



HAL
open science

User-centric assessment of maritime situation awareness solutions

Maximilian Zocholl, Clément Iphar, Richard Dréo, Francesca de Rosa,
Anne-Laure Jouselme, Cyril Ray, Elena Camossi

► To cite this version:

Maximilian Zocholl, Clément Iphar, Richard Dréo, Francesca de Rosa, Anne-Laure Jouselme, et al.. User-centric assessment of maritime situation awareness solutions. OCEANS 2019 - Marseille, Jun 2019, Marseille, France. pp.1-9, 10.1109/OCEANSE.2019.8867518 . hal-03853385

HAL Id: hal-03853385

<https://hal.science/hal-03853385>

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.

User-centric assessment of maritime situation awareness solutions

Maximilian Zocholl
NATO STO Centre for
Maritime Research and
Experimentation
La Spezia, Italy

maximilian.zocholl@cmre.nato.int

Clément Iphar
NATO STO Centre for
Maritime Research and
Experimentation
La Spezia, Italy

clement.iphar@cmre.nato.int

Richard Dréo
NATO STO Centre for
Maritime Research and
Experimentation
La Spezia, Italy

richard.dreo@cmre.nato.int

Francesca de Rosa
NATO STO Centre for
Maritime Research and
Experimentation
La Spezia, Italy

francesca.deRosa@cmre.nato.int

Anne-Laure Joussemme
NATO STO Centre for
Maritime Research and
Experimentation
La Spezia, Italy

anne-laure.joussemme@cmre.nato.int

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

Elena Camossi
NATO STO Centre for
Maritime Research and
Experimentation
La Spezia, Italy
elena.camossi@cmre.nato.int

Abstract—In this paper, we report on the evaluation methodology developed to support the experimental plan to assess the big data prototype developed under the umbrella of the *datAcron* project. The *datAcron* prototype was decomposed along several semantic levels of functionality: Maritime data, *datAcron* components, Maritime Situational Indicators (MSIs), and Maritime scenarios. Different experimental features have been selected for delivering representative results at all levels of system integration. The results were captured in an evaluation framework unifying the different evaluation criteria and measures into the following groups: (1) timeliness and scalability, (2) compression ratio, (3) classification quality and (4) clarity and effectiveness. By the three dimensions of *system decomposition*, *big data variations* and *evaluation criteria* an experimental space is opened, with respect to which all experiments are referenced. Major findings on the performance of the components of *datAcron* prototype are reported, together with other findings regarding the user assessment of the effectiveness and the clarity of the *datAcron* prototype.

Index Terms—Maritime big data, collision prevention, fishing use case, maritime use case, Maritime Situational Awareness

I. INTRODUCTION

The project *Big Data Analytics for Time Critical Mobility Forecasting (datAcron)* was funded by the European Union Horizon 2020 programme and run from 2016 to 2018. *datAcron* aimed at developing novel methods to detect threats and abnormal activities in streams of large numbers of moving objects in wide maritime and aerial areas. It involved eight (8) partners who jointly developed algorithms to address the big data challenges in both aerial and maritime environment, along the different topics of mobility pattern detection and forecasting [1], complex event recognition and forecasting [2],

This work was supported by project *datAcron*, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 687591.

visual analytics [3], system prototype and data management [4], as well as a maritime and aviation use case based validation of the components developed in the different work packages [5], [6].

The user-centric assessment of maritime situation awareness solutions, presented in this paper, is performed on the example of the big data prototype implemented for *datAcron* project. Other examples of analytic applications of AIS data can be found in [7] or from a big data perspective in [8] while an interesting example for the use of AIS data for evaluation purposes is proposed in [9].

The article is structured as follows: Firstly in Section II, a description of the *datAcron* prototype is provided in the evaluation scope. Secondly, the methods of experimentation are presented in Section III followed by the main results in Section IV and the method evaluation in Section V. Finally, a discussion and a conclusion complete the contribution.

II. EVALUATION SCOPE

The evaluation scope includes the different components of the *datAcron* system, shown in Figure 1 from a maritime perspective. These are the *Low Level Event Detection (LED)* for in-situ processing of data from different sources and the detection of spatio-temporal events with low complexity, the *Synopses Generator (SG)* for the compression of trajectory data with minimal loss of semantics, the *Semantic Integrator (SI)* for the linking of data from different sources in RDF, the *Complex Event Recognition (CER)* for the detection of events on single or between multiple vessels that build up on the results of the priorly described components, the *Complex Event Forecasting (CEF)* for the use with incomplete data, the *Future Location Predictor (FLP)* for the use with complete data and the *Interactive Visual Analytics (IVA)* for making the analytical

results explorable by domain experts and data scientists. Each component or a combination of multiple components calculate the so called Maritime Situational Indicators (MSI) around which the assessment is pivoting. These MSIs correspond to “anomalies” and “behaviours or events of interest” in which one or more vessels are involved and which are to be detected, predicted or forecast [10].

Besides of the components that are evaluated, Figure 1 depicts the two building blocks of the presented work. Firstly, the workflow between the components of datAcron system and the processed datasets is shown. Starting from a specifically prepared reference dataset based on Automatic Identification System (AIS) data, LED, SG and SI output a first set of MSIs and other enriched synopses which compress the reference dataset. These MSIs serve as input to CER, CEF and FLP components for the calculation of composed and complex MSIs. Secondly, Figure 1 depicts the relationship between the performed experiments on the MSI level and on the scenario level. Depending on these levels of experiments, the different maritime domain expert-users occupy different roles. To make them easily distinguishable in the following the domain experts are referred to as expert on the MSI level and as operator on the scenario level. On the MSI level, the expert is evaluating the computed MSIs by taking into account all data available and giving feedback for the improvement of *datAcron* components. On the scenario level, an operator is part of the experiment and enables the assessment of the complete *datAcron* system in simulated “operational context”.

III. METHODS OF EXPERIMENTATION

We followed a human-centric approach with the Maritime Situational Indicators (MSIs) playing a pivotal role; the evaluation space is defined to capture the evaluation results in a unified framework and partition the space for reducing the number of necessary experiments; two type of assessments have been designed, either purely computational without involvement of maritime surveillance experts and independent assessments of the components involving experts. Practically, Three complementary experiments have been performed: One on the MSI level and two on the scenario level, as detailed in [5].

On the MSI level, the accuracy of the MSIs detected by the different components was assessed by a domain expert, through their true or false positive or negative.

On the scenario level the *datAcron* system is tested simulating an operational context on the exemplary mission of collision avoidance, processing a specially prepared dataset. For decoupling the impact of MSI accuracy and the visualisation quality on the evaluated performance, two experiments are performed.

The first scenario-level experiment aims at measuring the impact of MSIs on the operators’ awareness of collision situations along the dimensions described in [11]. The performance of different experts in predicting and detecting collisions and near-distance situations with and without MSIs are compared¹.

¹In both cases AIS data and contextual data was available

Here, the MSIs are the outcome of a manual labelling process.

The setup of the second scenario-level experiment resembles the first scenario-level experimental setup with the difference that the MSIs are computed by the datAcron prototype. Here, the performance of different experts relying on calculated MSIs displayed on the screen is compared to their performance on manually labelled MSIs. In the following the three experiments are described in detail.

A. MSI-level experiment

The expert is given the reference AIS dataset in which MSIs detections are included, especially:

- 1) *Stop*, where a vessel stops (MSI #6);
- 2) *Underway*, which describes a vessel which is moving, or sailing (MSI #19);
- 3) *High speed*, describing a vessel which speed is above the cruise speed, i.e. above a threshold in a given area (MSI #7-11).

These events are priorly detected by either the *SG*, *CER* components or both. The experts’ task is to evaluate the accuracy of the outputs of these components. To accomplish the task, the expert has at his disposal the raw AIS dataset, plus many geographical features (e.g. electronic nautical charts), environmental data (e.g., sea state) and contextual information (e.g., vessel register). Furthermore, the expert knows the maritime region and the typical ships behaviours in the given area.

The expert used an ad-hoc combination of different software to support his analysis²:

- A database enhanced with spatial capabilities enabling basic analytics and spatial operations (e.g., the object-relational database PostgreSQL, including its spatial extension PostGIS);
- A Geographical Information System (GIS) for data visualisation and spatial analysis and filtering (e.g., the desktop GIS QGIS);
- Scripting languages for data analysis (e.g., Python and Matlab).

In the approach used by the expert, the support for data visualisation, in this case provided by the GIS software, is fundamental. The expert performed the following steps:

- 1) Import raw surveillance data (AIS), processed data (AIS data annotated with events), and supporting datasets, including AIS status codes for vessel types, vessel list, fishing vessel list, protected areas datasets, port database, weather conditions in the database;
- 2) Develop scripts to elaborate surveillance raw data and produce a *baseline* dataset, for comparison, and import it in the database;
- 3) Convert the data in the database in order to enable the spatial representation of the spatial features (AIS coordinates, vessel trajectories);

²We report this software here as an exemplification of the approach rather than as an expert recommendation.

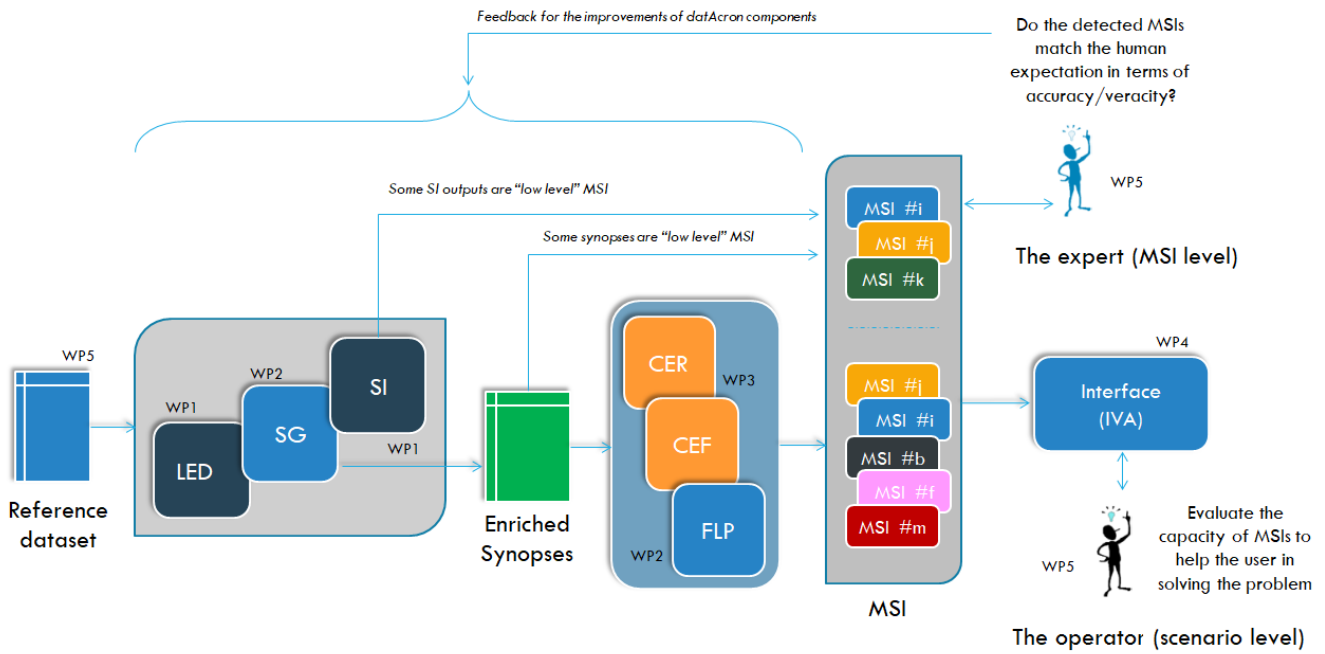


Figure 1: Maritime data workflow and user-driven assessment principle

- 4) Use the integrated database spatial capabilities to query relevant subsamples of the baseline and processed datasets;
- 5) Visualise the baseline and processed datasets in the GIS software. Use GIS analysis capabilities (filtering, spatial overlay) to compare the two datasets and highlight inconsistencies;
- 6) Eventually, develop ad-hoc scripting to calculate statistics on relevant features of the two datasets (e.g., speed), to be analysed and compared.

B. Scenario-level experiments

On the scenario level two experiments are prepared, one with manually labelled MSIs and one with calculated MSIs. Both experiments are performed in a live fashion with an operator in front of a screen and the following commonalities.

For testing the research hypotheses, 3 scenarios are displayed to different operators. Each corresponding dataset includes two near-distance situations, *i.e.* collision, near-collision or rendezvous each between two vessels plus additional vessels. The task of the operator is to avoid collisions if possible or to detect collisions after they took place. For this, the operator is asked to think aloud and to describe the visualised situations. All statements are recorded on paper by a data collector with the point in time and a confidence value. After a short structured brief explaining the task, the interaction with the operator is reduced to a minimum which is asking confidence values for the stated situational assessments.

Figure 2 proposes a taxonomy for near-distance situations. While rendezvous is an intrinsically intentional manoeuvre of two or more vessels, close quarter situations are unintended, dangerous and can result either in a near-collision/near-miss

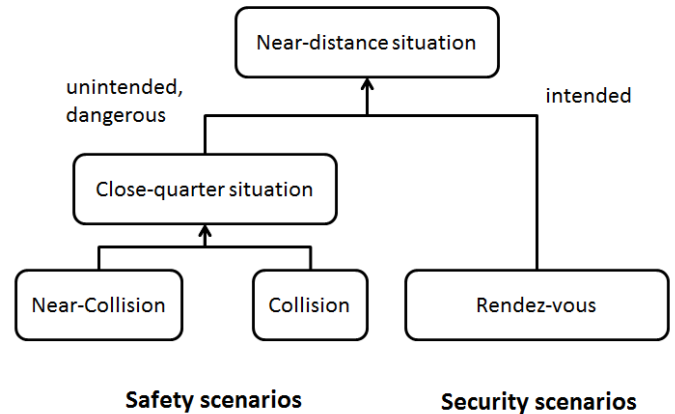


Figure 2: Near-distance situation taxonomy

or in an actual collision. For a functional representation of close-quarter situations see [12].

1) *Experimental Design:* Three scenario datasets are prepared each including two near-distance situations between two vessels plus additional vessels. Scenario 1 includes one collision and one rendezvous. Scenario 2 includes one collision and one near-collision, and scenario 3 includes one near-collision and one rendezvous.

The situations are described by AIS data and MSIs as specified in the following:

- Scenario 1 without MSI: Collision1 vs. rendezvous1
- Scenario 2 with MSI: Near-collision1 vs. rendezvous2

- Scenario 3 with MSI: Collision2 vs. near-collision2

For evaluating the effect of MSIs, the scenarios with and without MSIs are compared. For evaluating the effect of datAcron generated MSIs, scenario 2 is shown to the operator once with manually labelled MSIs and once with MSIs calculated by datAcron components. For evaluating the impact of the dataset creation method, the two near-collisions are compared. The near-collision is chosen as it is supposed to lie between Collision and rendezvous.

Each situation, e.g. Collision1 occurs once, so that the operator is shown different events in order to avoid learning effects. Each type of event occurs twice, so that conclusions are drawn both on the level of a specific situation and on the level of the type of situation.

2) *Results interpretation*: The qualitative interpretation of verbal comments by the operators is performed in two steps. Firstly, in the classification step, the their verbal description of the situation is mapped to one of the situation types of near-distance situations depicted in Figure 2, to ensure the comparability of the situation descriptions between different experts. Secondly, in the comparison step, the generalised situation types are compared to the situation type of the reference dataset. This step is performed both for the prediction phase and for the detection phase.

In the comparison step the the situation type assessed by the operator during the classification step is compared to the reference dataset and rated as true positive, true negative, false positive or false negative according to the following rules:

- True positive (TP): The description of the situation given by the operator coincides with or includes the situation type of the reference data with respect to the near-distance taxonomy in Table 2. For the inclusion, a distinction is made between prediction and detection. E.g. if a close-quarter situation is predicted by the operator (CQS_{op}) and the reference data includes a situation labelled as collision (C_{lab}), the prediction is rated as a true positive. Other events are near-collision (NC), near-distance situation (NDS), rendezvous (RDV) and no-collision ($NOC = RDV \cap NC$):
 - Prediction: $CQS_{op} \cap C_{lab} \rightarrow TP$, $CQS_{op} \cap NC_{lab} \rightarrow TP$, $NDS_{op} \cap RDV_{lab} \rightarrow TP$.
 - Detection: $NOC_{op} \cap NC_{lab} \rightarrow TP$, $NOC_{op} \cap RDV_{lab} \rightarrow TP$.
- True negative (TN): No negative statements are asked the operator, thus no true negatives are recorded. If the absence of positive statement is interpreted as a negative statement, a list of true negative detections can be computed for all pairs of vessels included in the respective dataset.
- False positive (FP): The description of the situation made by the operator is very specific, *i.e.* collision, near-collision, rendezvous or other events that are not included in the reference data, e.g. tugging, and does not coincide with the situation type of the reference data with respect

to the near-distance taxonomy in Figure 2, e.g. a tugging is described where no tugging takes place.

- False negative (FN): A situation in the reference data is not described as such, e.g. a collision is not described at all.

3) *Criteria and Measures*: The following criteria and associates measures were used:

- Timeliness: Time between cognition of a possible collision situation and hailing the vessel and the actual time of collision.
- Accuracy: number of TP, TN, FP, FN verdicts of the operator with respect to the situation and computed on both MSI and scenario level.
- Clarity: Self-confidence in the situational assessment between low (1) and high (5).

C. Scenario-level assessment with manually annotated MSIs

In the first scenario-level experiment situations with and without MSIs are compared. The MSIs are the result of a manual labelling process of domain experts and are assumed to be correct in the following.

In order to assess only the impact of the MSIs information on the operator situation awareness, regardless the detection quality of the *datAcron* components, the dataset has been labelled by 2 experts, ensuring an MSI detection quality that is assumed to be “perfect”. By this, the dataset is thus enriched by true MSI detections. The labelling process is based on the thresholds depicted in Table I. The annotation was performed for each AIS message and for all MSIs of interest. The MSIs are chosen for the labelling process in such a way that each component is represented by at least one MSI and each of these MSIs is detected by the *datAcron* components for at least one of the three scenarios. This selection process is of importance, as the results of this first scenario-level experiment represent the reference for the second scenario-level experiment.

Table I: Thresholds for MSI manual labelling

#	MSI	Threshold
2	Within a given area (TSS)	-
3	On a maritime route	-
4	Proximity to other vessels	100m
6	Null speed	< 0.5 knts
7	Change of speed	>25%
12	Change of course	4 degrees
15	No AIS reception	-
16	AIS reception interrupted	>1800s
19	Under way	-
23	Engaged in fishing	-
28	Rendezvous	-

D. Scenario-level experiment with calculated MSIs

For the second scenario experiment situations with manually labeled MSIs are compared to situations with MSIs detected by *datAcron* components. Therefore, the reference dataset is processed by the *datAcron* components and the resulting detections are displayed to the operator.

IV. RESULTS

The results are presented firstly for the MSI-level assessment, secondly for the assessment of the impact of MSIs on the situation awareness of experts and thirdly for the assessment of the impact of MSIs calculated by *datAcron* components.

A. Results of MSI-level assessment

In this experiment, the objective was to reproduce a “live” assessment at the MSI level. To this end, the expert gathered all the results produced the *datAcron* components during the other experiments. The expert then applied the methodology described in Section III-A. The assessment focused on the MSIs #2, #6, #7, #12, #19, #28 involved in the scenario and computed by *SG*, *SI*, *CER*. Detections of *datAcron* components were compared with raw, enriched and manually annotated data. Results of this comparison are reported in Table 3 and commented below.

MSI #2 (Within an area)

“The result is good, all events are detected. It appears that *SI* component limited the computation of *within an area* of the TSS to the polygons corresponding to the separation zones. However from a maritime situation perspective, I was expecting a kind of bounding box around these polygons in order to include the channels of the traffic separation scheme”.

MSI #6 (Null speed)

“While the *SG* component gives some rather consistent results, we see that the *CER* does not return the same thing. The big difference I think is the fact that the *CER* has rejected all the detections that are in the port of Brest. As for the rest, none of the *CER* detections correspond to the annotations. But they are not so inconsistent. The thresholds chosen are probably not unrelated to this difference. Note that after the collision, the *CER* does not see the boats stopped, unlike the annotators”.

MSI #7 / #12 (Change of speed/course)

“We can notice a big difference between the detected events and the annotations, particularly for the change speed event. Indeed, the change speed mainly corresponds to false alarms. We don’t have any explanation for this difference”.

MSI #19 (Underway)

“We notice that the detection of the underway event is done with a very few false alarms. But compared to the annotated dataset, the detection probability is only 20%”.

MSI #28 (RDV)

“The choice was made by the *CER* developers to consider the “rendezvous” as a punctual event, contrary to the experts who annotated more data in case of proximity between the involved vessels. The conclusion is that the two rendezvous situations in the annotated dataset were correctly detected,

while the apparent missed detection is very high”.

B. Scenario experiment on the impact of MSIs

1) Effect of MSIs on prediction and detection:

Table II summarises the True Positive, False Positive and False Negative Rates for collisions, near-collisions and rendezvous. The result indicates both a positive effect of MSIs on the task fulfilment of collision prediction and detection and on near-distance prediction and detection where both the true positive rate increases and the false positive and negative rates decrease.

2) Effect of MSIs on the prediction time:

Table III resumes the average time span between the TP prediction of a collision and near-distance situation and its occurrence. All observed effects indicate an extension of the prediction time for scenarios without MSIs, both for collisions and near-distance situations. Again, the sample size does not allow for the conclusion of significant correlations.

3) Effect of MSIs on the confidence of predictions and detections:

The range of possible confidence values is limited by low (1) and high (5) confidence. All stated confidence values are in the upper half of possible values, irrespectively of the veracity of the situational description that is specified with the confidence value. Comparing true and false detections with and without MSIs a slight increase of confidence in TP and a reduction in FP is visible for collisions. With a sample size of 21 observations, the difference in confidence between near-distance situations with and without MSIs for TP predictions and events is not significant.

C. Results of scenario-level assessment with MSIs calculated by *datAcron* prototype

The goal of this second scenario experiment is to compare the impact of MSIs labelled by experts and MSIs calculated by *datAcron* on the experts assessment of the situation. For this the MSIs displayed to the operator are once the results of the labelling process of domain experts, referred to as “labelled” situations, and once the detection results of *datAcron* components, referred to as “detected” situations. In contrast, prediction and detection are performed by the operator. While for the prediction the event didn’t occur in the reference dataset, yet, the detection requires the event happening or already having happened.

1) Effect of MSIs on prediction and detection:

With labelled data, the same situation is misclassified three times less than with detected events, cf. FNR in Table V. Table V also indicates lower true positive rates and higher false negative rates of operator assessments for situations

	SI		CER		SG		
	Within an area MSI#02	Null speed MSI#06	Rendez-vous MSI#28	Underway MSI#19	Change of course MSI#12	Change of speed MSI#07	Stopped MSI#06
Message count in annotated dataset	4077	3035	3035	3035	3035	3035	3035
Annotated detections	521	871	34	2224	712	96	871
Datacron detection	523	29	1	431	907	598	204
True positives	521	0	1	422	396	33	171
False positives	2	29	0	9	511	565	33
True negatives	3554	2135	3000	802	1812	2374	2131
False negatives	0	871	33	1802	316	63	700
Detection probability	1	0,00	0,03	0,19	0,56	0,34	0,20
Missed detection probability	0	1,00	0,97	0,81	0,44	0,66	0,80
False alarm probability	0,004	1,00	0,50	0,02	0,56	0,94	0,16

Figure 3: Expert-based Comparison of *datAcron* Results vs. Enriched and Annotated Data

Table II: Summary - True Positive, False Positive and False Negative Rates

Situation	Prediction			Detection		
	TPR	FPR	FNR	TPR	FPR	FNR
Collision without MSIs	0.33	0	0.67	0.33	0	0.67
Collision with MSI	1	0	0	0.67	0	0.33
Rendezvous without MSI	0.33	0.67	0	0.67	0.33	0
Rendezvous with MSI	0.67	0.33	0	0.67	0.33	0
Near-collision with MSI	1	0	0	0.83	0.17	0

Table III: Average time between TP prediction and event occurrence.

Situation	Prediction
	av. time (min)
Collision without MSIs	3
Collision with MSI	0.5
Near-distance situations without MSIs	4
Near-distance situations with MSI	2

with MSIs detected by *datAcron* components compared to situations with MSIs labelled by domain experts. This finding is observable both for the prediction and for the detection of near-distance situations.

2) Effect of MSIs on the prediction time:

Table VI describes the average time span between the TP prediction of a near-distance situation and its actual occurrence. The time span is larger for near-distance situations enriched with MSIs labelled by domain experts. As the situational datasets are shown to the operators accelerated by factor 3, the 1 minute difference between the average prediction time spans correspond to 3 minutes real time. Concluding, the detected MSIs reduce the time span between prediction and occurrence of the respective near-distance event.

3) Effect of MSIs on the confidence of predictions and detections:

Given that only positive assertions are counted, table VII does not allow a comparison of prediction confidence due to the lack of stated values. For all stated values the findings of the first scenario experiment are confirmed in the sense that for both positive true and false rated events the operators state a relatively high confidence value. For true positive detections, the labelled situations receive marginally higher confidence levels, which are not significant, given the small number of confidence values.

4) Effect of MSIs on the situational awareness:

Table VIII summarizes the differences in the situational awareness between manually labelled and detected MSIs. The availability of interactive functionalities and MSIs detected by *datAcron* components has a positive effect on the maritime situational awareness of operators. Especially, the information quantity is assessed to be larger, the situations are perceived to be less unstable, less complex and less variable. The operators feel less in the status of arousal, need to concentrate less and their attention is less divided. Further, the experts users

Table IV: Average confidence in prediction and detection.

Situation	Prediction			Detection		
	TP	FP	FN	TP	FP	FN
Collision without MSIs	-	-	-	3	4.5	5
Collision with MSI	4	-	-	4.5	3	5
Near-distance situation without MSIs	4	5	-	5	4.3	5
Near-distance situation with MSI	4.5	-	-	4.85	4.3	5

Table V: Difference in the operators accuracy between manually labelled and datArcon detected MSIs

Situation	Prediction			Detection		
	TPR	FPR	FNR	TPR	FPR	FNR
Near-distance situation labelled	0.83	-	0.17	0.57	0.14	0.29
Near-distance situation detected	0.5	-	0.5	0.33	-	0.66

Table VI: Average time between TP prediction and event occurrence.

Situation	Prediction
	Average time (min)
Near-distance situations labelled	2.3
Near-collision situations detected	1.3

Table VII: Average confidence in prediction and detection from low (1) to high (5) confidence.

Situation	Prediction			Detection		
	TP	FP	FN	TP	FP	FN
Near-distance situations labelled	5	-	-	5	5	-
Near-distance situation detected	-	-	-	4	-	-

estimate in average to have more spare mental capacity, even though experiment 3 is conducted consecutively to experiment 2 which would let assume that the spare mental capacity decreases. Thus, the proposed hypothesis can be rejected and replaced by the alternative hypothesis stating that interactive functionalities and MSIs detected by datAcron components increase the situational awareness. As discussed before, the average ratings are not representative for all operators. In comparison to experiment 2, the answers of the operators diverge more importantly.

Table VIII: Average Situational awareness self assessment for manually labelled MSIs and MSIs detected by datAcron prototype. Situational awareness rating cp. [11]: 1-Low, 7-High.

Dimension of situational awareness	labelled	datAcron detected
Instability of Situation	5.2	4.5
Complexity of Situation	4.2	3.2
Variability of Situation	4.5	3.8
Arousal	5.2	4.8
Concentration of Attention	4.2	3.5
Division of Attention	4.2	3.8
Spare Mental Capacity	5	5.2
Information Quantity	4.5	5.2
Familiarity with Situation	5.8	6.2

V. METHOD EVALUATION

For estimating the impact of the dataset design and the data labelling process on the obtained results, the results on two near-collision situations both enriched with MSIs are compared. In the following, the two near-collision situations are referred to as control situations. The choice of near-collision situations as control situation is justified by its similarity to collisions in the prediction phase, since both situations are perceived as close-quarter situations and by its similarity to rendezvous in the detection phase, since both situations allow the vessels the continuation of their route. The two control situations are included in different scenario datasets and occur in different locations with different vessels, different AIS trajectories and at different seconds of the scenario dataset. The effect of the AIS trajectory construction methodology and the subsequent labelling process are assumed to be negligible, if the difference between the two control situations is small compared to the differences between situations with and without MSIs. In order to estimate if the difference between the control situations is small, the same measures are used as for the comparison between scenarios with and without MSIs.

Table IX describes the similarly rated operator assessment of the two control datasets. The variance of the operators assessment due to the different design of the same maritime situation is smaller than the variance introduced by the addition of MSIs shown in Table II for the case of collision and rendezvous.

Table XI shows very similar average time spans between the two different control datasets. The difference of 0.5 minutes is small compared to the differences of 2.5 minutes for collisions and 2 minutes for near-distance situations in Table III.

Table X shows again very similar results on the confidence of differently rated predictions and detections of the control situations. Only one false negative detection of a near-distance situation exists, which is due to the classification of the first near-distance situation detection as a rendezvous. Despite this false classification the absolute difference in the average confidence of the rated situational assessments is smaller between the control situations than between situations with and without MSIs.

Concluding, the effect of the dataset design and labelling process, measured by the difference between two near-distance

Table IX: Control situations - True Positive, False Positive and False Negative Rates

Situation	Prediction			Detection		
	TP	FP	FN	TP	FP	FN
Near-collision with MSI	1	0	0	0.67	0.33	0
Near-collision2 with MSI	1	0	0	1	0	0

Table X: Control situations - Average confidence in prediction and detection.

Situation	Prediction			Detection		
	TP	FP	FN	TP	FP	FN
Near-distance situation 1 with MSIs	5	-	-	5	-	5
Near-distance situation 2 with MSI	5	-	-	5	-	-

Table XI: Control situations - Average time between TP prediction and event occurrence.

Situation	Prediction
	Average time (min)
Near-collision situations 1 with MSIs	2.5
Near-collision situations 2 with MSI	3

situations as control situations, is small compared to the differences found between situations with and without MSIs. This results appears to be valid especially for the prediction and assessment of the situations, for the confidence assigned to those predictions and assessments, as well as the remaining time between situation prediction and situation occurrence. As these findings support the assumption that the effect of the dataset design and labelling processes is small, the applied method is supposed to be suitable also for modelling and labelling the complementary situations and yielding situational datasets whose analysis with respect to an effect of MSIs on the given measures is well-founded.

A. Discussion

The above-presented results depict a consistent picture and emphasise its applicability of the methodology for the evaluation of big data systems in general. Specifically on the datAcron prototype, the summary of the results of the three experiments open three points for discussion:

- On the MSI level, the results of the accuracy assessment of the MSIs show a heterogeneous picture of datAcron component performances reaching from very good results with high true positive rates and low false negative and positive rates to very bad results with very low true positive rates and very high false positive and negative rates. The expert reported that the components outputs are sometimes very divergent from the annotations, which may rise some doubt on the subjective definition of event concepts.
- On the scenario level, the results of the first experiment are indicating that the MSIs improve the prediction and the detection of both collision and near-distance situations. MSIs are not extending the time between the correct prediction of a near-distance situation and its occurrence. MSIs are not changing the confidence of experts users in their predictions and detections of near-distance situations. Furthermore, MSIs are changing the

situational awareness of operators in two ways. Firstly and beneficially, MSIs increase the perceived information quantity of situations and operators state to be more concentrated on situations where MSIs are displayed. Especially MSI icons that correspond to COLREGs day and night shapes are easily recognised [13]. Secondly and adversely, MSIs increase also the perceived complexity and variability of situations and tend to reduce the spare mental capacity of operators.

- In the second scenario experiment the prediction accuracy, the prediction time span and the detection accuracy of the operators indicate to be better for situations enriched with MSIs labelled by domain experts than with MSIs detected by datAcron components. No difference in the confidence level of operators situational assessment, both because of the rough granularity of the confidence scale and because of the small difference of the recorded values. For the maritime situational awareness, the availability of interactive functionalities and MSIs detected by datAcron components have a positive effect. Especially, the information quantity is assessed to be larger, the situations are perceived to be less unstable, less complex and less variable.

VI. CONCLUSIONS

Concluding the proposition for a contribution on the user-centric assessment of maritime situation awareness solutions, the described methodology is valuable for two reasons. Firstly, the results captured by the proposed methodology are both plausible and consistent between the MSI- and scenario-level of the assessment. This result favours the possibility of a decomposition of big data solutions to maritime situational awareness both for the simultaneous development and evaluation in time critical development phases and for the deductive identification of components with potential for improvement. Secondly, the method for creating and labelling the reference dataset is evaluated in the experimental scope and is found to introduce much less variance than the effect of MSIs on the operator's situational awareness in terms of accuracy and prediction time. Although the findings obtained are not statistically significant, the results gathered during the different experiments delineate trends which are sound and suggest the applicability of the method for the evaluation of other big data solutions.

REFERENCES

- [1] Harris Georgiou, Sophia Karagiorgou, Yannis Kontoulis, Nikos Pelekis, Petros Petrou, David Scarlatti, and Yannis Theodoridis. Moving objects analytics: Survey on future location & trajectory prediction methods. *arXiv preprint arXiv:1807.04639*, 2018.
- [2] Elias Alevizos, Anastasios Skarlatidis, Alexander Artikis, and Georgios Paliouras. Probabilistic complex event recognition: A survey. *ACM Computing Surveys (CSUR)*, 50(5):71, 2017.
- [3] Gennady Andrienko, Natalia Andrienko, Wei Chen, Ross Maciejewski, and Ye Zhao. Visual analytics of mobility and transportation: State of the art and further research directions. *IEEE Transactions on Intelligent Transportation Systems*, 18(8):2232–2249, 2017.
- [4] Christos Doukeridis, Apostolos Glenis, Panagiotis Nikitopoulos, Giorgos Santipantakis, Akrivi Vlachou, George Vouros, Nikos Pelekis, Harris Georgiou, Kostas Patroumpasa, and Georg Fuchs. Integrated prototype (final), H2020 datacron d1.12, 2018.
- [5] Maximilian Zocholl, Richard Dréo, Cyril Ray, Anne-Laure Joussetme, Clément Iphar, Francesca de Rosa, Elena Camossi, Guillaume Keraudren, and Florian Rozé. Maritime final validation, H2020 datacron d5.6. 2018.
- [6] Iciar García-Ovies Carro, Enrique Iglesias Martínez, David Scarlatti, Javier López Leones, and Miguel García Martínez. Aviation final validation report, H2020 datacron d6.6, 2018.
- [7] Enmei Tu, Guanghao Zhang, Lily Rachmawati, Eshan Rajabally, and Guang-Bin Huang. Exploiting ais data for intelligent maritime navigation: A comprehensive survey. *arXiv preprint arXiv:1606.00981*, 2016.
- [8] Ronan Fablet, Nicolas Bellec, Laetitia Chapel, Chloé Friguet, René Garello, Pierre Gloaguen, Guillaume Hajduch, Sébastien Lefèvre, François Merciol, Pascal Morillon, et al. Next step for big data infrastructure and analytics for the surveillance of the maritime traffic from ais & sentinel satellite data streams. In *BiDS'2017-Conference on Big Data from Space*, pages 1–4, 2017.
- [9] Ramona Pelich, Nicolas Longépé, Grégoire Mercier, Guillaume Hajduch, and René Garello. Ais-based evaluation of target detectors and sar sensors characteristics for maritime surveillance. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(8):3892–3901, 2015.
- [10] Anne-Laure Joussetme, Cyril Ray, Elena Camossi, Melita Hadzagic, Christophe Claramunt, Karna Bryan, Eric Reardon, and Michael Ilteris. Maritime use case description, H2020 datacron d5.1, 2016.
- [11] RM Taylor. Situational awareness rating technique (sart): The development of a tool for aircrew systems design. In *Situational Awareness*, pages 111–128. Routledge, 2017.
- [12] Helmut Hilgert. Defining the close-quarters situation at sea. *The Journal of Navigation*, 36(3):454–461, 1983.
- [13] Convention on the international regulations for preventing collisions at sea (COLREGs). Standard, International Maritime Organization, 1972.