub: aishub.net
[6] MarineTraffic: marinetraffic.com
[7] Vesseltracker: Vesseltracker.com
[8] MSSIS: mssis.volpe.dot.gov

Figure 3: datAcron retained spatial area

Commercial services exist as well, such as: *FleetMoon* [9] or *Vessel Finder* [10]. These provide also free sample of data or limited amount of live AIS streams available for free.

4.1.1 Spatial coverage

A preliminary area of interest for the specification of the maritime use case is specified as depicted in Figure 5. It covers most of the European coasts. This area can be further enlarged, to include events of interest for the validation of datAcron's developments.

Within such a large area, the definition of *test areas* is required. A test area is a region where datAcron can have a more accurate knowledge of ships' movement and ground truth support. This is essential for initial experimentations and validation of algorithms before application at a European scale. For such an area, both terrestrial and satellite AIS data are required. It should contain an active fishing area. Additional contextual information (cartography, regulated areas, known fishing fleet...) have to be available. Knowledge on local fisheries and connections with operational entities (e.g., control center, local committee of fishery, port authorities, navy) are also necessary to establish ground truth.

The western part of France, around Brest city has been identified as a first potential test area that fulfills most of the aforementioned constraints. Moreover, the Brest bay itself has local regulations which enforce fishing vessels to use AIS permanently. Figure 4 illustrates position reports of fishing vessels in Brest Bay during one month.

Other areas can be considered later, according to scenarios and work package objectives.

4.1.2 IMISG AIS dataset

Terrestrial and satellite AIS data provided by IMISG will define the core AIS dataset for datAcron. The maritime.

[9] FleetMoon: `fleetmon.com`
[10] Vessel Finder: `vesselfinder.com`

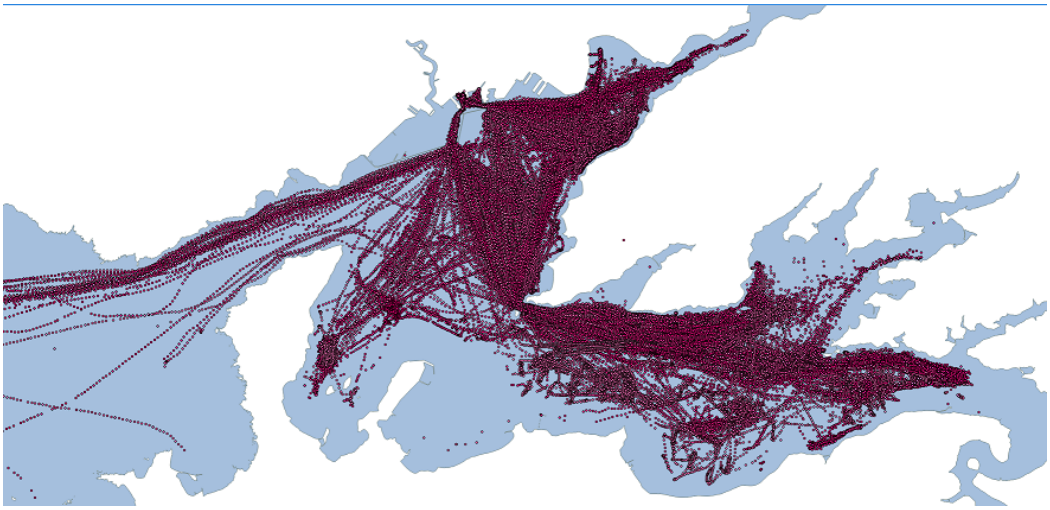


Figure 4: Fishing vessels in Brest Bay

Description of data sources

The AIS data provided by IMISG for the datAcron project has been sourced from a range of terrestrial and satellite AIS sources. The terrestrial sources are collated from various sources in Europe and decimated to limit the amount of data.

The satellite data is obtained from a range of ORBCOMM satellites of various generations and include an AIS receiver on the International Space Station (ISS), a range of older generation satellites and 11 new generation satellites that go to make up 19 sources of satellite data. The AIS equipped Low Earth Orbit (LEO) satellites are in a range of orbits around the earth. Each of the satellites can only download AIS data when there is a ground station within their coverage foot print. When just the European environment is considered, there are three satellite ground stations, Morocco, Italy and Norway, servicing the area of interest for which the datAcron project is consuming AIS data. This overlap of the satellite footprint and the coverage of a ground station primarily affects the delay between when an AIS signal is received on the satellite to when it is available to the datAcron partners via the satellite data collection and processing network.

Figure 5 illustrates terrestrial and satellite AIS data to be obtained and provided for datAcron:

Preparation of data sources

The satellite and terrestrial data are collected by various agencies and tagged with their source (terrestrial or satellite). To the satellite data, additional metadata is attached as indicated in the following data format (cf. paragraph *Data format*).

The terrestrial data arrives in an almost continuous stream. The satellite AIS data arrives in bursts as the satellite downloads the data to a ground station. The size of the burst is dependent primarily on two factors:

1. How much data is in the satellite to be downloaded
2. The type of satellite which determines the type of satellite to ground link

The AIS data contains the following ITU R-M. 1371-5 message types 1, 2, 3, 4, 5, 9, 18, 19, 21, 24 and 27. This allows the data to be used to track all of the following AIS devices:

- Class A AIS (SOLAS and inland waterway)
- Class B Carrier Sense AIS

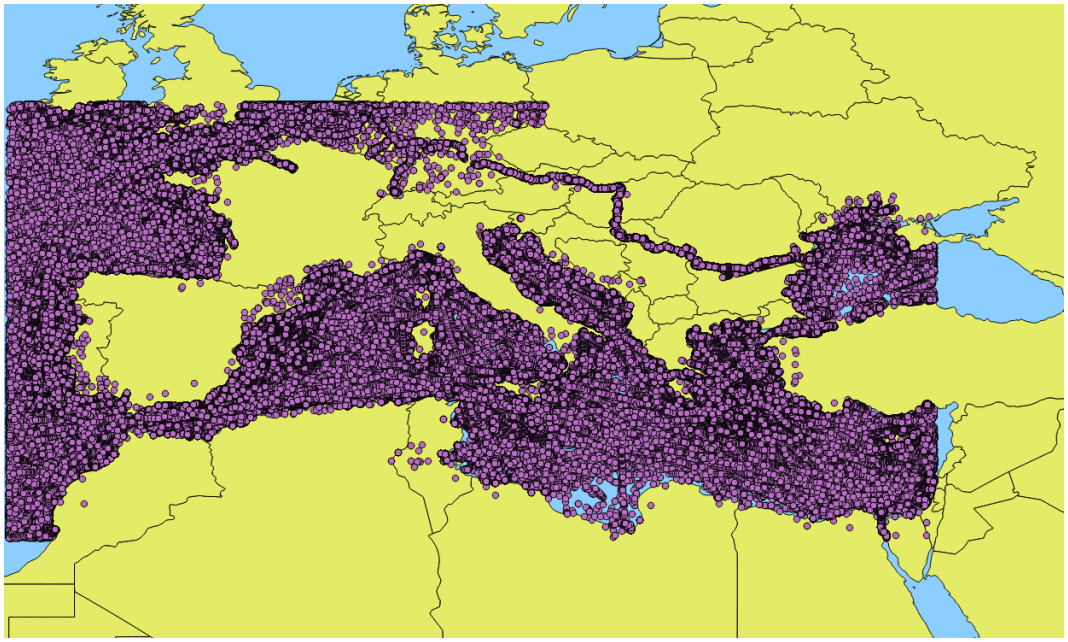


Figure 5: Geographical coverage of the maritime use case datasets (IMISG)

- Class B STODMA AIS
- SAR AIS (helicopters and other aircraft with fitted with AIS)
- AIS AtoN
- AIS-SART

The system that collects and processes the AIS data stream processing runs on a CloudStack environment that runs various services in real time (sub-second) for operational purposes, some services in sub-minutes for operational reporting services and then all received AIS data is stored in a large database (140TB) for business intelligence analytics.

In order to reduce the latency, the AIS data is collected and processed locally by geographically dispersed cloud services.

The front end data collectors have some basic checks on the data and include ensuring that the CRC is correct and for terrestrial AIS receivers, some of the long distance AIS reports are filtered out (the range of the terrestrial AIS receiver is often limited to +/-120NM). The AIS message flow is shown in Figure 6.

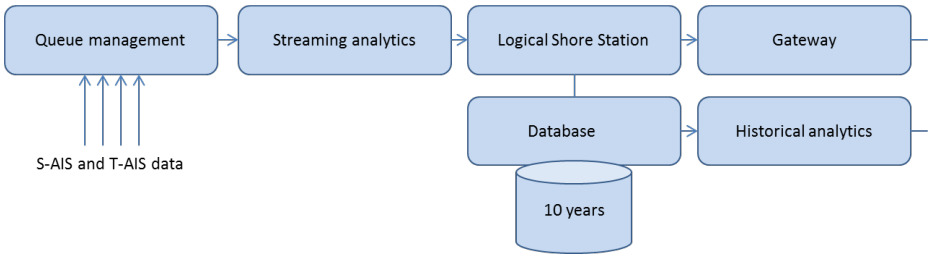


Figure 6: IMISG architecture

Once the AIS data has been collected, it is checked to ensure that the Cyclic Redundancy Check (CRC)[11] is correct and then the data is checked against the IEC 61162-1 specification to

[11]CRC) is an error-detecting code commonly used in digital networks and storage devices to detect accidental changes to raw data

ensure that it is compliant. The IEC 61162-1 allows a range of IEC 61162-1 data formats that includes the ITU R-M. 1371-5 messages that are specific to AIS as well as messages that can be used within an AIS centric Common Shore Side Architecture (CSSA) environment. None of these have been included in the datAcron AIS data stream.

Once the data has passed the initial compliance check, it is then tagged with a range of initial meta data (TAG or Comment block) that includes the port and time it was received and then when the message is stored, the time that it is stored in the primary data base is also made part of the tag.

The AIS data is then processed by a streaming data analytic application that, in real time (less than 500ms latency) validates the data and ensures that the object being tracked is performing within the configured boundaries and this includes the following:

- Message timing validation (terrestrial AIS messages only)
- Slot boundary detection (terrestrial AIS only)
- AIS vessel received station value detection
- Message reporting rate validation (minimum and maximum number of messages per time period)
- Improperly formatted MMSI
- Vessel Static Information validation
- AIS Message channel validation
- NMEA 0183 version 4 and / or IEC 61162-1 message type validation
- Area definitions (inclusive and exclusive areas)
- Duplicate / outlier vessel check

Each AIS message (both terrestrial and satellite) are tagged with a flag field that can be used by the datAcron partners to determine the accuracy of the data being used for further processing.

The AIS data is then stream to the datAcon partners and it is also stored in a large relational database (currently two sharded MySQL databases) to enable the data to be queried for operational, tactical and strategic reporting purposes.

Provision of data

The AIS data, once having been tagged and analysed, is served as a data stream via a data server. The AIS data is stored along with ownership and filter parameters.

To extract historical data, a query can be executed via the Human Machine Interface (HMI) or can be extracted using a REST web service.

Real time data is served within 1,000ms of being available. Connection to the AIS data stream is made by using a Secure proxy and the TCP/IP protocol.

Data format

The AIS data is provided in its raw and unparsed format as received from the AIS data sources, and can contain any of the 27 different message types as described in the ITU-R.M 1371-4 or NMEA 4.0 specification. The data includes a comment or TAG block which provides additional information for the IEC 61162-1 message. Data provided as single or multiple-line messages. Single line messages are formatted as illustrated in Figure 7. Multiple line messages are formatted as illustrated in Figure 8.

```
\s:SIMULATOR,c:1424419673,T:2015-02-20 08.07.53,e:10000010010000000000
i:|X=0|D=1|T=42055.3388029977|P=192.168.2.145:10000|R=IN|*hh\
!AIVDM,1,1,,,18157Rh00=0pPJ1svJ46T5Hf0L08,0*47<CR><LF>
```

The comment or TAG block includes the following parameters:

Identifier	Description
s:	The source of the message.
c:	The unix timestamp of the message when received (seconds since midnight, January 1st, 1970)
T:	The human readable timestamp of the message when received in yyyy-mm-dd hh.nn.ss
e:	The message error flag. Bits in this identifier are set to '1' if any of the 15 errors described in Annex A are true. If no error is present in the data, is this field excluded from the message output
i:	Proprietary data and contains the following fields, separated by a '|' character:

Identifier	Description
X=	Data source RX / TX capability = always set to '0'
D=	Data source 'delayed data flag' = always set to '1'
T=	Proprietary timestamp of the message
P=	The IP address and port where the message was received by the MSA
R=	The direction of the message

The IEC 61162-1 sentence is described below:

Field Name	Example	Description / Comment
	!AIVDM	VDM Message identifier
TotalSentences	1	Total number of sentences.
SentenceNumber	1	Sentence number of this sentence
SeqMsgNum		Sequential message ID. Always empty for a single line message
AISChannel		Channel ID (A, B, C or D or empty if unknown)
EncapsulatedMsg	18157Rh00=0pPJ1 svJ46T5Hf0L08	Bits of the data portion of the AIS message type 1, 2,3, 9,18, 19, 21, 24 or 27 (Each character represent 6 bits – encoding is per NMEA0183). The number of bits (and characters) depends upon the message type.
FillBitsNumber	0	Number of Fill bits appended
Chksum	**47**	Checksum ('12' is the checksum)
	<CR><LF>	Carriage Return and Line Feed

Figure 7: Single Line Messages

Line 1:

```
\g:1-2-1234,s:SIMULATOR,c:1424419673,T:2015-02-20 08.07.53,e:10000010010000000000
i:|X=0|D=1|T=42055.3388029977|P=192.168.2.145:10000|R=IN|*hh\
!AIVDM,2,1,8,,5P000Oh1IT0svTP2r:43grwb05q41P000Oh1IT0svTP2r:43grwb05q41P00,0*15<CR
><LF>
```

Line 2:

```
\g:2-2-1234*2hh\!AIVDM,2,2,8,,0Oh1IT0svT,0*7b<CR><LF>
```

The comment or TAG block includes the following parameters:

Identifier	Description
g:	Identify line 1 out of 2 lines of group 1234 (for example: `g:1-2-1234`)
s:	The source of the message
c:	The unix timestamp of the message when received (seconds since midnight, January 1st, 1970)
T:	The human readable timestamp of the message when received in yyyy-mm-dd hh.nn.ss
e:	The message error flag. Bits in this identifier are set to '1' if any of the 15 errors are true. If no error is present in the data, is this field excluded from the message output.
i:	Proprietary data and contains the following fields, separated by a '|' character:

Identifier	Description
X=	Data source RX / TX capability = always set to '0'
D=	Data source 'delayed data flag' = always set to '1'
T=	Proprietary timestamp of the message
P=	The IP address and port where the message was received by the MSA
R=	The direction of the message

The IEC 61162-1 sentences are described below:

Line 1:

Field Name	Example	Description / Comment
	\!AIVDM	VDM Message Identifier
TotalSentences	2	Number of sentences.
SentenceNumber	1	Sentence number of this sentence
SeqMsgNum	8	Sequential Message ID (0-9)
AISChannel		AIS Channel (A, B, C, D or empty)
EncapsulatedData	5P000Oh1IT0svTP2r:43grwb05q41P000Oh1IT0svTP2r:43grwb05q41P00	First part of the "data" section of the AIS Message
FillBits	0	Number of fill bits
Chksum	15	checksum

Line 2:

Field Name	Example	Description / Comment
	\!AIVDM	VDM Message identifier
TotalSentences	2	Number of sentences
SentenceNumber	2	Sentence number of this sentence
SeqMsgNum	8	Sequential Message ID (0-9)
AISChannel		AIS Channel (A, B, C, D or empty)
EncapsulatedData	0Oh1IT0svT	Last part "data" section of the AIS Message (per ITU M1371-3)
FillBits	0	Number of fill bits
ChkSum	7b	Checksum

Figure 8: Multiple Line Messages

4.1.3 NARI AIS dataset

Terrestrial AIS data will be provided by NARI. This will define a complementary AIS dataset for datAcron, especially for the test area. The maritime AIS data conforms to the NMEA 0183 version 4 specifications and is provided for partners along with timing information.

Description of data sources

The AIS data for the datAcron project is obtained by a single terrestrial receiver (SAAB R4) located in Brest City, France. Figure 9 illustrates its coverage (88% of received messages are located in a range of 50km).

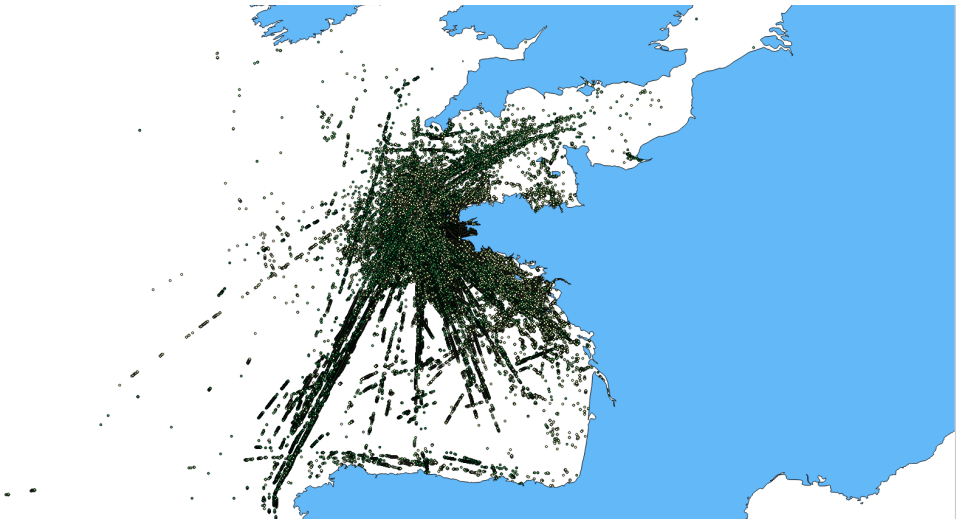


Figure 9: Coverage of NARI dataset

Preparation of data sources

The terrestrial data arrives in an almost continuous stream from one receiver. They are stored on-the-fly in flat files containing raw NMEA messages. The receiver add local time to each frame in UTC.

The AIS data contains all the ITU R-M. 1371-5 message types (1 to 27). This allows the data to be used to track all of the following AIS devices:

- Class A AIS (SOLAS and inland waterway)
- Class B Carrier Sense AIS
- Class B STODMA AIS
- SAR AIS (helicopters and other aircraft with fitted with AIS)
- AIS AtoN
- AIS-SART

The system is continuously collecting data and two years of AIS messages are available (since April 2014).

Provision of data

Data are available as flat files. On demand, a subset of these data can be prepared, parsed and provided as csv files

Data format

A file contains one day of AIS messages and is formatted as NMEA frames :@: Date and time in UTC.

An example of this format is : !AIVDM,1,1,,A,ENjCOGV10V4aRh:2ab@IP000000OmRw8-=mO8P10888fP:00,2*4F :@: Mon, 07 Dec 2015 23:00:01 +0000

Typical format of a csv file will be :

$MMSI_Number; Time; Longitude; Latitude; Heading; Speed; COG; ROT; shipCode$

$227635210; 2009-02-0509:16:25; -4.41620493; 48.28419876; 511.0; 0.00; 290.90; -128.00; 60$

$227635210; 2009-02-0509:16:34; -4.41620159; 48.28419876; 511.0; 0.00; 292.30; -128.00; 60$

$227635210; 2009-02-0509:16:44; -4.41620159; 48.28419495; 511.0; 0.00; 291.80; -128.00; 60$

$227635210; 2009-02-0509:16:54; -4.41620350; 48.28419495; 511.0; 0.00; 294.50; -128.00; 60$

Where parameters are adjusted according to the scenario and algorithms. All fields of all messages can be provided.

4.2 Additional Sources of Geo-Spatial Information

Complementary sources for vessel positioning information can be remote sensing 2D or 3D images for mapping the surfaces of the Earth and the Oceans, such as those provided by the European Space Agency[12]. ESA generates Synthetic Aperture Radar (SAR) data from satellite SENTINEL-1 for ship monitoring and oil pollution monitoring for marine safety. This imagery includes images of global landmasses, coastal zones and shipping routes in European waters in high resolution, coverage of the global oceans at regular interval [13]. It also provides information on Stripmap, Interferometric Wide Swath, Extra Wide Swath, and Wave[14].

Free and open access to Sentinel-1 and Sentinel-2 user products is offered through the Sentinels Scientific Data Hub[15]. Data are available in SAR SENTINEL-SAFE format[16], and can be managed with the NEST and PolSARPro Sentinel Toolboxes for working with SENTINEL-1 data products[17].

The dynamic data sources are summarised in Table 10. The consortium will mainly exploit data provided by consortium partners, i.e. both T-AIS and S-AIS messages, as described in this Section.

4.3 Contextual data sources

To define the context of the scenarios defined in D5.1, different information is necessary, including contextual information in support and to regulate navigation (e.g., navigation rules, nautical charts, regulated areas, facility areas), which depends on the activity that is ongoing and on

[12] (https://earth.esa.int)
[13] https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar
[14] https://scihub.copernicus.eu/userguide/WebHome#Sentinel_1_Data_Offer
[15] https://scihub.copernicus.eu/
[16] https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar/data-formats/sar-formats
[17] http://step.esa.int/main/download/

Information	Data Type	Data Type Description	ID		Data provider	Description and data provenance	Spatial coverage	Temporal coverage	Licence and Conditions of use	Costs	Data access and Format	Format description	Quality
Dynamic data and static information on vessels	Automatic Identification System messages	Terrestrial (T-AIS) and Satellite (S-AIS) information on 2D vessel voyages, including ship position (kinematic messages: position, speed, heading, COG, ROT), ship meta-information (static messages: ship ids, name, type, dimension), route-based information (destination, danger, ETA, draught)	P1	main sources	IMIS Global	Merged stream of AIS sentences from different providers: MSSIS, ExactEarth, AISHub	Europe	Hystorical data	Restricted used within datAcron consortium	NA	TCP/UDP streams raw and unparsed format as received from the AIS data sources, and can contain any of the 27 different message types as described in the ITU-R.M 1371-4 or NMEA 4.0 specification. Data includes a comment or TAG block which provides additional information for the IEC 61162-1 message ()	NMEA format: http://catb.org/gpsd/AIVDM.html See also full description of MARI Web AIS format "AIS Data sources, formats and processing for the datAcron project"	Each AIS message is tagged with a flag field that can be used by the datAcron partners to determine the accuracy of the data being used for further processing 1. Message timing validation (terrestrial AIS messages only) 2. Slot boundary detection (terrestrial AIS only) 3. AIS vessel received station value detection 4. Message reporting rate validation (minimum and maximum number of messages per time period) 5. Improperly formatted MMSI 6. Vessel Static Information validation 7. AIS Message channel validation 8. NMEA 0183 version 4 and / or IEC 61162 1 message type validation 9. Area definitions (inclusive and exclusive areas) 10. Duplicate / outlier vessel check
					NARI	AIS data from NARI AIS receiver	Brest area	Hystorical data (2 years 2014-2016)	Restricted used within datAcron consortium	NA	Flat files of AIS messages conformal to NMEAdata format. Parsed messages with a selection of AIS parameter can be provided on demand as csv files	NMEA format: http://catb.org/gpsd/AIVDM.html	Raw stream. On-demand pre processing
					DCNS Research					NA	NA		
					AIS HUB (www.aishub.net)	Merged stream of openly shared AIS sentences from volunteer providers	global	Live feed	All data received from AISHUB can be redistributed and shared	Free. Need to be part of the AIS Hub Network	TCP/UDP feed of raw NMEA data sentences. Data available also in XML / JSON format	NMEA format: http://catb.org/gpsd/AIVDM.html	raw stream. Requires pre-processing and filtering to remove invalid messages, and errors in data
				additional sources	Vessel Finder (www.vesselfinder.com)	Fused AIS feed of global shipping	global	Live feed	NA	Commercial services. Data sample available	NMEA sentences provided via TCP/IP or UDP connection. Real-Time API access to decoded position data in XML / JSON format	NMEA format: http://catb.org/gpsd/AIVDM.html JSON/XML format: https://www.vesselfinder.com/realtime-ais-data	NA
					Vesseltracker.com GmbH (www.vesseltracker.com)	Fused AIS feed from privately owned T-AIS receiver network and S-AIS constellation	global	Live feed. Historical data available (including reports and analytics)	NA	T-AIS of passenger vessels available for free (registration required). Costal AIS available to vesselTracker network partners. Free trials for commercial services	NMEA sentences, via API (SOAP and REST) and FTP	NMEA format: http://catb.org/gpsd/AIVDM.html	Auto cleanse and auto correct are applied
					FleetMoon (www.fleetmon.com)	Live AIS traffic (T-AIS and S-AIS), Port and Vessel database. Hystorical terrestrial AIS data. Historical data can include past vessel positions or past port call	global	T-AIS from 2007, S-AIS from 2013 on	NA	Commercial services. Limited Free plan available for T-AIS and vessels database	API, XML & Raw NMEA AIS Streams	NMEA format: http://catb.org/gpsd/AIVDM.html	
					Marine Traffic (www.marinetraffic.com)	Fused AIS feed from T-AIS and S-AIS networks	global	Live feed. Historical historical AIS positions and port calls	NA	T-AIS available for free (registration required). Free trials for commercial services	NMEA sentences provided via TCP/IP or UDP connection. CSV for hystorical data. Shared as Open Sea Map layer (http://openseamap.org/index.php?id=schiffstracking&L=1)	NMEA format: http://catb.org/gpsd/AIVDM.html CSV format: http://www.marinetraffic.com/en/p/ais-historical-data	NA
					Maritime Safety & Security Information System MSSIS (https://mssis.volpe.dot.gov)	Fused AIS feed from T-AIS and S-AIS networks	global	Live feed.	NA	AIS available to MSSIS network partners	NMEA sentences shared through Transview TV32 software.	NMEA format: http://catb.org/gpsd/AIVDM.html	NA
	Synthetic aperture radar (SAR) data	Remote sensing 2D or 3D images for mapping the surfaces of the Earth and the Oceans	P2	main sources	ESA (https://earth.esa.int)	SENTINEL-1 Synthetic Aperture Radar (SAR) data for ship monitoring and oil pollution monitoring for marine safety. Images of global landmasses, coastal zones and shipping routes in European waters in high resolution, coverage of the global oceans at regular interval(https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar). Includes information on Stripmap, Interferometric Wide Swath, Extra Wide Swath, and Wave. (https://scihub.copernicus.eu/userguide/WebHome#Sentinel_1_Data_Offer)	global, European		online resources are public releasable (registration required) Sentinel Toolboxes and full source code is distributed freely under the GNU GPL license	Free, full and open access to Copernicus Sentinel Data.	The Sentinels Scientific Data Hub provides complete, free and open access to Sentinel-1 and Sentinel-2 user products, starting from the In-Orbit Commissioning. Review (IOCR).https://scihub.copernicus.eu/ Sample data https://earth.esa.int/web/guest/data-access/sample-data#_56_INSTANCE_z3vD_matmp	SAR SENTINEL-SAFE format: https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar/data-formats/sar-formats NEST and PolSARPro software toolboxes for working with SENTINEL-1 data products Sentinel Toolboxes http://step.esa.int/main/download/	Calibration and validation described here https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar/cal-val

Figure 10: Dynamic datasets

the scenario of interest for the operator. For the detection and the forecast of illicit activities, events, incidents, as well as lists and registers created by official organizations and institutions may help defining patterns of list of vessel of interest for the specific activity. In particular, numerous data sources exist at European level to contextualise the sustainable development scenario, created as European contribution to global environmental programmes such as the Global Earth Observation System of System (GEOSS). This is the case, for instance, of the Copernicus Environmental monitoring programme.

In the rest of the section contextual sources and list potential datasets are described. We first introduce sources in support of navigation, together with information on fishing facts that can be used to develop ground truth in the experimental evaluation phase. Then we describe weather and ocean sources, that contextualize the analysis in support of the fishing monitoring scenarios.

4.3.1 Ports and Navigation rules, Vessel information and Fishing facts

Information and datasets in support of navigation, including navigation rules, navigation facilities, are described in Table 11. Vessel registers, including vessel databases, lists of blacklisted vessel, together and news sources for fishing incidents and facts are described in Table 12.

Port Information include ports position and physical characteristics, the facilities and services they offer. Example datasets are:

World Port Index (WPI)[18], an open, freely available and distributable port database maintained by the National Geospatial-Intelligence Agency that contains the locations, the physical characteristics and the facilities and services offered by major ports and terminals worldwide. It contains approximately 3700 entries. It can be useful to interpret events and activities done by vessels closeby ports (e.g., vessels loitering in a loitering area of a port that never enter the port). It is distributed as a PDF report, as an access database and as ESRI shape file.

EMODnet Main Port Statistics[19] by EUROSTAT and Eurofish International Organisation, which include ports data from EMODnet Human activities dataset, with information: passengers, vessels and goods and traffic in main port. Other datasets from the European Marine Observation and Data Network (EMODnet) portal[20], which is an entry point for harmonized marine data generated by 100 organisations, and are free of restrictions on use, include harmonized datasets for coastal maps, human activities such as ports and fishing areas, biological datasets, a digital terrain model for bathymetry, etc. .

EEA European coastline is a high resolution coast line map for analysis from EEA [21] created for highly detailed analysis for geographical Europe. The EEA coastline is a product derived from two sources: EU-Hydro and GSHHG.

Nautical Charts and Maps are cartographic information that are used in support of navigation and define navigation features such as buoys, traffic separation schemes, regulated and restricted areas, accurate bathymetry for coastal areas. Nautical charts can help understand sea traffic. In electronic form they are vector maps. Example datasets are:

Official nautical charts by the International Hydrographic Organisation. This charts are in IHO-S-57, a format for Electronic Navigation Charts (ENC). These maps are not free. Data scheme IHO-S-57 is widely documented [22], and a GDAL driver[23] is available for converting data in other vector formats.

[18] WPI:msi.nga.mil/NGAPortal/MSI.portal
[19] http://www.emodnet-humanactivities.eu/search-results.php?dataname=Main+Ports
[20] EMODnet portal: emodnet.eu
[21] EEA European coastline: www.eea.europa.eu/data-and-maps/data/eea-coastline-for-analysis-1
[22] IHO-S-57 www.s-57.com
[23] http://www.gdal.org/drv_s57.html

Type	Dataset	Description	ID		Dataset and Data provider	Description and data provenance	Spatial coverage	Temporal coverage	Licence and Conditions of use	Costs	Data access and Format	Format description	Quality
Contextual information	Port information	Port characteristics, facilities, services, statistics	C1	main sources	World Port Index (WPI) by National Geospatial-Intelligence Agency	Port database. Location and physical characteristics of, and the facilities and services offered by major ports and terminals world-wide (approximately 3700 entries)	global	updated to 2016	Public releasable and distributable	free	PDF report, Access database and shape file (can be converted in KML) from http://msi.nga.mil/NGAPortal/MSI.portal?_nfpb=true&_pageLabel=msi_portal_page_62&pubCode=0015		
				additional sources	EMODnet Main Port Statistics	Main European ports data from EMODnet Human activities dataset, with information: passengers, vessels and goods and traffic in main port. By EUROSTAT and Eurofish International Organisation	European	Updated up to 2013	Public releasable and distributable	free	http://www.emodnet-humanactivities.eu/search-results.php?dataname=Main+Ports		harmonized at European level
					EEA European coastline	High resolution coast line map for analysis	European	Updated to 2016	Public releasable and distributable	free	http://www.eea.europa.eu/data-and-maps/data/eea-coastline-for-analysis-1	ESRI shapefile and INSPIRE compliant metadata	harmonized at European level
	Nautical charts	Bathymetry and navigation facilities, traffic separation schemes, waiting areas	C2	main sources	OpenSeaMap (www.openseamap.org/index.php?id=openseamap&L=1)	Nautical chart, including beacons, buoys and other navigation aids as well as port information, repair shops and chandlerys	global		ODBL open data common (http://opendatacommons.org/licenses/odbl/CC Share Alike 2.0. Data freely available, usable and distributable	free	IHO-S-57 files from Open Street Map http://planet.openstreetmap.org/	IHO S-57 for Electronic Navication Charts (ENC): www.s-57.com for documentation (GDAL driver for S57 manipulation available http://www.gdal.org/drv_s57.html)	
				additional sources	International Hydrographic Organisation	Nautical charts and navigation features, buoys, traffic separation schemes, regulated and restricted areas, accurate bathymetry for coastal areas	global			commercial	IHO-S-57 files	IHO S-57 for Electronic Navication Charts (ENC): www.s-57.com for documentation (GDAL driver for S57 manipulation available http://www.gdal.org/drv_s57.html)	
	Navigation rules and guidelines	Legal Navigation rules	C3	main sources	IMO	International Regulations for Preventing Collisions at Sea 1972 (Colregs)	global		public	free	PDF http://www.mar.ist.utl.pt/mventura/Projecto-Navios-I/IMO-Conventions%20(copies)/COLREG-1972.pdf, https://en.wikipedia.org/wiki/International_Regulations_for_Preventing_Collisions_at_Sea Information on Future IMO regulation: http://www.lr.org/en/_images/213-35746_Future_IMO_legislation.pdf (by LLOYDS register)		
					IMO	Fishing Vessel Safety Code and Voluntary Guidelines	global				http://www.imo.org/en/OurWork/Safety/Regulations/FishingVessels/Pages/Default.aspx		
		Regulations and guidelines for fishing activities			FAO	Conduct codes, Port State Measures Agreement (PSMA)	global						
	Regulated areas	Fishing areas	C4	main sources	European Regulated Fishing Areas	European Regulated Fishing Areas from European Atlas of the Sea	European	Updated to 2013	Public releasable and distributable	free	WebGIS European Atlas of the Sea http://ec.europa.eu/maritimeaffairs/atlas/maritime_atlas/#lang=EN;p=w;pos=33.156:50.493:4;bkgd=5:1;gra=0;mode=1;theme=52:0.8:1:0;time=2013;		
					FAO and ICES Fishery areas from EMODnet	FAO Fishery Statistical Areas, Fish Catches by FAO Fishery Statistical Areas, and ICES Statistical Areas through Human activities datasets from EMODnet	European		Public releasable and distributable	free	WMS and WFS services: http://www.emodnet-humanactivities.eu/search-results.php?dataname=FAO+Fishery+Statistical+Areas, http://www.emodnet-humanactivities.eu/search-results.php?dataname=Fish+Catches+by+FAO+Fishery+Statistical+Areas, http://www.emodnet-humanactivities.eu/search-results.php?dataname=ICES+Statistical+Areas PDF FAO http://www.fao.org/3/a-az126e.pdf	Web Map Service: http://www.opengeospatial.org/standards/wms Web Feature Service: http://www.opengeospatial.org/standards/wfs	harmonized at European level
		Marine protected areas	C5	main sources	Natura2000 from European Environmental Agency (EEA)	Natura2000 Database of European Marine Protected Areas (MPAs), including marine areas from European Environmental Agency. Cf. also MPAs designated under the Regional Sea Conventions	European		Public releasable and distributable	Free	Access database (MDB) and CSV file from http://www.eea.europa.eu/data-and-maps/data/natura-6, http://www.eea.europa.eu/data-and-maps/figures/mpas-designated-under-the-regional		
				additional sources	Environmental datasets from European Environmental Agency (EEA)	Environmental datasets harmonized at European level, including biodiversity, waste, sediments, high resolution coastline	European			Free	http://www.eea.europa.eu/data-and-maps/find/global?c12=marine&search=Search#c1=Data&c1=Graph&c1=Indicator&c1=Interactive+data&c1=Interactive+map&c1=Map&c6=&c0=12&b_start=0&c12=marine		

Figure 11: Contextual datasets: Ports and Navigation Rules

Open Sea Map [24] is an open and free nautical chart, including beacons, buoys and other navigation aids as well as port information, repair shops and chandlerys. It can integrate the World Port Index supporting the maritime use case with information on port facilities and provide also information on vessel routes close to ports. It is an Open Street Map project and data are freely available, usable and distributable according to the ODBL open data common licence [25].

Navigation rules of interest for datAcron include the International Regulations for Preventing Collisions at Sea 1972 (Colregs)[26] from IMO. IMO and FAO have developed a series of guidelines for fishing vessel safety, such as the Fishing Vessel Safety Code and Voluntary Guidelines[27] and Conduct codes, Port State Measures Agreement (PSMA).

Regulated areas of interest for the Fishing Monitoring use case include fishing areas and marine protected areas where fishing is allowed and forbidden, respectively. For Europe, reference datasets are:

European regulated fishing areas are included in the recent European proposal on the conservation of fishery resources and the protection of marine ecosystems through technical measures. These are given as thematic as well as spatial information, together with the coordinates of the regulated fishing areas. This dataset can be used as a contextual information for the IUU fishing scenario.

Fishery statistical areas and fish catches from FAO[28] and ICES[29].

Marine protected areas in Europe are defined in the *NATURA2000* ecological network of protected areas and freely downloadable from the European Environmental Agency (EEA) website [30]. Other environmental biodiversity datasets (*e.g.*, marine biodiversity, waste, sediments) harmonized at European level and freely downloadable in different formats from the EEA website [31] that can be useful to further develop the maritime sustainable development scenario.

Vessel registers of interest for datAcron Fishing monitoring use case include, for instance, database of fishing vessels. Datasets potentially of use are:

The European Commission Community Fishing fleet register[32], a European register of official fishing vessels maintained by the European Commission. This dataset can be used as a contextual information for the IUU fishing scenario.

IMISG vessel database, the vessel database developed by IMISG, that will be shared with the consortium for internal use.

Equasis database of world' merchant fleet and shipping companies[33], developed on voluntary basis by a number of maritime administrations. It is free without any restriction on use, but not downloadable in bulk.

[24] Open Sea Map: www.openseamap.org, Downloadable from: planet.openstreetmap.org
[25] opendatacommons.org/licenses/odbl/CC Share Alike 2.0
[26] http://www.mar.ist.utl.pt/mventura/Projecto-Navios-I/IMO-Conventions%20(copies)/COLREG-1972.pdf
[27] http://www.imo.org/en/OurWork/Safety/Regulations/FishingVessels/Pages/Default.aspx
[28] http://www.emodnet-humanactivities.eu/search-results.php?dataname=FAO+Fishery+Statistical+Areas and http://www.emodnet-humanactivities.eu/search-results.php?dataname=Fish+Catches+by+FAO+Fishery+Statistical+Areas
[29] http://www.emodnet-humanactivities.eu/search-results.php?dataname=ICES+Statistical+Area
[30] NATURA2000: natura2000.eea.europa.eu
[31] EEA biodiversity data: biodiversity.europa.eu/data
[32] Community fleet register http://ec.europa.eu/fisheries/fleet/index.cfm
[33] Equasis http://www.equasis.org/

Type	Dataset	Description	ID		Dataset and Data provider	Description and data provenance	Spatial coverage	Temporal coverage	Licence and Conditions of use	Costs	Data access and Format	Format description	Quality
Contextual information	Vessel registers	Vessel Registers	C6	main sources	European Commission Community Fishing fleet register	Community Fishing fleet register including declared fishing vessels	European		open data	free	web service or text file, http://ec.europa.eu/fisheries/fleet/index.cfm		
					IMISG Vessel database	Vessel database	Global		Restricted used within datAcron consortium				
					Equasis, Maritime administrations adhering to the MoU	Database of world' merchant fleet and shipping companies	Global		Registration required, only for personal use and not to be included in a retrieval system	free	web service http://www.equasis.org/		
					LLOYDS list intelligence vessel database(http://www.lloydslistintelligence.com/)	Vessel register from LLOYDS insurance company, shipping companies and ports	global		commercial distribution	commercial	http://www.lloydslistintelligence.com/		most complete and reliable register (12000 vessels, 163,500 shipping company, 2800 ports)
	Blacklists	List of vessels involved in IUU fishing	C7	main sources	Blacklist of IUU fishing vessels, Greenpeace	Official Blacklist of IUU fishing vessels compiled by Greenpeace and from publicly available official registries of IUU vessels and companies (Official List, GreenPeace list, fused list including more than 200 vessels)	Global		public	free	Online resources. Fused list http://www.greenpeace.org/international/en/campaigns/oceans/pirate-fishing/Blacklist1/Browse-all-blacklists/	Semistructured information	Includes also official sources of information. Information on vessel identifiers can be partial.
	Fisheries and vessel facts	Vessel and fishing news	C8		VesselTracker (www.vesseltracker.com)	Live feed of vessel news about incidents, illegal activities	Global	live	public	free	Web service https://www.vesseltracker.com/en/News/Home.html	unstructured, web site aggregator of news from media	can be spoofed
					FleetMoon (https://www.fleetmon.com)	Live feed of vessel news about incidents, illegal activities	Global	live	public	free	https://www.fleetmon.com/maritime-news/	unstructured, web site aggregator of news from media	
					Media	News from media and social media about detection of illegal activities and accidents at sea	Global	live	public		online resource, PDF documents	unstructured, web site content	
		Accident reports, official reports on illegal activities	C9		Various authorities	Technical reports by authorities about detection of illegal activities and accidents at sea	Global		public	free	online resource, PDF documents	unstructured, web site content	official source of information
					Interpol	Interpol purple notices: reports on illegal activities detected by authorities worldwide	Global	live. Archive 2013-2015 on fishing illegal activities	public	free	online resource, PDF reports e.g., http://www.interpol.int/INTERPOL-expertise/Notices/Purple-notices-%E2%80%93-public-versions/2014	Semistructured text including type of incident, date, location, country and description of modus operandi	official source of information
					OFFICE OF NAVAL INTELLIGENCE	Worldwide Threat to Shipping (WTS) messages: The Worldwide Threat to Shipping (WTS) message provides info on piracy threats to, and criminal action against, merchant vessels worldwide, to inform merchant mariners and naval forces. Includes also report of discovered illegal activities, such as smuggling	Global	2010, 2011, 2014, 2016	public	free	TXT and PDF files	Unstructured text, including location, date and a textual description of the attack	official source of information
		Piracy reports	C10	additional sources	ASAM	database of reports on piracy events	Global	updated to 2014	public		Access database (MDB), Esri shape file (SHP), Google KMZ	Access database (MDB), Esri shape file (SHP), Google KMZ	Contains almost 7000 records. Include information on area, date, type of aggression and type of vessel
					IMB	map of piracy events (attacks, attempted,boarded)	Global	2011	public		Comma Separated Value (CSV) and PDF	CSV format: incident id, latitude, longitude, vessel status, date	Partial. Include information on type of vessels
					Interpol	Global Maritime Piracy Database from Interpol	Global		public				official source of information
						Pirate attacks database	Global	2007, other dates	public		Comma Separated Value (CSV)	CSV format: Longitude, Latitude, Steaming, Location, Anchored, latitude, longitude, CSV format: Not Stated, Knives, Other Weapons, Longitude, Guns, Latitude,	Partial, no vessel identifier

Figure 12: Contextual datasets: Vessel registers and lists

LLOYDS list intelligence[34], the vessel register from LLOYDS insurance company, including information on shipping companies and ports

Vessels blacklists are official list of vessels that have been involved in some illegal activities. Greenpeace provide, free for use and open to the public, the *Blacklist of IUU fishing vessels*, a list of fishing vessels involved in IUU based on information extracted from publicly available official registries of IUU vessels and companies[35]. The European Union strategy to fight IUU fishing[36] include to issue two blacklists. The first includes vessels engaged in IUU fishing as identified by Regional Fisheries Management Organisations and the second the list of States that are lenient towards them.

Vessels and fishing facts sources include official reports drafted by authorities on navigation incidents, fisheries facts and illegal activities, as well as news from media and social media. This information can be used to define ground truth datasets. Most of these sources are unstructured:

Web services of vessel and maritime news, such as news from VesselTracker[37] and FleetMoon[38].

Worldwide Threat to Shipping (WTS) messages from the Office of Naval Intelligence provides mainly info on piracy threats to, and criminal action against, merchant vessels worldwide, to inform merchant mariners and naval forces and include report of discovered illegal activities, such as smuggling. WTS messages are available in .txt and .pdf files.

Interpol purple notices[39] are reports, available on the web, on illegal activities detected by authorities worldwide, including reports on illegal fishing events.

Reports and datasets on piracy events are compiled by ASAM, IBM, Interpol.

4.3.2 Weather and ocean data

Weather data and ocean data from forecast models and from observations (e.g., in-situ sensor data), which are openly available from several providers, can help validate analysis results and reduce false alarm rate, for example identifying sea and weather conditions that force vessels to change direction or modify their normal behaviour. They can also be used to characterize seasonal trends in traffic routes, and to contextualise movement parameters such as speed of vessels.

The reference source of harmonized oceanographic data in Europe is by far the Copernicus Marine Environment and Monitoring Service (CMEMS) [40], developed by the EU as part of the European Programme for the establishment of a European capacity for Earth Observation and Monitoring. This operative service provides an interactive catalogue of updated oceanographic products, including Regional and Global Ocean model (forecast and reanalysis) and observations produced by the network of oceanographic centres in Europe. 140 data products world wide can be downloaded for free after registration, including data on the Mediterranean, the North West Coast of Europe and the Iberian Peninsula. Depending on the dataset, historical forecast data may be available, as well as reanalysis of past forecast.

In particular, weather and ocean datasets that can be of helpful to the maritime use case include:

Mediterranean Sea Physics Analysis and Forecast (MFS) data products by INGV Mediterranean Forecasting System[41]. The physical component (ocean variables) produced daily by the

[34] LLOYDS database http://www.lloydslistintelligence.com/
[35] http://blacklist.greenpeace.org/0/vessel/list
[36] http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1466764546838&uri=URISERV:pe0005
[37] VesselTracker news: https://www.vesseltracker.com/en/News/Home.html
[38] FleetMoon news: https://www.fleetmon.com/maritime-news/
[39] http://www.interpol.int/INTERPOL-expertise/Notices/Purple-notices-%E2%80%93-public-versions/2014
[40] Copernicus Marine Environment and Monitoring Service: marine.copernicus.eu
[41] MediterraneanForecastingSystem-MFS(http://medforecast.bo.ingv.it)

Mediterranean Forecasting System (Med-currents) include 10 day-forecast on information on Currents (velocity), Temperature, Salinity, Sea-Level (Sea surface height), Significant Wave Height, Wind stress. The model solutions are corrected by the variational assimilation of temperature and salinity vertical profiles and along track satellite Sea Level Anomaly observations[42]. The quality assessment[43] of the system is monitored weekly by the calculation of the root mean square statistics of difference between observations and model background fields (so-called misfits)[44]. The system is also evaluated by considering independent data at fixed stations around the Mediterranean Sea (http://calval.bo.ingv.it) MFS is distributed through the CMEMS catalogue[45]. It is updated daily. Historical data are available from 2013-01-01, and variable values are available as daily mean and hourly mean. Its geographical coverage is Latitude North 45.937, Latitude South 30.187, Longitude East 36.25, Longitude West -15, with a resolution of 1/16 degree (*i.e.*, ca. 6-7 km). The format of files downloadable from the CMEMS catalogue is netCDF, a binary multidimensional array format widely used in the oceanographic domain[46].

Global Ocean Wind observations distributed through the CMEMS service[47]. Observations are distributed every 6 hours, with horizontal resolution of 0.25x0.25 degrees. Daily and monthly mean are available from 2007. Data are in netCDF.

*Forecast and observations for the Adriatic and for the Ionian Sea.*Forecast and simulation data on sea conditions are produced by the *Adriatic Forecasting System (AFS)* [48] by the *Ionian Forecasting System* [49], which are based on Princeton Ocean Model (POM). Variables include Currents, Temperature, Salinity and Sea Level. AFS makes available one week of data, but simulation are availables from 2010, For IFS three months of data are available. The *IONIO in-situ database* includes observations from buoys and ship of opportunity [50], with historical data from 1986. Data are in netCDF.

EMODnet datasets include energy data from *Seabed Habitat* (wind, waves and currents) [51] and physics data (Sea water temperature, salinity or density, water currents, level, Waves and winds, Light attenuation, Atmospheric parameters at sea level, HF radar data) [52]. EMODnet data are harmonized at European Level and are distributed as netCDF or through geospatial map services (OGC Web Map Service).

The *National Oceanic and Atmospheric Administration (NOAA)* shares through its catalogue [53] public releasable global meteorological and oceanographic datasets from cooperating networks of ships and buoys.

The *European Centre for Medium-Range Weather Forecasts (ECMWF)* distributes Ocean Wave and Atmospheric model data, mainly historical. The Ocean Wave ERA-Interim model data [54] are available daily until 2015, and the ERA-20C model[55] available until 2010.

[42] MFS data product documentation http://marine.copernicus.eu/documents/PUM/CMEMS-MED-PUM-006-001.pdf

[43] http://marine.copernicus.eu/documents/QUID/CMEMS-MED-QUID-006-001.pdf

[44] http://medforecast.bo.ingv.it/mfs-copernicus-evaluation/

[45] MFS: marine.copernicus.eu/web/69-interactive-catalogue.php?option=com$_$csw&view=details&product$_$id=MEDSEA$_$ANALYSIS$_$FORECAST$_$PHYS$_$006$_$001$_$a

[46] netCDF documentation: http://www.unidata.ucar.edu/software/netcdf/docs/ (full documentation and APIs), data model http://www.unidata.ucar.edu/software/netcdf/docs/netcdf_data_set_components.html (data model)

[47] marine.copernicus.eu/web/69-interactive-catalogue.php?option=com_csw&view=details&product_id=WIND_GLO_WIND_L4_NRT_OBSERVATIONS_012_004

[48] www.ionioproject.eu

[49] ionioproject.hcmr.gr

[50] www.mediterraneanmarinedata.eu/ionio/home.htm

[51] http://www.emodnet-seabedhabitats.eu/default.aspx?page=1934

[52] http://www.emodnet-physics.eu/Map/service/Catalogue.aspx

[53] www.ndbc.noaa.gov/data

[54] apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/

[55] ERA-20C apps.ecmwf.int/datasets/data/era20c-daily/levtype=sfc/type=an/)

Atmospheric ECMWF datasets are availablein different fomats (WMO, FM-92 GRIB, NetCDF) from web[56] and trhough the ECMWF Web API [57].

4.4 Relation Between Data and Scenarios

The next figures summarise the scenarios as related to the datAcron work packages and big data challenges. Some examples of relevant information needs are provided, as well as some relevant data sources which could be used.

4.5 Synthetic and pseudo-synthetic datasets

In Deliverable D5.1, we proposed that the quality of the Recognized Maritime Picture can be assessed according to the main five criteria of *Completeness*, *Accuracy*, *Clarity*, *Continuity* and *Timeliness*. This does not mean that these will be the only criteria to be considered, but that these are recognized by the operational community for target detection and tracking, and we propose to use them as high-level criteria for MSI detection. That does not exclude other performance criteria closer to the processing, that will be proposed and defined.

Each of the five criteria above will be defined relatively to

(1) a given area,

(2) a given period of time, and

(3) a given Maritime Situational Indicator (MSI).

Hence, for a given scenario, the user expects datAcron algorithms to provide answers to the relevant MSIs with a quality defined by these five dimensions.

Also, because the user chooses the MSIs to detect the scenario-related events (collision, vessel in distress, smuggling, etc), another layer of performance criteria related to human factor tasks while dealing with scenario-events, will be considered. Figure 17 illustrates the two levels of assessment of datAcron: the MSI level and the scenario level. The datAcron algorithms will be evaluated along both the operational and technical criteria (some may overlap).

The set of real data to be processed will need to be supplemented by either additional information or synthetic data. Indeed, while the criteria of *Clarity*, *Continuity* and *Timeliness* can be assessed without access to any ground truth, the criteria of *Completeness* and *Accuracy* some ground truth information is required.

We will thus need to create datasets, either *purely synthetic* based on some motion models or *pseudo-synthetic* by modifying existing real data with a controlled process. The methodology will be further described in the experimental plan, and includes:

- Having experts directly labelling vessel trajectories with ground truth information, such as “on-route” and “off-route” vessels (see also Section 4.1.1);
- Generating purely synthetic MSIs of any kind based on proper models (motion models, statistical error models, etc);
- Modifying parts of real datasets and
  - including synthetic anomalies or MSIs, much in line with what was proposed in [4] and illustrated in Figures 18,

[56] apps.ecmwf.int/datasets/
[57] apps.ecmwf.int/datasets/

Type	Dataset	Description	ID		Data provider and dataset	Description and data provenance	Spatial coverage	Temporal coverage	Licence and Conditions of use	Costs	Data access and Format	Format description	Quality
Contextual Information (Sea and Weather Conditions)	Ocean conditions (forecast, simulation and observations)	Data on sea physical state including sea level, waves and currents, which can affect navigation	E1	main sources	Mediterranean Sea Physics Analysis and Forecast, Mediterranean Forecasting System - MFS (http://medforecast.bo.ingv.it)	Physical component of the Mediterranean Forecasting System (Med-currents) produced by INGV (http://medforecast.bo.ingv.it/). Includes information on Currents (velocity), Temperature, Salinity, Sea-Level (Sea surface height), Significant Wave Height, Wind stress. The model solutions are corrected by the variational assimilation of temperature and salinity vertical profiles and along track satellite Sea Level Anomaly observations.	mediterranean and close atlantic ocean. (Spatial resolution 0.06 degree, from -5500.0m to 0.0m, CRS=EPSG:5714)	10 day forecast with daily updates (Temporal resolution: Daily mean, Hourly mean). Historical data from 2013	Public releasable	Free. Registration to Copernicus Marine Environment Monitoring Service (CMEMS, marine.copernicus.eu) required	NetCDF files shared through marine.copernicus.eu Interactive catalogue and services, FTP and download with Phyton scripts (MOTU client) from marine.copernicus.eu/web/69-interactive-catalogue.php?option=com_csw&view=details&product_id=MEDSEA_ANALYSIS_FORECAST_PHYS_006_001_a. OGC Web Map Service (WMS) and Catalogue Service for the Web (CSW) services are also available	NetCDF binary multidimensional array format: http://www.unidata.ucar.edu/software/netcdf/docs/ (full documentation and APIs), http://www.unidata.ucar.edu/software/netcdf/docs/netcdf_data_set_components.html (data model) MFS data product dccumentation http://marine.copernicus.eu/documents/PUM/CMEMS-MED-PUM-006-001.pdf	Quality information http://marine.copernicus.eu/documents/QUID/CMEMS-MED-QUID-006-001.pdf, The quality assessment of the system is monitored weekly by the calculation of the root mean square statistics of difference between observations and model background fields (so-called misfits): http://medforecast.bo.ingv.it/mfs-copernicus-evaluation/ The system is also evaluated by considering independent data at fixed stations around the Mediterranean Sea (http://calval.bo.ingv.it)
				main sources	Regional and Global Ocean model data (forecast and reanalysis) and observations from Copernicus Marine Service (marine. copernicus.eu)	Physical variables of ocean state (Currents, Temperature, Salinity, Sea level , Significant Wave Height)	Global. Regional: Mediterranean, North Sea, Baltic Sea, Iberian Sea, Black Sea, Artic Sea	Daily updates. Historical data available	Public release	Free. Registration to Copernicus Marine Environment Monitoring Service (CMEMS, marine.copernicus.eu) required	NetCDF files shared through marine.copernicus.eu Interactive catalogue and services, FTP and download with Phyton scripts (MOTU client). Catalogue Service for the Web (CSW)	NetCDF binary multidimensional array format: http://www.unidata.ucar.edu/software/netcdf/docs/ (full documentation and APIs), http://www.unidata.ucar.edu/software/netcdf/docs/netcdf_data_set_components.html (data model)	
				additional sources	Global Ocean Wind	Winds	global. Hor. res of .25x.25 degrees	daily and monthly mean from 2007	Public release	Free. Registration to Copernicus Marine Environment Monitoring Service (CMEMS, marine.copernicus.eu) required	NetCDF files shared through marine.copernicus.eu Interactive catalogue and services, FTP and download with Phyton scripts (MOTU client). marine.copernicus.eu/ web/69-interactive-catalogue.php?option= com_csw&view=details&	NetCDF binary multidimensional array format: http://www.unidata.ucar.edu/software/netcdf/docs/ (full documentation and APIs), http://www.unidata.ucar.edu/software/netcdf/docs/netcdf_data_set_components.html (data model)	
				additional sources	Adriatic Forecasting System - AFS (www.ionioproject.eu) Adriatic Ionian Forecasting System - AIFS (http://oceanlab.cmcc.it/aifs/) Southern Adriatic Northern Ionian coastal Forecasting System - SANIFS (http://oceanlab.cmcc.it/sanifs/)	Forecast and simulation data on sea conditions based on Princeton Ocean Model (POM). Includes Currents, Temperature, Salinity and Sea Level	adriatic and ionian sea	one week (last 7 days). AFS Simulations available from 2010	Public releasable	Free. Require registration (ocean-lab@cmcc.it)	NetCDF files via OpenDAP or FTP from http://tds.tessa.cmcc.it/thredds/catalog.html	NetCDF binary multidimensional array format: http://www.unidata.ucar.edu/software/netcdf/docs/ (full documentation and APIs), http://www.unidata.ucar.edu/software/netcdf/docs/netcdf_data_set_components.html (data model)	NA
				additional sources	Ionian Forecasting System - IFS (ionioproject.hcmr.gr)	Forecast model data from SANI2 (Southern Adriatic – Northern Ionian Sea 2) model, based on Princeton Ocean Model (POM). Includes Temperature, Salinity, Currents, Sea Surface Height, Significant Wave Height	ionian sea	last 3 months	Public releasable	Free	NetCDF files, available via OpenDAP from tethys.hcmr.gr/opendap/hyrax/ionio/	NetCDF binary multidimensional array format: http://www.unidata.ucar.edu/software/netcdf/docs/ (full documentation and APIs), http://www.unidata.ucar.edu/software/netcdf/docs/netcdf_data_set_components.html (data model)	NA
				additional sources	IONIO In-situ database (www.mediterranean-marinedata.eu/ionio/home.htm)	Observations from buoys and ship of opportunity	adriatic sea	From 1986	Public releasable and Confidential data	Authorization is required for download	NetCDF and Medatlas files from web interface http://www.mediterranean-marinedata.eu/ionio/home.htm	NetCDF binary multidimensional array format: http://www.unidata.ucar.edu/software/netcdf/docs/ (full documentation and APIs), http://www.unidata.ucar.edu/software/netcdf/docs/netcdf_data_set_components.html (data model) MEDATLAS format: http://www.ifremer.fr/mater/formats.htm	NA
				additional sources	EMODnet Energy data from Seabed Habitat	Seabed Habitat datasets, including Energy data: wind, waves and currents harmonized at European Level. Based on data from	european		Public releasable	Free	WMS service http://www.emodnet-seabedhabitats.eu/default.aspx?page=1934	Web Map Service: http://www.opengeospatial.org/standards/wms	NA
				additional sources	EMODnet Ocean Physics data	Sea surface temperature, Sea level, Salinity or density, water currents, Waves and winds, Light attenuation, Atmospheric parameters at sea level, HF radar data from Spain maps of total velocity of the surface current. Provided by EuroGOOS associates and their regional operational	european	last 60 days	Public releasable	Free	NetCDF from THREDDS catalogue http://thredds.emodnet-physics.eu/thredds/catalog.html Web GIS: http://www.emodnet-physics.eu/map/, WMS, WFS and other Web Services available HF Radar data catalogue http://thredds.emodnet-physics.eu/thredds/HFRADARCatalog.html	NetCDF binary multidimensional array format: http://www.unidata.ucar.edu/software/netcdf/docs/ (full documentation and APIs), http://www.unidata.ucar.edu/software/netcdf/docs/netcdf_data_set_components.html (data model) Web Map Service:	
				additional sources	ECMWF Ocean Wave model data	ERA-Interim model (ocean waves, atmospheric variables), ERA-20C model	global	ERA-Interim: daily until 2015. ERA-20C model: daily	Public, limitations apply	Free, limitations apply on use. Registration is required https://apps.ecmwf.int/registration/.	apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/ and apps.ecmwf.int/datasets/data/era20c-daily/levtype=sfc/type=an/		

Figure 13: Contextual datasets: wheather and ocean datasets

		User information needs				Big data challenges				
Use Case: Monitoring fishing activities		[WP.1] Scalable integration and management of data from disparate and heterogeneous sources	[WP.2] Real-time detection and forecasting of trajectories	[WP.3] Real-time event recognition and forecasting	[WP.4] Real-time interactive visual analytics	Variety		Veracity	Volume	Velocity
3	Illicit activities					Types	S			
3.1	Migrants/ refugees/ Human trafficking		Estimation of	Detection of migrants patterns (fishing vessel that is not fishing and is travelling in common migrants/ smuggling routes, outside fishing areas, not transmitting AIS outside territorial waters)	Visualization of smuggling/migrants routes	AIS	P1	need to compare AIS and SAR to detect small vessels that are not transmitting AIS		
						SAR	P2			
						Vessel register, check vessel characteristics (size, flag of convenience, not equipped for fishing, vessel is "single box"	C6			
						Smuggling or migrants routes from	C8, C9	News and reports are not complete		
3.2	IUU fishing, ish trafficking, fishing spoofing, ...		Estimation of vessel position in an area where illegal fishing is applied	Recognition of fishing patterns (drifting within regulated areas, switching off AIS, static vessel information matching with fishing vessel)		AIS	P1	Deception of AIS signal		
				Any inconsistency between type of vessel and behaviour		IUU blacklist	C7			
						Fishing areas	C4			
						News and reports on IUU fishing	C8,C9			
						Vessel register, check vessel characteristics (type, flag, owner)	C6			
						Reports on illegal fishing events	C8,C9			

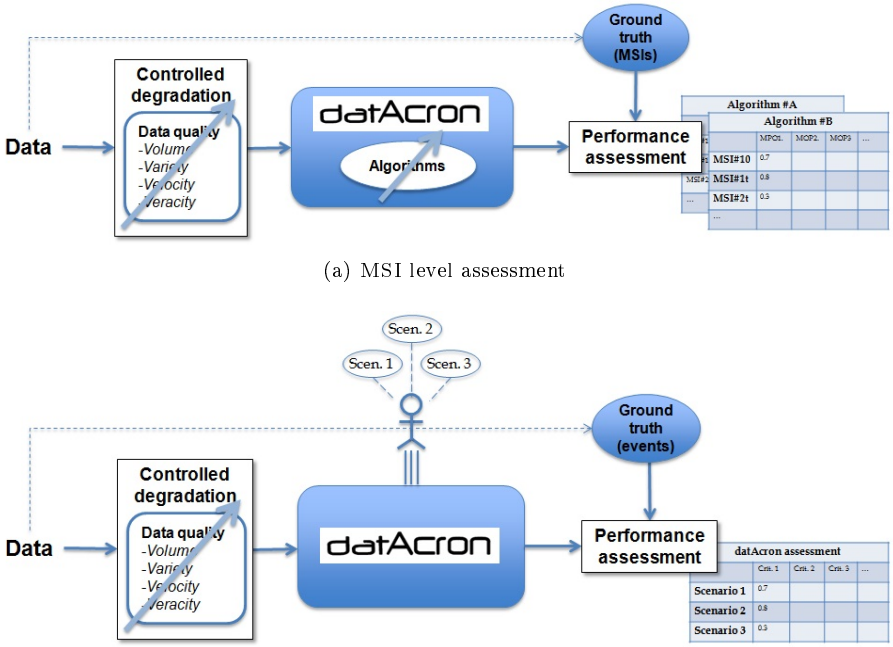
Figure 14: Illicit Activities

		User information needs				Big data challenges				
	Use Case: Monitoring fishing activities	[WP.1] Scalable integration and management of data from disparate and heterogeneous sources	[WP.2] Real-time detection and forecasting of trajectories	[WP.3] Real-time event recognition and forecasting	[WP.4] Real-time interactive visual analytics	Variety		Veracity	Volume	Velocity
1	Secured fishing					Types	S			
1.1	**Collision prevention**: Protect fishing vessels from possible collision with other large vessels	*See the Big data challenges columns - Veracity: data cleaning*	Fishing vessels current position	Detection of fishing areas from historical data	Visualisation of fishing areas: fishing density vs. seasons vs. types of fish vs. time	AIS data possibly from different sources (terrestrial, satellite), different types of AIS (class A, B), with partially overlapping coverage.	P1	No AIS emission when fishing, AIS turning off or duping	AIS current streams, historical streams	Varying temporal resolution of AIS contacts: e.g., fishing vessel every 30', cargo every 5'
				Recognition of fishing patterns: drifting within regulated areas, switching off AIS in regulated areas, static vessel information matching with fishing vessel	Adaptive selection of the area of interest					
					Visualisation of preferred MPA management options (adding or removing protected areas)	Fishing areas, area of interest for collision prevention scenario	C4	Variability of fish location according for instance to season		

Figure 15: Secured Fishing

		User information needs				Big data challenges				
	Use Case: Monitoring fishing activities	[WP.1] Scalable integration and management of data from disparate and heterogeneous sources	[WP.2] Real-time detection and forecasting of trajectories	[WP.3] Real-time event recognition and forecasting	[WP.4] Real-time interactive visual analytics	Variety		Veracity	Volume	Velocity
2	Sustainable development					Types	S			
2.1	**Protection of marine protected areas(MPAs)**: Protect specific areas from fishing activities and possibly maritime traffic		Estimation of fishing vessels location, starting from previous vessel positions (e.g., fishing vessel leaving a port, heading towards a protected area and stopping AIS outside a regulated fishing area)	Recognition of fishing patterns (drifting within regulated areas, switching off AIS in	Visualisation of preferred MPA management options (adding or removing protected areas)	AIS	P1	No AIS emission when fishing, AIS turning off or duping		
						Vessel register	C6			
						Port database	C1			
						Map of the protected areas	C5			
						Fishing areas	C4			
			Prediction of future fishing vessels location for interception			AIS	P1			
2.2	**Fishing pressure on areas**: Predict and prevent the pressure on fishing			Recognition of fishing patterns (drifting within regulated areas, switching off AIS in regulated areas, static vessel information matching with fishing vessel)	Visualisation of preferred MPA management options (adding or removing protected areas)	AIS	P1	No AIS emission when fishing, AIS turning off or duping		
						SAR imagery	P2			
						Vessel register	C6			
						fishing areas	C4	Variability of fish location according for instance to season		
						Nautical charts	C2			
				Detection of fishing areas from historical data	Visualisation of fishing areas: fishing density vs. seasons vs. types of fish vs. time	AIS	P1			
						SAR imagery	P2			

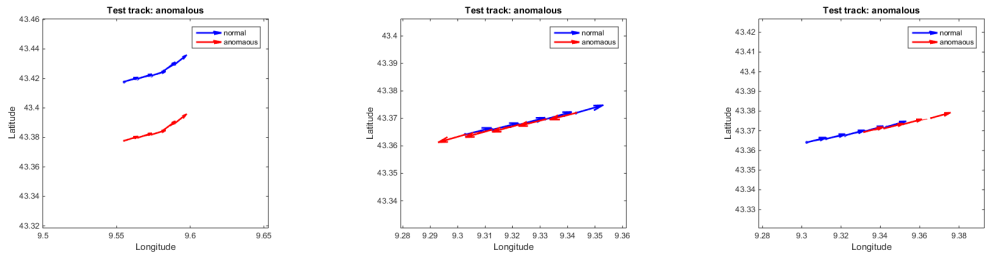
Figure 16: Sustainable Development



(a) MSI level assessment

(b) Scenario level assessment

Figure 17: Sketch of datAcron evaluation



(a) Positional anomaly: shifted track (b) Directional anomaly: reverse flow track (c) Kinematic anomaly: high speed track

Figure 18: Example of pseudo-synthetic anomalies

– realigning real events to our area and timescale of interest. For instance, real data corresponding to a real collision happened in another part of the world could be re-aligned in time and space with our area of interest.

References

[1] L. Cazzanti and A. Davoli. Big data architectures in support of computational maritime situational awareness – A case study in port traffic analysis. Technical Report CMRE-FR-2015-021, NATO CMRE, 2015.

[2] Jim Llinas, Anne-Laure Jousselme, and Geoff Gross. Context as an uncertain source. In L. Snidaro, J. Garcia, J. Llinas, and E. Blasch, editors, *Context Enhanced Information Fusion: Improving real world performance with domain knowledge*. Springer, 2015.

[3] Pete McBreen. Using use cases for requirements capture. www.mcbreen.ab.ca, McBreen.Consulting, 1998.

[4] Giuliana Pallotta and Anne-Laure Jousselme. Data-driven detection and context-based classification of maritime anomalies. In *Proceedings of the 18th International Conference on Information Fusion*, Washington, D. C. (USA), July 2015.

[5] C. Ray, C. Iphar, A. Napoli, R. Gallen, and A. Bouju. Deais project: Detection of ais spoofing and resulting risks. In *OCEANS'15 MTS/IEEE, Genoa, Italy*. IEEE, 2015.