



City Research Online

City, University of London Institutional Repository

Citation: Claramunt, C., Ray, C., Salmon, L., Camossi, E., Hadzagic, M., Joussetme, A-L., Andrienko, G., Andrienko, N., Theodoridis, Y. & Vouros, G. (2017). Maritime data integration and analysis: Recent progress and research challenges. *Advances in Database Technology - EDBT*, 2017-March, pp. 192-197. doi: 10.5441/002/edbt.2017.18

This is the published version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/19854/>

Link to published version: <https://doi.org/10.5441/002/edbt.2017.18>

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Maritime Data Integration and Analysis: Recent Progress and Research Challenges

**Christophe Claramunt,
Cyril Ray, Loïc Salmon**
Naval Academy Res. Inst.
Lanvéoc, France

**Elena Camossi,
Melita Hadzagic,
Anne-Laure Jousset**
CMRE, La Spezia, Italy

**Gennady Andrienko,
Natalia Andrienko**
Fraunhofer Institute IAIS
Sankt Augustin, Germany

**Yannis Theodoridis,
George A. Vouros**
University of Piraeus
Piraeus, Greece

ABSTRACT

The correlated exploitation of heterogeneous data sources offering very large historical as well as streaming data is important to increasing the accuracy of computations when analysing and predicting future states of moving entities. This is particularly critical in the maritime domain, where online tracking, early recognition of events, and real-time forecast of anticipated trajectories of vessels are crucial to safety and operations at sea. The objective of this paper is to review current research challenges and trends tied to the integration, management, analysis, and visualization of objects moving at sea as well as a few suggestions for a successful development of maritime forecasting and decision-support systems.

Keywords

Big Spatio-temporal Data, Moving Objects, Maritime Information Systems, Event detection, Forecasting, Uncertainty

1. INTRODUCTION

The maritime environment has a huge impact on the global economy and our everyday lives. Specifically, Maritime Situation Awareness (MSA) and surveillance systems have been attracting increasing attention due to their importance for the safety and efficiency of maritime operations. Safety and security are constant concerns of maritime navigation, especially when considering the continuous growth of maritime traffic around the world and persistent decrease of crews on-board. For instance, preventing ship accidents by monitoring vessel activity represents substantial savings in financial cost for shipping companies (e.g., oil spill cleanup) and averts irrevocable damages to maritime ecosystems (e.g., fishery closure). This has favoured and led to the development of automated monitoring systems, such as the Automatic Information Systems (AIS) and institutional initiatives for maritime data infrastructures [31]. However, the necessary correlated exploitation of large data sources offering historical and streaming maritime data is still a crucial computational issue. For instance, a typical volume of radio and satellite-based worldwide maritime data represents an estimated 18 millions positions per day [16] (see Figure 1 for an illustration of AIS coverage at the global level).

©2017, Copyright is with the authors. Published in Proc. 20th International Conference on Extending Database Technology (EDBT), March 21-24, 2017 - Venice, Italy: ISBN 978-3-89318-073-8, on OpenProceedings.org. Distribution of this paper is permitted under the terms of the Creative Common License CC-by-nc-nd 4.0.

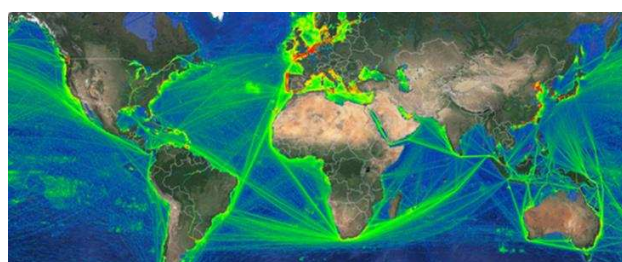


Figure 1: Worldwide AIS positions acquired by satellites (ORBCOMM)

Beside the indisputable value of information extracted from the AIS, the correlated exploitation of additional and heterogeneous sources is unavoidable to overcome the lack of veracity and incompleteness of the data. Thus, additionally to volume, velocity and variety, veracity of maritime data poses significant challenges. In particular, AIS messages are vulnerable to manipulation and subject to hacking [44], due to the unsecured channel of transmission, which weakens the whole system and the safety of navigation [35]. AIS data can thus contain deliberate falsifications and undergo spoofing [36], such as identity fraud, obscured destinations, or GPS manipulations [43]. According to [44], approximately 50% of AIS static data transmissions have errors of any kind. Vessels involved in illicit activities such as illegal fishing, deliberately avoid transmitting their information, while others may simply want to keep secret their fishing area to others.

Moreover, despite large available volumes, AIS data at open seas or at the border of Exclusive Economic Zones (EEZs) may be sparse, or delayed due to either low coverage or to multi-level processing issues. The data sparseness, latency, possibly manipulated, and the poor quality of movement data in general [1], render very challenging the design of information systems to support MSA processing AIS data and detecting abnormal behaviours. Thus, in addition to the need of real-time processing of large volume of data of high velocity, maritime surveillance systems should also have the ability to process and correlate many data sources, ideally of wide variety to compensate any lack of veracity of the data. For instance, Long Range Identification and Tracking (LRIT) and Vessel Monitoring Systems (VMS), Synthetic aperture radar (SAR) imagery can be used to verify AIS emission and detect anomalies [19].

Indeed, vessel trajectories are quite unique with respect to terrestrial trajectories: they are in only a limited way constrained by rigid network infrastructures, landmarks (e.g., ports), prefixed

waiting / meeting points, and more difficult to observe and monitor. The objective of this paper is to review recent development and research challenges of methods, operational frameworks and systems oriented at large towards moving objects at sea. Most of the ideas and proposals developed in this paper are generated by the datAcron European funded project, which aims to advance the management and integrated exploitation of voluminous and maritime data sources, so as to significantly advance the capacities of systems to promote safety and effectiveness of critical operations for large numbers of moving entities in large geographical areas [12].

Our perspective towards an integrated maritime information infrastructure is presented in Figure 2. The different components identified cover the integration of in-situ streaming data, trajectories detection and forecasting, recognition and identification of complex events and the development of visual analytics interfaces for maritime experts and decision-makers.

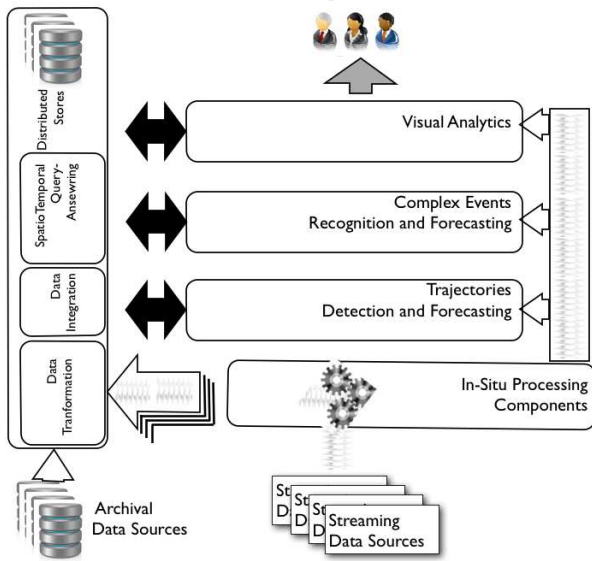


Figure 2: Towards an integrated maritime information infrastructure [12]

The rest of the paper is organised as follows. Section 2 introduces the main issues and challenges behind the integration of very large and heterogeneous maritime data and briefly surveys recent progress in information fusion, in-situ processing and database integration. Section 3 presents recent progress and remaining directions to explore in maritime event pattern and abnormal behaviour detection, trajectory analysis and visualisation. Section 4 explores issues, challenges and trends in maritime decision support and forecasting. Finally, Section 5 draws the conclusions.

2. MARITIME DATA INTEGRATION AND MANAGEMENT

The search for successful MSAs implies multiple data sources, such as surveillance sensors, automated processors, maritime institutional databases (e.g., navigation rules, protected areas), ocean and weather data, and “soft” data in unstructured formats (e.g., social media, intelligence reports). Nonetheless, advances in the spatio-temporal data analytics field with application on the urban domain (e.g., the very recent work over NYC Urban collection [10]) are not easily applicable in the maritime case. Efficient integration and management of maritime data is instrumental in effectively exploiting the available data, but it

nevertheless entails some challenges that will be discussed in this section.

2.1 In-situ data processing

To face challenges due to the volume, velocity and variety of data sources, in-situ processing aims to scale, by shortening the time needed for detecting patterns of interest within a single- or cross-streaming process; addressing this challenge has been the focus of a great deal of academic research and industry efforts in recent years [11]. For instance, a framework for a distributed stream processing architecture supporting in-situ processing has been presented in [5]. However, such approaches have to become communication efficient and have to learn abilities for automatic model adaption for handling concept drift.

In-situ processing for the detection of patterns can be worthily-investigated towards cross-streaming data integration, and integration of streaming data (e.g., regarding a specific vessel) with contextual information (e.g., weather data) given that detected patterns may further be joined and aggregated, producing output streams that provide semantically and contextually rich information, further enabling effectiveness in detection and predictive analytics.

Closely related to the in-situ processing paradigm is the computation of data synopses. In particular, the computation of trajectories synopses for individual vessels is challenging, given that state of the art techniques [29] have achieved a compression ratio of 95% over AIS vessel traces. The challenge here is to address high levels of data compression without compromising the accuracy of the prediction / detection components.

2.2 Streaming data integration

In close relation to the computation of data synopses, a major objective is to develop appropriate components for integrating and summarizing maritime streaming data sources producing a scalable framework for cross-streaming data integration. Summarized streams can be semantically integrated with archival data as well as with detected and forecasted vessel trajectories and events.

The database community has proposed many efficient algorithms for link discovery applied to RDF data [32] [39]. Major shortcomings of these approaches are (a) the restriction to RDF properties of specific (mostly numerical) types, (b) the not-proved ability to integrate in real-time (cross-) streaming with archival data. Viewing the annotation of trajectories with contextual data as part of link discovery task, a specific challenge is the computation of semantic trajectories, taking advantage of trajectory-specific works (e.g., [34]). Distributed paradigms for streaming data, such as Storm, Spark Streaming, and Flink Streaming provide richer sets of primitive for incremental data sources, however, they do not provide full support for the integration of heterogeneous and historical data sources as well as for semantic enrichment and querying facilities [21] [37]. Moreover, these systems still lack specific spatio-temporal primitives necessary to deal with moving object trajectory data management [15].

2.3 Streaming data management

Recent works focus on the volume of spatial data to be processed, developing systems specifically oriented to the spatial domain and particularly moving objects [29] [38]. Other works have been extended to process data “on the fly” to handle data velocity and provide fast response time in a “moving object context” [34]. Nevertheless, these systems are oriented either towards a “posteriori analysis” characterized by long processing times or

“on the fly processing” which can provide approximate answers to queries.

Provision of integrated views of data for exploring and querying streaming and archival data sources in real-time has also been addressed in the context of the Internet of Things (IoT): Live Knowledge Graphs, backed with scalable and elastic software stack can deal with millions of static records and billions of streaming triples per hour in real time [22] [26]. However, support of spatio-temporal queries and real-time integration of disparate data sources enabling scalability for massive amounts of dynamic data still remains a challenge. Concerning the representation and querying of spatial information, several RDF stores have begun integrating spatial query processing and reasoning [27]. Their performance still falls largely behind standard spatially-enabled DBMS's. Parallel and distributed platforms [46], key-value stores [33] and main memory systems, such as TriAD [20] and Trinity [47] have been developed for RDF data. However, current RDF stores with spatial and/or temporal support are not tailored to offer efficient trajectory-oriented data management, due to the volatile, multi-dimensional, and inherently sequential nature of such data (e.g., Strabon [24]).

2.4 Maritime data fusion

The information fusion literature addresses extensively the integration and combination of information from cooperative and non-cooperative maritime data sources. Information (data) fusion originally focused on “low-level” processing mainly from signals or images, from which vessel tracks (and trajectories) are built, new sensor measurements (contacts) are associated to tracks, objects corresponding to tracks are recognised and identified. The corresponding challenges include alignment of data in space and time, multi-resolution issues, at the same time, handling contextual and semantic differences.

The more recent trend in information fusion is higher-level processing with tasks, such as situation and impact assessments, closer to the decision maker, where the semantics has a higher importance, and involving human sources (hard and soft fusion) [13]. The integration and fusion of maritime data and information from various sources can overcome some of the single source processing issues (e.g., compensating for the lack of coverage and increasing accuracy). However, this also requires a suitable management of conflicting information, which may be either due to unintentional malfunction of sensors or to deliberate deception deemed of interest. The cross-fertilisation of database and fusion techniques will contribute to MSA just like the InFuse framework [14] or, more recently, the architecture proposed in [18].

2.5 Maritime data semantics and ontologies

A few semantic approaches, including vocabularies, taxonomies and ontologies have been proposed as tentatives to bridge the gap between low level data from maritime sensors and maritime domain semantics [25], for example to enhance the integration of maritime information [6], to model ships' behaviour [41], for patterns identification [2], abnormal behaviour detection [42], and prediction [7]. Indeed, semantics' representation and exploitation is preferably addressed at the application level, because existing semantic approaches and technologies are not adequate to address the requirements of multi-mission and multi-task MSAs. In particular, approaches for semantic multi-domain interoperability able to integrate heterogeneous information sources (e.g., surveillance data, weather and ocean data, registers, bulletins) and combine multi-representation formalisation of multiple contexts need to be developed.

Regarding the integrated exploitation of disparate data sources

in the maritime domain, semantic representation of maritime information in multi-scale and at multiple granularity levels brings new business opportunities as well as new research challenges. For instance, processing maritime data as linked stream data requires integration and joint processing of this data with quasi-static data from the Linked Data Cloud or other open data sources, in soft-real-time, that usually are at different scales and granularity levels. As additional examples, data from Earth Observation sensors and VMS data are at a lower temporal resolution than surveillance data from VTS radar and AIS, which have revisit times of few minutes; VTS radar spatial resolution is poor compared to GPS position accuracy from AIS, which is assumed to be around 10m; freely available meteorologic data have spatial resolution of few kilometres, and estimated and measured environmental variables are provided with hourly and daily means. In addition, using open data sources in the maritime domain is a challenge itself because of the different policies of European countries regarding the provision of environmental and other marine data to the users, that, depending on the data sources, can be classified at national level and not freely distributable at the necessary resolution.

2.6 Discussion

Support to real-time (semantic) integration, storage and spatio-temporal querying of disparate maritime data sources enabling scalability for massive amounts of dynamic data is a challenging task. A series of specific challenges are as follows:

- producing a scalable, fault-tolerant framework for cross-streaming data integration and processing of maritime data from multiple streaming sources, via the real-time computation of data synopses, achieving high rates of data compression;
- reconstruction of vessel trajectories and computation of events and multi-scale visualizations of data and patterns via advanced analytics techniques;
- incremental integration of maritime data, allowing advanced management and query answering of spatio-temporal data;
- automatic, real-time semantic annotation and linking of maritime data towards generating coherent views on integrated cross-streaming and archival data;
- efficient distributed management and querying of integrated trajectory and contextual data.

3. EVENT PATTERN DETECTION AND TRAJECTORY ANALYSIS

3.1 Event detection

Detection of anomalous vessel movements and analysis of suspicious vessel trajectories are crucial assets for improving the security of vessel traffic. The range of possible events of interest is very large, from detecting vessels in distress and collisions at sea to discovering illegal fishing and any other illicit activities occurring at sea such as contrabands and smuggling.

Detecting events and patterns of interest in the maritime domain requires, as a first step, correlating vessel trajectories with data expressing entities' characteristics, geographical information, weather data, patterns of mobility in specific areas, regulations, intentional data (e.g. planned routes) etc., in a timely manner, while addressing the challenges described in the previous section. As part of the data integration task, the annotation of trajectories with contextual data is expected to provide a more robust solution towards the semantic annotation of trajectories [34], the interlinking of events with trajectories and with other contextual

data that are of particular importance for maritime domain awareness and decision-making. These present major challenges as they do concern both archival and streaming data, and they do require dealing with spatio-temporal features at multiple scales and dimensions (e.g., for determining the similarity among trajectories, or for relating events to trajectory segments) [4].

The specific challenges of event detection and suspicious pattern identification in the maritime domain are driven by MSA objectives. In particular, the development of early warning anomaly detection algorithms supporting maritime operators in the identification of the potentially suspicious of dangerous activities in the maritime areas under surveillance encompasses many challenges, such as:

- real-time reconstruction of vessel trajectories, supported by real-time analysis of multiple and voluminous streams of data on possibly conflicting vessel positions;
- algorithms for the prediction of anticipated vessel trajectories at different time scale, which is fundamental to achieve early warning maritime monitoring;
- machine learning methods supporting the identification and the formalization of events and patterns that are of interest to maritime security operators, able to observe and learn from their behaviour;
- algorithms for complex event (and outlier) recognition and prediction in real-time, dealing with heterogeneous, fluctuating and noisy voluminous data streams of moving entities in large geographic areas, taking advantage of data analytics results over archival data.

3.2 Visual analytics

MSA may greatly benefit from the development of Visual Analytics (VA) methods oriented to the maritime domain. VA methods, being more oriented than traditional analysis approaches towards addressing human factors and enhancing user perception, may help obtaining better analysis results through a more effective integration of unformalized operative knowledge and expertise, which are of fundamental importance in surveillance activities [30]. Specific VA research challenges are as follows:

- interactive data exploration of both archival (data-at-rest) and streaming (data-in-motion) spatio-temporal data, with varying levels of resolution and quality;
- exploration of real-time maritime moving entities integrating contextual and historical information at varying levels of resolution, supporting operators in early alerting validation;
- scalable spatio-temporal analytical querying, such as drill-down / zoom-in and on user-defined spatio-temporal regions of interest for surveillance;
- interactive pattern extraction (and assessment of data quality) considering both data-in-motion and data-at-rest, able to visually integrate information on sensor performance to validate early alerts obtained by the analysis tools;
- user-guided model building and validation, aiming at visual steering of modelling tools enabling interactive selection of model types, tuning model parameters, and analysis of model residuals in multiple dimensions, including space and time;
- building situation overview and situation monitoring, capable of computing an overall operational picture of mobility at desired scales and levels of detail, both in spatial and temporal dimensions. Monitoring needs to provide alarms and explanations if observations significantly deviate from models.

4. TOWARDS A MARITIME DECISION SUPPORT AND FORECASTING SYSTEM

The variety of data sources is expected to provide an improved MSA to the operator taking advantage of the complementarity and redundancy provided. For instance, the knowledge captured in databases of records of events like piracy events or incidents at sea, the lists of vessels of interest such as blacklisted vessels, may provide the relevant context to understand and explain some events of interests. If the processing and correlation of data and information from different sources and databases can also overcome for some incompleteness in databases and compensate for sparseness, it however may reveal some inconsistencies that need to be managed. For instance, ship information from the MarineTraffic¹ database may conflict with that from Lloyds²: the length may differ slightly, or the flag may be different due to a lack of update in one source. In this regards, additional knowledge on sources' quality may help solving the issue. An example of estimating and exploiting the AIS reliability is proposed in [8], while for enhancing the reliability of AIS, the vessel identity verification method has been used by the US Coast Guard (USCG)'s Maritime Information for Safety and Law Enforcement (MISLE) and Vessel Documentation System (VDS) [44].

The correlation with social media information [3] can furthermore help in establishing links with external events for a better global picture. However, most of these sources contain natural language information (possibly in different languages), which needs to be automatically processed, analysed, interpreted and finally correlated with other data from physical sensors. The fusion of human generated information ("soft") with sensor data ("hard"), thus named "hard and soft fusion", has been recently widely addressed [28] and brings promising avenue to the MSA problem [17], in keeping the human at the core of the processing. Hence, the design of an efficient information system for Maritime Domain Awareness and decision support should consider the maritime data quality issues in their entirety and diversity to ideally resolve them or at least to not occult them and rather inform the operator of some possible output uncertainty.

Probabilistic databases are certainly a promising avenue for the maritime domain [3] [23], which allows to deal for instance with empty fields very common in marine data, approximate values or uncertain fields. Besides, the consideration of the open-world assumption is unavoidable if one wants to provide a realistic outcome to the user [9]. Indeed, the AIS database clearly violates the closed-world assumption since, according to Windward [43], 27% of ships do not transmit data at least 10% of the time ('go dark'). Consequently, querying for instance *rendez-vous* events from an AIS database will return only those events reflected by the AIS data. Considering that anything which is not in the AIS database remains possible is thus crucial to maritime anomaly detection. Moreover, the extension to other uncertainty representations such as evidence or possibility theories is certainly desirable for maritime anomaly detection and event forecasting in order to cope with the different nature of uncertainty (probabilistic, subjective, vague, ambiguous, etc) due to the variety and poor veracity of the sources.

Although no clear guidelines exist so far for the selection of the appropriate uncertainty framework and aggregation (or fusion) rule, it is acknowledged that the choice depends on the nature, interpretation or type of uncertainty and information, and on the sources quality and independence [45]. Considering second-order

¹ <http://www.marinetraffic.com>

² <http://www.lloydslistintelligence.com>

uncertainty seems also unavoidable if one wants to properly account for the imperfection of data in the estimation of patterns-of-life, in developing approximate models of vessels' motion or in recognition of vessels, but also if one wants to communicate to the user faithful information. Interestingly, similar uncertainty challenges recently arose within the visualization community. A great challenge is to enable reasoning under uncertainty (in all its forms) uncertainty throughout the processes of sensemaking, decision-making, and action-taking [23].

In the aim of developing trustful and useful decision support systems, the human must be considered in his/her two main roles of (i) source of information and (ii) decision maker. The underlying challenges are then, on the one hand, to properly capture the human generated information including the associated uncertainty assessment so it can be meaningfully aggregated with other information from physical sensors or databases, and on the other hand, to ensure that the system outputs meaningful, interpretable and unambiguous results on which the user can take an informed decision. The design of information systems should provide a flexible architecture to ensure both the utility of the output provided and adaptability to dynamic and unforeseen events, and to changing user's needs. For instance, an explicit consideration of context provides an understanding of normalcy as a reference for anomaly detection (i.e., *pattern-of-life*) [40]. It helps detecting, and distinguishing between, spoofed information and deception, reducing the set of possible hypotheses (e.g., classes) for threat classification; it also provides information about sources' quality such as reliability or truthfulness.

Finally, the development of decision support systems for maritime event detection and forecasting should provide (1) the necessary simplicity to the processes by a judicious filtering of information suited to the users' needs, (2) the suitable flexibility and adaptability for the algorithms implementation by separating between the events of interest and their surrounding context, (3) the adequate uncertainty representation and processing considering the sources' quality and uncertainty's origin and (4) the expected human-system synergy for a better understanding of the system' outputs with associated explanations and simpler queries tuned to specific needs.

5. CONCLUSION

While most of current research in spatial databases and geographical information systems addresses issues often related to phenomena and practices related to the land domain, we believe that the maritime environment also provides many application opportunities and research challenges that still deserve to be addressed. This paper surveys a series of current computational issues still opened for a successful integration, manipulation and analysis of maritime information, with a specific focus on trajectories of moving objects at sea. We explore and suggest several research development directions that might contribute to a better use of voluminous and disparate sets of maritime data available so far. Due to the diversity and complexity of the problem, a successful solution should involve an integration of complementary contributions from different scientific domains, let us mention amongst many research areas ontology and conceptual data models at the data integration level, data mining and visual analytics for the ability to discover patterns within large volume of data, machine learning for streaming data, information fusion for the ability to combine information from different sources and deal with uncertainty, human factor and decision-aided systems. This is why we believe that not only many opportunities are still open for the extended database

community, but also avenues for further experimentations and interactions with the maritime world at large.

6. ACKNOWLEDGMENTS

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.

7. REFERENCES

- [1] G. Andrienko, N. Andrienko and G. Fuchs, Understanding movement data quality. *Journal of Location Based Services*, 10(1), 31-46, 2016.
- [2] H. Arenas, B. Harbelot and C. Cruz, A semantic analysis of moving objects using as a case study maritime voyages from eighteenth and nineteenth centuries. In *Proc. of GEOProcessing*, pp. 45-50, 2014.
- [3] M.L. Ba, S. Montenez, R. Tang, and T. Abdessalem, Integration of web sources under uncertainty and dependencies using probabilistic XML. In *Proc. of Database Systems for Advanced Applications*, pp. 360-375, 2014.
- [4] M. Balduzzi, A. Pasta and K. Wilhoit, A security evaluation of AIS automated identification system. In *Proc. of ACSAC*, pp. 436-445, 2014.
- [5] S. Bothe, V. Manikaki, A. Deligianakis and M. Mock, Towards flexible event processing in distributed data streams. In *Proc. of EPPForDM EDBT/ICDT Workshop*, 2015.
- [6] S. Brüggemann, K. Bereta, G. Xiao and M. Koubarakis, Ontology-based data access for maritime security. In *Proc. of ESWC*, pp. 741-757, 2016.
- [7] R. N. Carvalho, R. Haberlin, P. Cesar, G. Costa, K. B. Laskey and K. Chang, Modeling a probabilistic ontology for maritime domain awareness. In *Proc. of FUSION*, pp. 1-8, 2011.
- [8] D. Ceolin, W. R. van Hage, G. Schreiber and W. Fokkink, Assessing trust for determining the reliability of information. In *Situation Awareness with Systems of Systems*, Springer, New York, 209-228, 2013.
- [9] I.I. Ceylan, A. Darwiche and G. Van den Broeck, Open-world Probabilistic Databases. In *Proc. of KR*, pp. 339-348, 2016.
- [10] F. Chirigati, H. Doraiswamy, T. Damoulas and J. Freire, Data polygamy: the many-many relationships among urban spatio-Temporal data sets. In *Proc. of SIGMOD*, pp. 1011-1025, 2016.
- [11] G. Cugola and A. Margara, Processing flows of information: From data stream to complex event processing. *ACM Computing Surveys*, 44(3), article 15, 2012.
- [12] datACRON EU H2020 project: Big Data Analytics for Time Critical Mobility Forecasting. URL: <http://datacron-project.eu>.
- [13] D. Dubois, W. Liu, J. Ma and H. Prade, The basic principles of uncertain information fusion: An organised review of merging rules in different representation frameworks, *Information Fusion*, 32A, 12-39, 2016.
- [14] O. Dunemann, I. Geist, R. Jesse, K.-U. Sattler, and A. Stephanik, A database-supported workbench for information fusion: INFUSE. In *Proc. of EDBT*, pp. 756-758, 2002.
- [15] A. Eldawy and M. F. Mokbel, The era of big spatial data. In *Proc. of ICDE Workshops*, pp. 42-49, 2015.

- [16] EMSA, Automated behaviour monitoring (ABM) algorithms – operational use at EMSA. In *Proc. of MKDAD Workshop*, pp. 12-16, 2016.
- [17] R. Falcon, R. Abielmona, S. Billings, A. Plachkov, and H. Abbass, Risk management with hard-soft data fusion in maritime domain awareness. In *Proc. of CISDA*, pp. 1-8, 2014.
- [18] M. Fidali and W. Jamrozik, Concept of database architecture dedicated to data fusion based condition monitoring systems. In *Proc. of BDAS*, CCIS 424, pp. 515–526, 2014.
- [19] M. Guerriero, P. Willett, S. Coraluppi and C. Carthel, Radar/AIS data fusion and SAR tasking for Maritime Surveillance. In *Proc. of FUSION*, 2008.
- [20] S. Gurajada, S. Seufert, I. Miliaraki and M. Theobald: TriAD: a distributed shared-nothing RDF engine based on asynchronous message passing. In *Proc. of SIGMOD*, pp. 289-300, 2014.
- [21] S. J. N. Hughes, M. D. Zimmerman, C. N. Eichelberger and A. D. Fox, A survey of techniques and open-source tools for processing streams of spatio-temporal events. In *Proc. of IWGS@SIGSPATIAL*, 6:1-6:4, 2016.
- [22] M. F. Husain, J. P. McGlothlin, M. M. Masud, L. R. Khan and B. M. Thuraisingham, Heuristics-based query processing for large RDF graphs using cloud computing. *IEEE Trans. Knowl. Data Eng.* 23(9), 1312-1327, 2011.
- [23] A.-L. Joussemme and P. Maupin, Uncertainty representations for information retrieval with missing data. In *Fusion Methodologies in Crisis Management*, Springer, 87-104, 2016.
- [24] K. Kyzirakos, M. Karpasiotakis, and M. Koubarakis. Strabon: A semantic geospatial DBMS. In *Proc. of ISWC*, pp. 295-311, 2012.
- [25] A. Leadbetter, T. Hamre, R. Lowry, Y. Lassoued and D. Dunne, Ontologies and ontology extension for marine environmental information systems. In *Proc. of EISSAP at EnviroInfo*, Bonn, 2010.
- [26] D. Le-Phuoc, H.N. Mau Quoc, H.N. Quoc, T.T. Nhat and M. Hauswirth, The graph of things: a step towards the Live Knowledge Graph of connected things. *Journal of Web Semantics*, 37, 25-35, 2016.
- [27] J. Liagouris, N. Mamoulis, P. Bouros, and M. Terrovitis. An effective encoding scheme for spatial RDF Data. *PVLDB*, 7(12), 1271-1282, 2014.
- [28] J. Llinas, Challenges in information fusion technology capabilities for modern intelligence and security problems. In *Proc. of EISIC*, pp. 89-95, 2013.
- [29] J. Lu and R. H. Guting, Parallel secondo: practical and efficient mobility data processing in the cloud. In *Proc. of Big Data*, IEEE Press, pp. 17-25, 2013.
- [30] A. M. MacEachren, Visual analytics and uncertainty: It's not about the data. In *Proc. of EuroVis Workshop on Visual Analytics*, 2015.
- [31] Maritime CISE, A Common Information Sharing Environment for Maritime Surveillance in Europe <https://webgate.ec.europa.eu/maritimeforum/en/frontpage/1046> (accessed in June 2016)
- [32] A.-C. Ngonga Ngomo. On link discovery using a hybrid approach. *J.Data Semant*, 1, 203-217, 2012.
- [33] N. Papailiou, I. Konstantinou, D. Tsoumakos and N. Koziris: H2RDF: adaptive query processing on RDF data in the cloud. *WWW (Companion Volume)*, 397-400, 2012.
- [34] C. Parent, S. Spaccapietra, C. Renso, G. Andrienko, N. Andrienko, V. Bogorny, M. L. Damiani, A. Gkoulalas-Divanis, J. Macedo, N. Pelekis, Y. Theodoridis and Z. Yan, Semantic Trajectories Modeling and Analysis. *ACM Computing Surveys*, 45(4), article 42, 2013.
- [35] K. Patroumpas, A. Artikis, N. Katzouris, M. Voudas, Y. Theodoridis and Y. Pelekis, Event recognition for maritime surveillance. In *Proc. of EDBT*, pp. 629-640, 2015.
- [36] C. Ray, C. Iphar, A. Napoli, R. Gallen and A. Bouju, DeAIS project: Detection of AIS spoofing and Resulting Risks. In *Proc. of OCEANS'15 MTS/IEEE*, 2015.
- [37] L. Salmon and C. Ray, Design principles of a stream-based framework for mobility analysis. *GeoInformatica*, 20(2), 1-25, 2016.
- [38] S. Shekhar, V. Gunturi, M. R. Evans, and K. Yang, Spatial big-data challenges intersecting mobility and cloud computing. In *Proc. of MobiDE*, pp. 1-6, 2012.
- [39] Silk (<http://wifo5-03.informatik.uni-mannheim.de/bizer/silk>) by Univ. of Mannheim
- [40] L. Snidaro, J. Garcia and J. Linas, Context-based information fusion: a survey and discussion. *Information Fusion*, 25, 16-31, 2013.
- [41] W. R. van Hage, V. Malaisé, G. de Vries, G. Schreiber, and M. van Someren, Combining ship trajectories and semantics with the simple event model (SEM). In *Proc. of EiMM*, pp. 73-80, 2009.
- [42] A. Vandecasteele and A. Napoli. An enhanced spatial reasoning ontology for maritime anomaly detection. In *Proc. of IEEE SOSE*, pp. 247-252, 2012
- [43] Windward, AIS Data on the High Seas: An Analysis of the Magnitude and Implications of Growing Data Manipulation at Sea, <http://www.windward.eu/wp-content/uploads/2015/02/AIS-Data-on-the-High-Seas-Executive-Summary-Windward-October-20-2014.pdf> (accessed in June 2016).
- [44] D. Winkler, Enhancing the reliability of AIS through vessel identity verification, US DOT Datapalooza, https://www.fhwa.dot.gov/2015datapalooza/presentations/Safety4_Winkler.pdf (accessed in June 2016).
- [45] Z. Yu, Y. Liu, X. Yu, and K. Q. Pu, Scalable distributed processing of K-nearest neighbor queries over moving objects. *IEEE Trans. Knowl. Data Eng.*, 27(5), 1383-1396, 2015.
- [46] X. Zhang, L. Chen, Y. Tong and M. Wang: EAGRE: Towards scalable I/O efficient SPARQL query evaluation on the cloud. In *Proc. of ICDE*, pp. 565-576, 2013.
- [47] K. Zeng, J. Yang, H. Wang, B. Shao and Z. Wang: A Distributed Graph Engine for Web Scale RDF Data. *PVLDB* 6(4), 265-276, 2013.