



**HAL**  
open science

## Use case design and big data analytics evaluation for fishing monitoring

Cyril Ray, Elena Camossi, Richard Dréo, Anne-Laure Jouselme, Clément Iphar, Maximilian Zocholl, Melita Hadzagic

► **To cite this version:**

Cyril Ray, Elena Camossi, Richard Dréo, Anne-Laure Jouselme, Clément Iphar, et al.. Use case design and big data analytics evaluation for fishing monitoring. OCEANS 2019 - Marseille, Jun 2019, Marseille, France. pp.1-8, 10.1109/OCEANSE.2019.8867554 . hal-03853392

**HAL Id: hal-03853392**

**<https://hal.science/hal-03853392v1>**

Submitted on 15 Nov 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Use case design and big data analytics evaluation for fishing monitoring

Cyril Ray  
French Naval Academy  
Brest, France  
cyril.ray@ecole-navale.fr

Elena Camossi  
NATO STO CMRE  
La Spezia, Italy  
elena.camossi@cmre.nato.int

Richard Dréo  
NATO STO CMRE  
La Spezia, Italy  
richard.dreo@cmre.nato.int

Anne-Laure Joussetme  
NATO STO CMRE  
La Spezia, Italy  
anne-laure.joussetme@cmre.nato.int

Clément Iphar  
NATO STO CMRE  
La Spezia, Italy  
clement.iphar@cmre.nato.int

Maximilian Zocholl  
NATO STO CMRE  
La Spezia, Italy  
maximilian.zocholl@cmre.nato.int

Melita Hadzagic  
OODA Technologies  
Montreal, Quebec, Canada  
melita.hadzagic@ooda.ca

**Abstract**—This work reports on the maritime use case development and evaluation of the prototype of the project *Big Data Analytics for Time Critical Mobility Forecasting (datAcron)*, which was funded by the European Union Horizon 2020 programme. *datAcron* aimed at developing novel methods to detect threats and abnormal activities in streams of large number of moving objects in wide maritime and aerial areas. The *datAcron* maritime use cases, which were scoped in the *Maritime Work Package (WP5)*, geared on fishing activities monitoring, aligning with the European Union Maritime Security Strategy. The maritime use cases aimed at providing research partners with maritime domain knowledge to frame in some operational context novel approaches designed during the project. Moreover, the maritime use cases provided a basis for the *datAcron* maritime prototype setup and evaluation of big data solutions. Six scenarios falling under three use cases were therefore developed, identifying relevant data sources, situational indicators as well as corresponding operational tasks. This paper overviews the design, development and evaluation of the *datAcron* maritime use case.

**Index Terms**—Maritime big data, Maritime Situational Awareness (MSA), Maritime Situational Indicators (MSI), fishing monitoring, maritime use case, collision avoidance, open maritime data, evaluation framework

## I. INTRODUCTION

Vessel Traffic Systems receive and visualise streams of vessel positions, which together contribute to a real-time picture of the maritime traffic. This large *volume* of data arrives to the systems with high *velocity* and integrates *various* sensors observations with no guaranteed *veracity*, which is challenging to analyse for operators and decision makers. *Maritime Situation Awareness (MSA)* is obtained assessing the maritime traffic against contextual information, also inferred from the analysis of historical data (*e.g.* maritime patterns-of-life, routes and loitering areas). Besides, MSA requires not only detecting, tracking and classifying vessels but also

This work was supported by project *Big Data Analytics for Time Critical Mobility Forecasting (datAcron)*, which has received funding from the European Unions Horizon 2020 research and innovation programme under grant agreement No. 687591. The authors would like to thank LCdr Eric Reardon, Cdr Mike Ilteris, Karna Bryan and Christophe Claramunt for their contribution in the original ideas of the use cases and scenarios.

predicting their behaviour [4]. The main research objectives of this three-year project *Big Data Analytics for Time Critical Mobility Forecasting (datAcron)*<sup>1</sup> that has concluded in 2018 comprised the development of highly scalable methods for advancing spatio-temporal data integration and management to support the generation of high-rate compressed moving object trajectories, *i.e.* *synopses* [10], from spatio-temporal streaming data and their accurate reconstruction, along with the real-time recognition and forecasting of critical events and their interactive visualisation and analysis. The *datAcron Maritime Work Package* bolstered the scientific and system developments to focus on relevant operational questions and assessed the the integrated prototype of the project against a challenging and realistic fishing monitoring use case, aligning with key features of the European Union Maritime Security Strategy<sup>2</sup> published in December 2014. Furthermore, the developed use case emphasised the role of the user and his/her interaction with the system, allowing the consideration of the user’s needs in the design of the algorithms, and a human-centric design of the experiments for the prototype evaluation.

In the rest of the paper, we outline the *datAcron* maritime researches, especially: the design of the maritime use case (Section II), the list, preparation and curation of data (Section III), necessary to the development and evaluation of the prototype, the specification of experiments (Section IV), the set-up of the maritime prototype in a realistic maritime setting (Section V), and the evaluation of the system (Section VI). The paper concludes in Section VII.

## II. MARITIME USE CASE DEFINITION

We detail below the *datAcron* maritime use cases and associated scenarios which supported the development of vessel movement analytics algorithms and the prototype setup. While “[a] use case is the story of how the business or system and the user interact”, while the “[s]cenarios tell the full story”

<sup>1</sup>*datAcron* website: <http://www.datacron-project.eu> (accessed April 2019)

<sup>2</sup>EUMSS Action Plan: [http://ec.europa.eu/maritimeaffairs/policy/maritime-security/doc/20141216-action-plan\\_en.pdf](http://ec.europa.eu/maritimeaffairs/policy/maritime-security/doc/20141216-action-plan_en.pdf) (accessed April 2019)

TABLE I  
*datAcron* MARITIME USE CASES AND SCENARIOS WITH RELEVANT MSIS (ELABORATED FROM [9], EXCERPT FROM [8])

Use case	Scenario	Description	MSIs (e.g.)
Secure fishing	Collision prevention	Protecting fishing vessels from collision with large vessels (cargos, tankers, ferries)	MSI#3 On a maritime route MSI#4 Proximity of other vessels MSI#6 Null speed MSI#12 Change of course MSI#21 Movement mobility affected MSI#23 Engaged in Fishing MSI#26 Loitering MSI#27 Dead in water/driftng
	Vessel in distress Man Over Board (SAR)	Provide early assistance to a vessel in distress	MSI#4 Proximity of other vessels MSI#6 Null speed MSI#16 AIS emission has interrupted MSI#21 Movement mobility affected MSI#23 Engaged in Fishing MSI#25 In SAR operation MSI#26 Loitering MSI#27 Dead in water/driftng
Sustainable development	Protection of ecological areas	Protect specific areas from illegal fishing activities	MSI#2 Within a given area MSI#6 Null speed MSI#13 Course not compatible with expected destination MSI#16 AIS emission has interrupted MSI#18 AIS error detection MSI#26 Loitering
	Fishing pressure	Estimate and predict fishing pressure, identify areas at risk	MSI#2 Within a given area MSI#4 Proximity of other vessels MSI#6 Null speed MSI#16 AIS emission has interrupted MSI#23 Engaged in fishing MSI#26 Loitering
Maritime security	Migration and human trafficking	Detect possible human trafficking involving fishing vessels (or the like)	MSI#13 Course not compatible with expected destination MSI#15 No AIS emission/reception MSI#16 AIS emission has interrupted MSI#17 Change of AIS static information
	Illicit activities	Detect suspicious activities involving fishing vessels	MSI#1 Close to a critical infrastructure MSI#4 Proximity of other vessels MSI#5 In stationary area MSI#7 Change of speed MSI#12 Change of course MSI#13 Course not compatible with expected destination MSI#15 No AIS emission/reception MSI#17 Change of AIS static information MSI#18 AIS error detection MSI#26 Loitering MSI#28 rendezvous

[5]. The *datAcron* maritime use case, which is described in detail in [9][8], focused on maritime surveillance for fisheries monitoring to prevent maritime accidents and preserve the marine environment. It has been developed in collaboration with operational experts to capture domain requirements, such as the need for live monitoring of fishing vessels with the surrounding traffic, while providing warning of critical events, highlight the benefit of contextually enhanced data analytics.

Fishing exploits common natural resources and needs to be regulated to safeguard fair access, sustainability and widespread profitability. Fishing safety is an important issue, as for instance, fishing vessels may fish in areas with dense traffic, like traffic lanes and waiting areas, or sometimes intentionally switch off their position reporting devices (e.g. the Automatic Identification System - AIS, the Vessel Monitoring System - VMS) while fishing, endangering themselves and the surrounding traffic.

Preserving the maritime environment from illegal fishing and ensuring fishing safety requires live processing and prediction of fishing vessel trajectories, identifying movement patterns (e.g. *close-encounters*, *change in speed* and *course*), detecting vessel activities (e.g. *fishing*, *loitering*, *tugging*), forecasting potential collisions between surrounding ships within a typical time scale of 5 to 15 minutes.

In *datAcron*, six fishing monitoring scenarios have been considered. For each scenario, the user information needs are expressed through combinations of *Maritime Situational Indicators (MSIs)*. MSIs match maritime events of interest, such as behavioural patterns and singularities in vessel traffic

data. 28 MSIs have been formalised in *datAcron* (cf. [8]). This list of MSIs is a synthesis of outcomes of workshops gathering user's elicitation and reported in the literature (e.g. [2], [17]). These MSIs have then been filtered according to their ability to be automatically detected or predicted by the *datAcron* components, processing mainly AIS data. For instance, any MSI referring to visual sighting has been excluded from this list.

Among the three use cases (Secure fishing, Sustainable development, Maritime security), the *collision prevention scenario* has been the focus of the project. The *collision prevention scenario* addresses the task of protection of fishing vessels from collision with large vessels (cargos, tankers, ferries). Using the *datAcron* system, maritime operators should be able to warn fishing vessels at risk of collision and warn large vessels heading to fishing areas. As illustrated in Table I, relevant MSIs in this scenario help the operators identify vessels potentially at risk (e.g. *vessels engaged in fishing*, *loitering* or whose which *movement mobility is affected*, with *null speed*, *drifting*, that are positioned *on a maritime route*), large vessels doing unexpected manoeuvres (e.g. *vessels that suddenly change course*) or vessels in *proximity of other vessels* which could participate in Search And Rescue (SAR) operations to rescue the crew of a *drifting* vessel. Table I describes the three use cases and associated six *datAcron* scenarios and lists meaningful samples of the associated MSIs. A complete specification is available in [8].

TABLE II  
*datAcron* MARITIME DATA (TABLE ADAPTED FROM [16], [18])

Type	Source	Originator	Format	Spatial extent	Temporal extent	Size	Avg stream rate
Surveillance	Automatic Identification System	Naval Academy	CSV	NW of France	Oct 2015-Mar 2016 (6 months)	19.680.743 msg (1.05 GB)	76 msg/min
		IMIS Global	CSV	Europe	Jan 2016	80.169.806 msg 8.5 GB	1.830 msg/min
		IMIS Global	JSON	Europe	-	-	400 KB/min 1.830 msg/min
Weather	Sea state	SHOM.fr, Ifremer	CSV	NW of France	Oct 2015-Mar 2016 (6 months)	79.652.684 obs (1463 forecasts) 3.02 GB	1 forecast/3h
	Weather forecast	UK Met Office (based on NOAA)	CSV	NW of France	Oct 2015-Mar 2016 (6 months)	71.516 obs from 16 stations 5 MB	1 obs/h
Contextual	Geographical	<i>e.g.</i> nautical charts	ESRI shapefiles	Europe	-	22 features 1.4 GB	-
	Port Registers	World Port Index SeaDataNet	ESRI shapefiles	World	-	5754 ports 70 MB	-
	Ship Registers	European Commission Agence Nationale des Fréquences	CSV	-	-	166.683 distinct ships	-

### III. DATA PREPARATION AND CURATION

The research developments of the maritime *datAcron* prototype leveraged maritime datasets that were prepared to challenge the components of the big data system in terms of *velocity*, *veracity*, *variety* and *volume*, stressing the *datAcron* prototype in realistic conditions as outlined in the fishing monitoring use case. With this aim, we performed a board inventory and detailed characterisation of available data sources that could contribute to improve Maritime Situation Awareness of the user fulfilling the task described by the use cases.

In order to increase the clarity and accuracy of the maritime picture and to overcome possible signals deception (*e.g.* positioning system manipulation and spoofing of collaborative surveillance systems like the AIS [12]), maritime operators use complementary and redundant information, including surveillance, commercial and public data sources. Besides, the analysis of streaming data from multiple sensors is essential to detect critical events as soon as they occur at sea.

We identified over forty data sources, classified within sixteen categories (*cf.* [13]). Based on this study, a representative heterogeneous maritime dataset was delivered to the project to support the maritime developments. Table II summarises the data prepared for this purpose, detailing, for each data piece, type, source, originator, file format (with CSV, Comma Separated Value; JSON, JavaScript Object Notation), spatial and temporal extent, size, and approximate average stream rate (with *msg*, messages; *obs*, observations).

The dataset included both a batch dataset, limited in volume but composed of a large variety of maritime data, and a live stream of parsed AIS messages, meant to challenge the big data system under a large volume of surveillance data with high velocity (*cf.* [16]). The AIS is an electronic system that enables ships to broadcast their dynamic (*e.g.* position, speed, heading), voyage-related (estimated time of arrival, destination) and static (*e.g.* name, type, dimensions, international identifier) information via radio communications.

Originally designed for collision avoidance, the AIS is now the main collaborative source of surveillance data. For the *datAcron* maritime use case, AIS data have been identified as the main data source for generating the maritime moving objects trajectories and their compressed representation (synopses) to be analysed by the *datAcron* event detection and forecasting algorithms. Therefore, the preparation of the *datAcron* dataset has been based on data collected from this positioning system, leveraging resources available to workpackage partners.

Although positional data are essential for *datAcron* algorithms, they are not sufficient to solve the maritime surveillance missions described in the *datAcron* maritime use case, which often require the integration of AIS with contextual information like weather and sea state conditions (*e.g.* forecast and nowcast, or observations), regulations (*e.g.* traffic separation scheme, Electronic Nautical Charts), and registries complementing vessel information (*e.g.* vessel and port registries). For all contextual data, European institutions and projects (*e.g.* SeaDataNet<sup>3</sup>, Copernicus<sup>4</sup>, EMODnet<sup>5</sup> and Ifremer<sup>6</sup> and the European Commission<sup>7</sup>) have been the preferred data providers and distributors. Considering licenses of each piece of maritime data prepared for *datAcron*, a revised public version of the batch dataset has been published online, and will remain available for further reuse [14]. Data obtained by various types of sensor technologies were processed, cleaned up from inconsistencies, transformed, harmonised and aggregated. Besides, a technical note describing which nautical charts and objects are useful for *datAcron* scenarios and the necessary scripts to process them has been shared publicly

<sup>3</sup>Pan-European Infrastructure for Ocean & Marine Data Management (SeaDataNet) data search: <https://www.seadatanet.org> (accessed April 2019)

<sup>4</sup>Copernicus Maritime Environment Monitoring Service: <http://marine.copernicus.eu> (accessed April 2019)

<sup>5</sup>European Marine Observation and Data Network (EMODnet): <http://www.emodnet.eu> (accessed April 2019)

<sup>6</sup>Institut français de recherche pour l'exploitation de la mer (Ifremer) <https://www.ifremer.fr> (accessed April 2019)

<sup>7</sup>Joint Research Centre Data Catalogue: <https://data.jrc.ec.europa.eu> (accessed April 2019)

[15].

To support the evaluation of the *datAcron* algorithms, the development dataset has been enriched with contextual historical information. Raw data have been processed to extract maritime patterns and clusters, specifically maritime routes, which facilitate the operators assessment. Additionally, a library of modification functions has been designed and implemented to systematically degrade and enrich the AIS batch dataset. The modifications applicable to each AIS field have been categorised along *kinematic*, *coverage* and *spoofing* dimensions. This library of functions provides a rich set of basic constructs to build different modification patterns in a non unique manner [6]. Having different ways to produce equivalent patterns or to provide a wider diversity in similar patterns is desirable to create realistic synthetic datasets. This library of functions has been used to prepare experiment datasets and to create reference information, useful for the evaluation of the *datAcron* prototype.

#### IV. EXPERIMENTS SPECIFICATION

The experiment methodology aimed at assessing the ability of the *datAcron* prototype, setup within the maritime use case, to efficiently help the operator accomplishing the specific surveillance task. The *datAcron* algorithms are generic in order to be applied to different domains and scenarios specifically, to the air and maritime domains. Equivalently, the methodology for experimenting, evaluating and validating the *datAcron* prototype and components had to ensure the replicability of the results, and generic enough to be applicable to the range of use cases previously described.

With these objectives, we designed and proposed an evaluation methodology and an initial maritime experimental plan (cf. [3]). To facilitate the collaborative development and ensure the efficient integration of the *datAcron* prototype towards meeting the operational objectives, the experimental plan was developed along the following centrepieces, as illustrated in Figure 1:

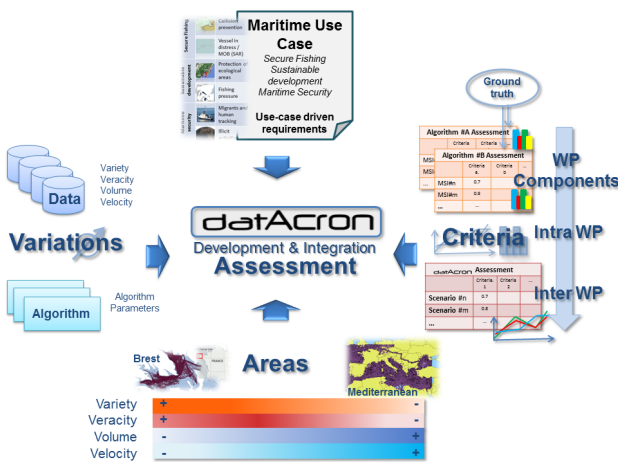


Fig. 1. Centrepieces of the maritime *datAcron* prototype assessment

- i) *Scenario-driven evaluation*, ensuring that the maritime domain benefits from the *datAcron* outcomes;
- ii) *Interconnected stages for collaborative development, integration and assessment*, to facilitate inter-team cooperation;
- iii) *Functional levels of assessment* (technical, functional and human-driven), enabling an incremental integration of the application context;
- iv) *Data-driven evaluation*, to address different big data dimensions foci.

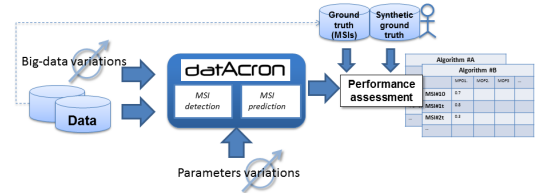


Fig. 2. Assessment at *MSI-level* (elaborated from [9])

MSIs played a pivotal role in the evaluation, expressing information needs in a maritime surveillance task and acting as a bridge between the operator and the system, hence worked as an interface between the *datAcron* teams involved in the development of the different system components. By design, they represent atomic functional units against which the maritime prototype is designed, hence evaluated. As such, the three functional levels of assessment correspond to either *under-MSI* functionalities, which include for instance data management operations (e.g. the search of and access to stored maritime data, the creation of vessel trajectory synopses, and the generation of vessel routes), or directly to *MSIs* detection and forecasting (cf. Figure 2). At the last level of assessment, namely *scenario-level*, the system is used by the operator to accomplish a mission as described in one of the scenarios, through the detection or prediction of the *MSIs* helping him/her complete the task (cf. Figure 3).

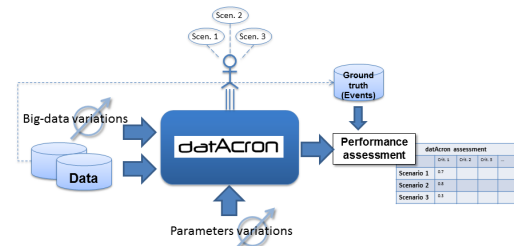


Fig. 3. Assessment at *scenario-level* (elaborated from [9])

The method was refined along the work, to be applicable to any maritime use case and scenario. At the scenario level, as exemplified in Figure 3, it includes the following steps:

- i) Define the operational scenario for evaluation;
- ii) Identify the relevant scenario's *MSIs* and the *datAcron* components responsible for *MSIs*' detection or forecasting;

- iii) Specify the detailed experimental plan in terms of: research questions, components to test, big data variations to apply, assessment criteria and measures to use for evaluation;
- iv) Design and prepare the data to support the scenario and the data variations specified in the experiment (e.g. create a dataset containing heterogeneous data, either with ground truth events or synthetic reference data, then degrade, annotate, or enrich the scenario dataset to challenge the *datAcron* algorithms under the big data variations specified in the experimental plan);
- v) Design and implement an architectural schema to describe the *datAcron* components and their interconnections, and their role in the scenario's workflow;
- vi) Execute the experiments, collect the results in a unified schema, then analyse them to produce the evaluation.

The approach proposed allowed a comprehensive and structured coverage of the four big data challenges, the evaluation criteria (both computational and involving humans), as well as the traceability and repeatability of the results. The methodology developed during the *datAcron* project can be further reused and improved to support evaluation activities of future collaborative research projects.

## V. MARITIME *datAcron* PROTOTYPE SET-UP

The *datAcron* maritime prototype set-up was designed to support the scenario-driven evaluation, then developed in two steps: An initial set-up together with some experiment rehearsal, and a final set-up concluding with a week of experiments at the end of the project. The first set-up was tested on a partially developed prototype. The first prototype set-up was designed and tested with three maritime experts in an experimental setting focused on a collision avoidance scenario. Based on the experts' feedback, a second prototype set-up was organised few months later. As outcome, the experimental laboratory sketched in Figure 4 was delivered to challenge the *datAcron* analytics and visualisation with live data (real data complemented with pseudo-synthetic events [6]), simulating a near-operational environment (cf. [7]). This objective differs from a simple demonstration where results are just shown to users, often in a favourable situation.

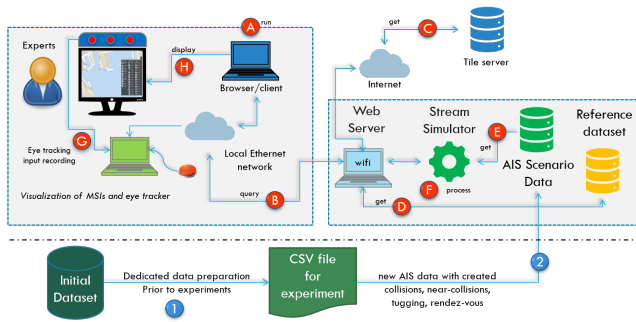


Fig. 4. Architecture of the maritime *datAcron* prototype set-up

The final set of experiments focused on a collision scenario and involved three maritime experts, two cadets from the French Navy and a maritime expert who also assessed the quality of MSIs detection along the project. For each experiment, the final prototype set-up involved the following steps:

- i) Choice of the area of interest (realistically matching a maritime area under monitoring by Navy and Coast Guard operators);
- ii) Specification of the methods to use for generating the data to populate the scenario;
- iii) Creation of the scenario;
- iv) Data annotation (i.e. manually or automatically label the surveillance stream with MSIs).

The final prototype set-up and experiment executions corresponded to the workflow schematically represented in Figure 5. As illustrated in the figure, given an initial dataset created as a selection (i.e. projection in time and space) of the whole dataset, events were injected in the surveillance stream, producing a realistic story to match the maritime surveillance challenges of the collision avoidance scenario. The resulting data stream consisted of real surveillance tracks enriched with maritime events including:

- collisions between real vessel (trajectory) and a synthetic vessel (trajectory);
- collisions between real vessel (trajectory) and a real vessel, whose trajectory is shifted in time and space;
- synthetic near-collisions;
- shift in time of real tugging cases;
- shift in time and space of existing trajectories in order to simulate specific behaviours (e.g. fishing patterns);
- simulation of rendez-vous behaviours.

The resulting stream was then annotated with MSIs, either manually by domain experts or automatically, i.e. by the detection and forecasting *datAcron* components [10], [11]. This annotated dataset was converted into a CSV file, and was also streamed in Apache Kafka<sup>8</sup> to be made available to the *datAcron* visualisation interface for the use of the experiments. This workflow is repeatable for any of the scenarios defined in the *datAcron* maritime use case.

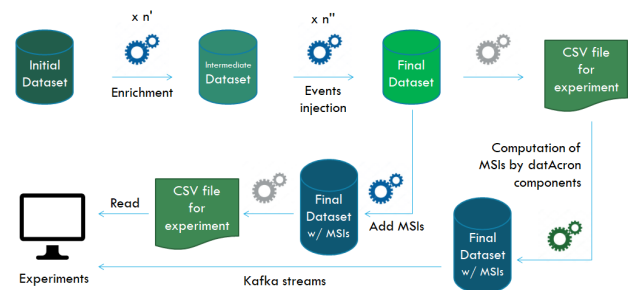


Fig. 5. Scenario definition and displaying

<sup>8</sup>Apache Kafka, A distributed streaming platform: <https://kafka.apache.org> (accessed April 2019)

The prototype set-up required suitable icons for displaying MSIs to the expert users. Icons were fundamental to help the users familiarise with the computational aspects of the *datAcron* system and to enhance their situational awareness during the experiments. Although the typical icons used in the domain of maritime surveillance are familiar to the experts, they do not cover the MSIs identified for the *datAcron* projects, especially the complex events. Consequently, the set of standard icons has been complemented according with new designs. It appeared however that these new icons would require some training to guarantee a proper interpretation by the users.

## VI. SYSTEM EVALUATION

The prototype evaluation, which is described in detail in [21], [20], followed the execution of the maritime experimental plan as described in Section IV. This experimental plan underwent refinements and improvements until the execution of the final experimental evaluation, in order to include new experiments and to incorporate the latest *datAcron* developments. It was executed with the maritime prototype set-up as for Section V, processing the data prepared as described in Section III.

To facilitate the evaluation and the analysis of the experiment results, we adopted a unified evaluation framework consisting of an evaluation space defined along three dimensions:

- i) Big data dimensions (*i.e.* which among volume, velocity, variety and/or veracity is made varying, and how much);
- ii) Evaluation criteria and specific measures as output of the experiment;
- iii) *datAcron* components involved in the experiment execution.

The space was then partitioned for reducing the number of the necessary experiments, according to a semantic-based decomposition and recombination approach that allows for the aggregation of component evaluation results at inter-component level. A design of experiments ontology was defined for supporting the selection of experiments, which exploits domain and inter-domain specific restrictions on the factor combinations in order to select a representative subset from the very large number of possible experiments (*cf.* [19], [1] for details).

As illustrated in Figure 6, the evaluation was defined according to a human-centred design, with maritime domain experts involved at two different levels:

- i) at *MSI-level*, an expert evaluated *off-line* the accuracy of the MSIs detected, *i.e.* whether the MSIs detected or forecast by *datAcron* components matched his expectation;
- ii) at *scenario-level*, we measured the capability of the *datAcron* prototype to help the experts solve the maritime surveillance task at hand.

The expert-driven evaluation of MSIs accuracy was also carried out along the project, assisting the partners involved in the development of the detection and forecasting *datAcron* components. This specific evaluation was conducted simultaneously with the computation-driven assessment performed

by *datAcron* partners at the component (*under-MSI*) level (*cf.* Section IV).

The expert involved in the MSIs' accuracy evaluation checked the component results and evaluated them according to his expertise and knowledge of the area covered by the data, compensating the lack of ground truth on real maritime events, as labelled datasets are challenging to obtain in extensive and realistic use cases. Whenever the number of events was too large for a complete manual assessment, a representative sample of events was evaluated instead. As a result, a series of qualitative confusion matrices was produced, measuring the quality of the evaluated algorithms based on the events that were correctly detected (*i.e. true positives*), miss-detected (*i.e. false positive*), and in some cases correctly undetected (*i.e. true negative*), as illustrated in Figure 7. The detailed approach followed by the expert is described in [21], [20].

The final user-centric scenario-driven evaluation encompassed a set of experiments assessing the *datAcron* system in an operational context. The experiments evaluated the capability of the system to improve the situational awareness of maritime security operators involved in a surveillance mission, *i.e.* to help identify and prevent collisions involving fishing vessels as described in the *datAcron* collision avoidance scenario. Expert users, including maritime experts and cadets from different countries, participated to the evaluation.

The first experiment used a gaming approach to assess the user perceived relevance and usefulness of MSIs, decoupling the information content of information layers (context, MSIs, raw AIS) from the outcome of the *datAcron* components and from the visualisation tool (see details in [21]). Another series of experiments was conducted live, with the user in front of a screen showing the scenarios and the data prepared and processed for the experiments by the *datAcron* components. Live experiments (*cf.* Figure 8) assessed the impact of MSIs on the operators detection of near-distance events (*i.e. tugging, collision, near-collision, rendezvous*), both *per se* and taking full advantage of the *datAcron* functionalities, *i.e.* the interactive visual analytics and the detection and forecasting capabilities. All the experiments are reported in [21] and described in detail and analysed in [21], [20].

## VII. CONCLUSION

The *datAcron* maritime use case was designed to support *datAcron* innovations demonstrating their applicability in a realistic maritime surveillance operational scenario. In this paper, we have described its design, together with the activities undertaken towards this objective within the *datAcron* Maritime workpackage, including the data inventory, preparation and curation, the maritime prototype set-up and the design and execution of the experimental evaluation. Six scenarios corresponding to three use cases (two each) of Secure fishing, Sustainable development and Maritime security have been proposed, while the collision avoidance scenario under the Secure fishing use case was retained as the main one. A set of 28 Maritime Situational Indicators as been proposed to both capture user information needs and provide targets to event

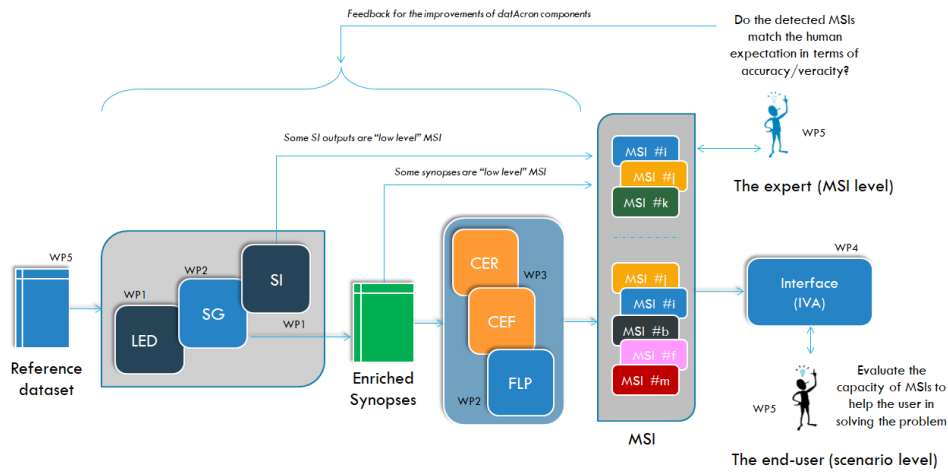


Fig. 6. Workflow of human-centered expert-driven evaluation in *datAcron*: maritime experts are involved in the accuracy evaluation of MSIs (top-right), and in the final evaluation at scenario-level (bottom-right) [21]

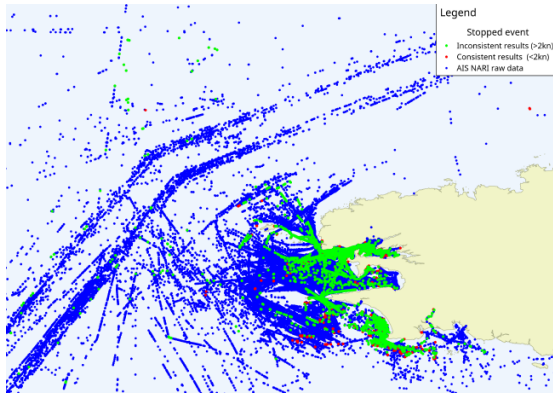


Fig. 7. Evaluation of accuracy of MSIs (example, *Stop* events)

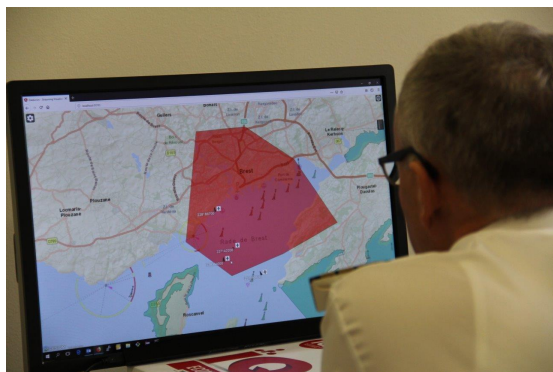


Fig. 8. Scenario-driven evaluation: live experiment with expert evaluating the *datAcron* prototype

detectors to be designed by the *datAcron* and integrated in the different components. The scenario was described with relevant MSIs, for which big data analytics solutions were proposed and implemented. The algorithms thus designed have been exercised on an heterogeneous dataset gathering timely AIS data and other contextual data. The open heterogeneous

maritime dataset has been made available to the public to enable other further maritime experimentation in realistic operational settings, challenging specifically the *variety* dimension of big data. The maritime prototype has been set-up following a human-centric approach involving maritime expert and tying together the scenario definition, data preparation, visualisation and human evaluation. Finally, the evaluation of the maritime prototype has been conducted according to a methodology summarised in this paper and further described in [21]. Future work will address the deeper formalisation of the assessment process, using this *datAcron* experimental plan as a blueprint to be applied to the evaluation of other big data systems.

## REFERENCES

- [1] M. Zocholl and E. Camossi and A.-L. Joussemme. Test case development for big data solution evaluation. In *Maritime Big Data Workshop, CMRE, La Spezia, Italy, 9 May 2018 (MBDW), La Spezia, Italy, 9-0 May 2018 (Session MARITIME BIG DATA IN DATACRON)*, May 2018.
- [2] S. Andler, M. Fredin, F. Gustafsson, J. va Laere, M. Nilsson, and P. Svenson. SMARTracIn - a concept for spoof resistant tracking of vessel and detection of adverse intentions. In *SPIE Defense, Security, and Sensing*, Orlando, FL, 2009.
- [3] E. Camossi, A.-L. Joussemme, C. Ray, M. Hadzagic, R. Dréo, and C. Claramunt. Maritime experiments specification, H2020 datacron d5.3, 2017.
- [4] C. Claramunt, C. Ray, E. Camossi, A. Joussemme, M. Hadzagic, G. L. Andrienko, N. V. Andrienko, Y. Theodoridis, G. A. Vouros, and L. Salmon. Maritime data integration and analysis: recent progress and research challenges. In V. Markl, S. Orlando, B. Mitschang, P. Andritsos, K. Sattler, and S. Breß, editors, *Proceedings of the 20th International Conference on Extending Database Technology, EDBT 2017, Venice, Italy, March 21-24, 2017.*, pages 192–197. OpenProceedings.org, 2017.
- [5] J. Heumann. Tips for writing good use cases. *IBM - Transforming software and systems delivery*, May 2008. White paper.
- [6] C. Iphar, A.-L. Joussemme, and C. Ray. Pseudo-synthetic datasets in support to maritime surveillance algorithms assessment. In *VERITA workshop, Metz, France, 22 January*, 2019.
- [7] C. Iphar, C. Ray, M. Zocholl, A.-L. Joussemme, R. Dréo, and E. Camossi. Maritime datacron prototype set up, H2020 datacron d5.5, 2018.
- [8] A.-L. Joussemme, C. Ray, E. Camossi, M. Hadzagic, C. Claramunt, K. Bryan, E. Reardon, and M. Ilteris. Maritime use case description, H2020 datacron d5.1, 2016.



- [9] A.-L. Joussemle, E. Reardon, M. Hadzagic, E. Camossi, K. Bryan, M. Iteris, C. Ray, and C. Claramunt. A maritime use case in support to collaborative research. In M. Vespe and F. Mazzarella, editors, *Maritime Knowledge Discovery and Anomaly Detection Workshop*, Ispra, Varese, July 2016. European Commission Joint Research Centre, European Commission Publication Office.
- [10] K. Patroumpas, E. Alevizos, A. Artikis, M. Vodas, N. Pelekis, and Y. Theodoridis. Online event recognition from moving vessel trajectories. *Geoinformatica*, 21(2):389–427, 2017.
- [11] M. Pitsikalis, I. Kontopoulos, A. Artikis, E. Alevizos, P. Delaunay, J.-E. Pouessel, R. Dreó, C. Ray, E. Camossi, A.-L. Joussemle, and M. Hadzagic. Composite event patterns for maritime monitoring. In *Proceedings of the 2018 Hellenic Conference on Artificial Intelligence (SETN'18), Rio Patras, Greece*, pages 192–197, July 2018.
- [12] C. Ray. Data variety and integrity assessment for maritime anomaly detection. In *1st International Conference on Big Data and Cyber-Security Intelligence (BDCSIntell 2018), Hadath, Lebanon, December 13-15, 2018*, volume 2343 of *CEUR Workshop Proceedings*, 2018.
- [13] C. Ray, E. Camossi, A.-L. Joussemle, M. Hadzagic, C. Claramunt, and E. Batty. Maritime data preparation and curation, H2020 datacron d5.2, 2016.
- [14] C. Ray, R. Dréo, E. Camossi, and A.-L. Joussemle. Heterogeneous integrated dataset for maritime intelligence, surveillance, and reconnaissance (version 0.1) [data set], 2018.
- [15] C. Ray and A. Grancher. *Integration of Nautical Charts in Maritime Dataset 10.5281/zenodo.1167595*, Feb. 2018.
- [16] C. Ray, C. Iphar, R. Dréo, W. Kleynhans, E. Camossi, A.-L. Joussemle, M. Zocholl, E. Batty, Q. Roche, and A. Metzger. Maritime data preparation and curation (final), H2020 datacron d5.4, 2018.
- [17] J. Roy and M. Davenport. Exploitation of maritime domain ontologies for anomaly detection and threat analysis. In *Proc. of the IEEE int. Waterside Security Conference (WSS)*, 2010.
- [18] G. A. Vouros, A. Vlachou, G. M. Santipantakis, C. Doulkeridis, N. Pelekis, H. V. Georgiou, Y. Theodoridis, K. Patroumpas, E. Alevizos, A. Artikis, C. Claramunt, C. Ray, D. Scarlatti, G. Fuchs, G. L. Andrienko, N. V. Andrienko, M. Mock, E. Camossi, A. Joussemle, and J. M. C. Garcia. Big data analytics for time critical mobility forecasting: Recent progress and research challenges. In M. H. Böhlen, R. Pichler, N. May, E. Rahm, S. Wu, and K. Hose, editors, *Proceedings of the 21th International Conference on Extending Database Technology, EDBT 2018, Vienna, Austria, March 26-29, 2018.*, pages 612–623. OpenProceedings.org, 2018.
- [19] M. Zocholl, E. Camossi, A. Joussemle, and C. Ray. Ontology-based Design of Experiments on Big Data Solutions. In A. Khalili and M. Koutraki, editors, *Proceedings of the Posters and Demos Track of the 14th International Conference on Semantic Systems (SEMANTiCS 2018), Vienna, Austria, September 10-13, 2018.*, volume 2198 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2018.
- [20] M. Zocholl, R. Dréo, C. Ray, A.-L. Joussemle, C. Iphar, F. de Rosa, E. Camossi, G. Keraudren, and F. Rozé. Maritime final validation, H2020 datacron d5.6, 2018.
- [21] M. Zocholl, C. Iphar, E. Camossi, F. de Rosa, A.-L. Joussemle, C. Ray, and R. Dréo. User-centric assessment of maritime situation awareness solutions. In *OCEANS 2019, Marseille, France*, 2019.