Defence Research and Development Canada | Recherche et développement pour la défense Canada

CAN UNCLASSIFIED

DRDC | RDDC
technologysciencetechnologie

Semantic Integration of Real-Time Heterogeneous Data Streams for Ocean-related Decision Making

Renata Dividino, Amilcar Soares, Stan Matwin
Institute for Big Data Analytics, Dalhousie University

Anthony W. Isenor, Sean Webb
DRDC – Atlantic Research Centre

Matthew Brousseau
Institute for Big Data Analytics, Dalhousie University

IST-160 Specialists' Meeting on Big Data and Artificial Intelligence for Military Decision Making
NATO Science and Technology Board, Bordeaux, France

STO-MP-IST-160
Pagination info: S1-3 - 1–12

Date of Publication from Ext Publisher: May 2018

Defence Research and Development Canada

External Literature (N)
DRDC-RDDC-2018-N082
June 2018

CAN UNCLASSIFIED

Canada

IMPORTANT INFORMATIVE STATEMENTS

Semantic Integration of Real-Time Heterogeneous Data Streams for Ocean-related Decision Making

Renata Dividino[1], Amilcar Soares[1], Stan Matwin[1], Anthony W. Isenor[2], Sean Webb[2], and Matthew Brousseau[1]

1- Institute for Big Data Analytics, Dalhousie University, Halifax, CANADA

{dividino, amilcar.soares, m.brousseau}@dal.ca, stan@cs.dal.ca

2 - Defence Research and Development Canada (DRDC) Atlantic Research Centre, Halifax, CANADA

{Anthony.Isenor, Sean.Webb}@drdc-rddc.gc.ca

ABSTRACT

Information deluge is a continual issue in today's military environment, creating situations where data is sometimes underutilized or in more extreme cases, not utilized, for the decision-making process. In part, this is due to the continuous volume of incoming data that presently engulf the ashore and afloat operational community. However, better exploitation of these data streams can be realized through information science techniques that focus on the semantics of the incoming stream, to discover information-based alerts that generate knowledge that is only obtainable when considering the totality of the streams. In this paper, we present an agile data architecture for real-time data representation, integration, and querying over a multitude of data streams. These streams, which originate from heterogeneous and spatially distributed sensors from different IoT infrastructures and the public Web, are processed in real-time through the application of Semantic Web Technologies. The approach improves knowledge interoperability, and we apply the framework to the maritime vessel traffic domain to discover real-time traffic alerts by querying and reasoning across the numerous streams. The paper and the provided video demonstrate that the use of standards-based semantic technologies is an effective tool for the maritime big data integration and fusion tasks.

Keywords*. Data Integration, Data Streams, Streaming Reasoning, Maritime Situation Awareness, Semantic Web, IoT Platforms.*

1.0 INTRODUCTION

Maritime sensors relevant to defence and security, such as the Automatic Identification System (AIS), produce large volumes of spatially distributed data every day. When considered with data from weather forecasting systems, such as those providing global-scale wind, wave height, wave period, or pressure data, the potential for understanding cause and effect between environment and vessel increases, at the expense of managing the incoming data volume. Unfortunately, with the limitations of standard technologies to adequately store, access, integrate and process this significant amount of data, information is sometimes underutilized or in more extreme cases, not utilized at all. The ability to handle and process stream data and to reduce the overload of information is essential in many of the maritime monitoring scenarios.

Maritime monitoring systems or Real-time Maritime Situation Awareness (RMSA) systems rely on sensor observations to monitor the marine environment, with these observations originating from heterogeneous and spatially distributed Internet of Things (IoT) infrastructures. When data from these different sensors are combined, there is potential to predict unusual events, and potentially prevent negative defence, security or environmental events. Integrating such heterogeneous streams together, and further linking them to previously stored information in databases or from Web sources is typically very challenging. In part, this is due to the continuous volume of incoming data that presently engulf the ashore and afloat operational community. However, it is also due to the decentralized, self-organizing nature of the information, i.e., data is outside the control of only one organization, it is distributed and it is prone to corruption when shared or exchanged among users. Subsequently, the information is highly variable in quality and is described using different terminology, different structures (or even a combination of structured and unstructured data), and formats. We believe that the efficiency and prediction accuracy of a RMSA system relies on the ability of such a system to integrate and inter-link heterogeneous data sources. This, however, can only be achieved by the adoption of standards and common models that promote a shared and common understanding of the domain and enable comprehensive data communication between people and application systems.

Furthermore, RMSA systems need to ensure the data streams are continuously integrated and efficiently processed when analyzing and forecasting unusual events. Hence, one of the major challenges in today’s data management and processing environment is to deal with the multi-faceted (e.g., volume, variety, etc.) aspects of the data itself. Better exploitation of these data streams can be realized through information science techniques that focus on the semantics of the incoming stream, and link such streams to background knowledge to discover information-based alerts that generate knowledge that is only obtainable when considering the data in its totality. Semantic technologies arise to address these problems. These new technologies are used for modeling, linking and integrating data across data sets. In particular, existing semantic models [9,2,11,8,18] have been proposed to "bridge" the gap between low-level sensor descriptions and maritime domain semantics. Such approaches aim to provide a standard view of the different data sources, and thus to facilitate integration, link and (re-) use across the data sources.

Approaches have emerged that combine semantic technologies with stream and event processing techniques. Research in these areas has proposed an ecosystem of solutions to perform real-time processing over heterogeneous and distributed data streams, more specifically, to disseminate [10], to query [14,1,5], and to reason [7,12] over the data streams.

In this paper, a framework is presented for the integration of the heterogeneous data streams produced by different maritime and marine sensors from IoT infrastructures, and for stream querying and processing. The heterogeneity of data is addressed as is stream processing using Semantic Web solutions. First, semantic models and mappings for data representation and integration of sensor streams with other information sources are presented. These use existing core and domain ontologies along with Semantic Web technologies. In using such technologies, the goal is to ensure the interoperability of the heterogeneous data. Then, continuous stream processing is used to identify unusual events and issue warnings/alerts to the user. To achieve these goals C-SPARQL [1] is used, this being a language for real-time and continuous queries over streams of resource description framework (RDF) data. The resulting system can perform incremental data processing on streams and rich background (static) knowledge, i.e., slowly change information which is represented in a language with formal semantics, such as ontologies. The main advantage of applying Semantic Web technologies to the RSMA scenarios is the possibility of automatic reasoning (instead of ad-hoc solutions) on stream data, this supporting the process of unusual event detection. A demonstration video is provided at https://bigdata.cs.dal.ca/resources.

2.0 SITUATION AWARENESS RUNNING EXAMPLES

Even though the framework can be used in different Real-time Maritime Situation Awareness scenarios, in the following three scenarios are presented for forecasting maritime traffic events. These scenarios will be use throughout the paper to guide us to collect the suitable data and to explain the system modules.

Scenario 1: Object movement associated with static geospatial areas. This scenario considers vessel movement and vessel speed limit regulations for defined geospatial areas. Different areas have different imposed speed restrictions, these restrictions imposed as a result of safety and environmental concerns (e.g., protection of marine mammals). The framework should send warnings whenever incoming data indicate a vessel is not respecting the speed limit restrictions for the particular area of operation.

Scenario 2: Object movement associated with dynamic areas. This scenario considers vessel movement and temporally variable areas as represented by weather patterns. Marine forecasts provide temporally varying areas that are monitored by the RMSA system to identify those areas where weather conditions are deteriorating for a given period of time. Such conditions should be considered when planning vessel operation.

Scenario 3: Object movement associated with quasi-static areas. This scenario considers vessel movement and quasi-static areas as represented by ice concentration. The Canadian Ice Service provides ice data on a weekly basis in the form of ice concentration, ice thickness, amount of open water, etc. Such conditions are important when related to attributes specific to the vessel, such as the vessel's ability to navigate through ice-infested water.

3.0 MARINE DATA

To demonstrate the scenarios described in the previous section, it is essential that the system manage and process both dynamic data streams and static or pre-existing information. Both dynamic and static forms are described, containing different levels of granularity, different frequency of collection, different structures, and diverse formats. Schemas used by the system are also described.

3.1 Streaming Data Observations

Automatic Identification System (AIS) is a common maritime data source that was originally designed for collision avoidance. Broadcasted AIS messages contain information regarding the particular vessel including ship identifying information, ship location, speed, heading, a rate of turn, destination and estimated arrival time. Since about 2010, AIS data has started to become publicly available on the Internet, and this has triggered the development of new applications and services, particularly in the field of vessel tracking and screening.

Weather Observations from Environment Canada provide current and historical weather data (http://climate.weather.gc.ca). Weather data and related atmospheric information (e.g. temperature, precipitation, degree days, relative humidity, wind speed and direction) are available for numerous locations across Canada, and are updated hourly.

Ice Observations from the Canadian Ice Service (CIS) are represented as regional sea ice charts used for marine navigation, climate research, and input to the Global Digital Sea Ice Data Bank (GDSIDB) (http://nsidc.org/data/G02171). The ice charts are created through the manual analysis of in situ, satellite, and aerial reconnaissance data. The ice charts are updated daily and have information on ice concentration, stage of development, and ice form, following World Meteorological Organization (WMO) terminology.

3.2 Marine Static Knowledge

IMO Vessels Codes: The International Maritime Organization (IMO) number is a unique reference for ships and registered ship owners and management companies. IMO numbers were introduced to improve maritime safety and security and to reduce maritime fraud. For vessels, the IMO number remains linked to the hull for its lifetime, regardless of a change in name, flag, or owner.

Linked Open Data (LOD) Cloud: The Linked Open Data Cloud is a set of connected and distributed data sources across the Web that provides public access to structured and exchangeable data. LOD sources provide information about, such as geographical places, points of interest, and sensor observation. Some of the most popular LOD sources include the DBpedia (dbpedia.org), GeoNames (geonames.org), and OpenStreetMap (openstreetmap.org). For this work data from DBpedia is used in all scenarios. DBpedia provides a complementary service to Wikipedia by making Wikipedia-knowledge available in LOD form and allowing sophisticated queries against the Wikipedia information and data sets linked to it.

3.3 Schemas

Semantic Sensor Network (SSN) Ontology: The core ontology of the framework is the SSN ontology [4] which provides a framework to describe sensors, observations, and related concepts. The SSN ontology can describe sensor capabilities, measurement processes, observations, and deployments. The SSN ontology used is in ontology web language (OWL) form [15]. SSN was created by the W3C Semantic Sensor Network Incubator Group (The SSN-XG) and initially designed for the ocean sensor observations domain.

Domain ontologies: Domain ontologies contain fundamental ideas of a particular domain, providing potential terms for describing knowledge dedicated to the particular domain. In this case, domain ontologies are used for describing concepts and relationships about sensor-specific instrumentation, e.g. the climate domain ontology describes the terms and vocabulary defined by WMO standards which are used to describe the data of 1480 climate stations in Canada; likewise the ice domain ontology describes ice total concentration, stage of development, and ice form follows the WMO terminology. Different work in the literature [9, 2,11,18], have proposed vocabulary extensions, taxonomies, and ontologies in the maritime domain.

4.0 REAL-TIME MARITIME SITUATION AWARENESS SYSTEM

The following section contains a description of the data integration framework for the heterogeneous data streams produced by the different IoT infrastructures, the static information, and the real-time querying capability of the system.

4.1 Stream Publication and Ingestion

The following utilizes marine devices that directly or indirectly provide data streams to the network, in multiple data formats. Although formats vary across streams, they remain consistent for a single stream and are both structured and machine readable. As a means to join the various streams, the Resource Description Framework (RDF)[19] is used to represent objects and relationships, as RDF offers a flexible, graph-based model.

The system also utilizes JSON for Linking Data (JSON-LD). JSON-LD (https://json-ld.org) is introduced as a means to provide interoperability between JSON representations and Web-based data streams in RDF. To ensure interoperability among the streams, the unstructured or semi-structured streams are transformed into JSON-LD streams using the Triplewave [10] system. TripleWave is a reusable, generic, and open-source tool that enables the publication of JSON-LD/RDF streams, in real-time, on the Web.

This transformation is based on a set of mapping rules. Additionally, standard vocabularies are incorporated thus resulting in consistent terminology across the many contexts. As well, ontologies are utilized to describe *contextual relationships* behind the defined vocabulary, thus allowing richer bilateral mappings. In particular, customized mapping rules are created using the relational database (RDB) to RDF Mapping Language (R2RML) and based on the SSN Ontology and particular domain ontologies. Each device observation (e.g., AIS, weather/climate, ice) is represented as semantically enriched and interoperable stream. Figure 1 and 2 show an overview of the SSN ontology and some of the mappings created to semantically enrich the climate stream.

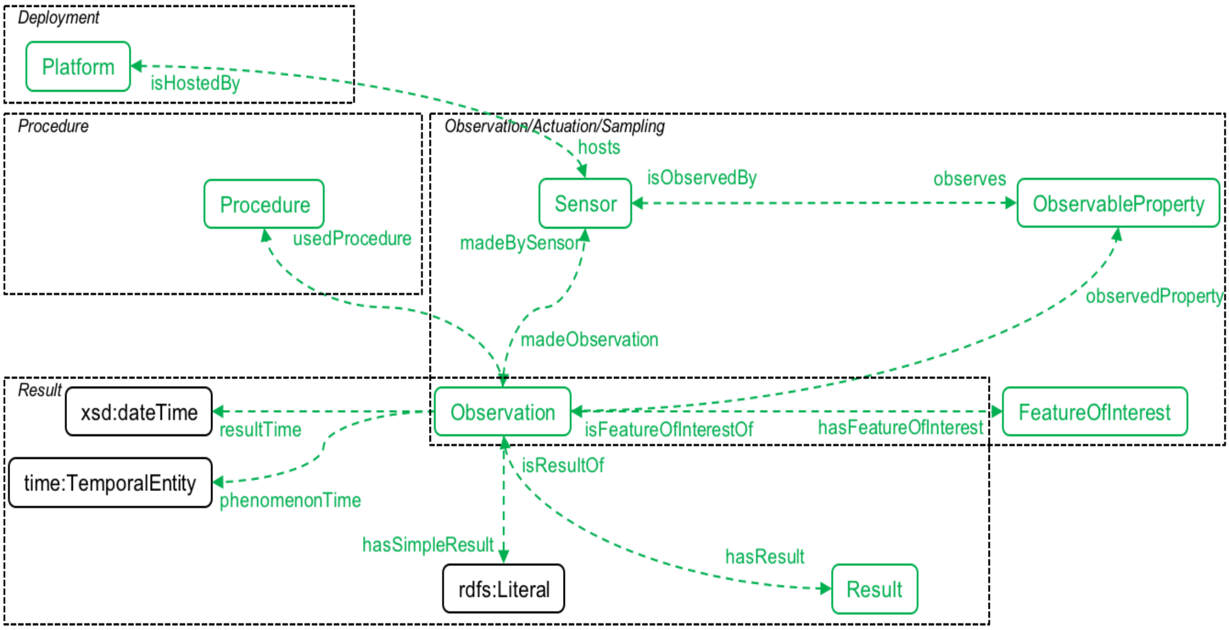

Figure 1: Overview of the SSN classes and properties (observation perspective). Copyright © 19-October-2017 World Wide Web Consortium, (MIT, ERCIM, Keio, Beihang). http://www.w3.org/Consortium/Legal/2015/doc-license". The original can be found in [20].

```
WeatherMap
  rr:subjectMap [
    rr:template "http://dal.ca/resource/Climate_Sensor_{Climate_Identifier}";
    ];
  rr:predicateObjectMap [
    rr:predicate rdf:type;
    rr:objectMap [ rr:constant ssn:Sensor]
    ];
  rr:predicateObjectMap [
    rr:predicate ssn:madeobservation;
    rr:objectMap [ rr:parentTriplesMap :ObservationTemperatureDataMap]
    ].

ObservationTemperatureDataMap
  rr:subjectMap [
    rr:template
"http://dal.ca/resource/Climate_Sensor_{Climate_Identifier}_Observation_Temperature_{timestamp}";
    ];
  rr:predicateObjectMap [
    rr:predicate rdf:type;
    rr:objectMap [ rr:constant ssn:Observation]
    ];
  rr:predicateObjectMap [
    rr:predicate ssn:observedProperty;
    rr:objectMap [ rr:parentTriplesMap :TemperatureDataMap ]
    ];
  rr:predicateObjectMap [
    rr:predicate ssn:observationResult;
    rr:objectMap [ rr:parentTriplesMap :ObservationResultTemperatureDataMap ]
    ];
  rr:predicateObjectMap [
    rr:predicate ssn:observationSamplingTime;
    rr:objectMap [ rr:parentTriplesMap :ObservationSamplingTimeMap ]
    ].

:ObservationResultTemperatureDataMap
  rr:subjectMap [
    rr:template
"http://dal.ca/resource/Climate_Sensor_{Climate_Identifier}_Observation_Result_Temperature_{timestamp}";
    ];
  rr:predicateObjectMap [
    rr:predicate rdf:type;
    rr:objectMap [ rr:constant ssn:ObservationResult]
    ];
  rr:predicateObjectMap [
    rr:predicate ssn:hasValue;
    rr:objectMap [ rr:column "temperature" ;rr:datatype xsd:float]
    ].
```

Figure 2: Example of R2RML mapping rules for climate stations

The details of Figure 2 are as follows. The first map, WeatherMap, states that a device is of type ssn:Sensor (described in the SSN ontology), and this sensor made observations. These observations are mapped in the

ObservationTemperatureDataMap map which states that every observation is of type ssn:observation, observes a property (in this case the temperature property), has a value or a result, and has been observed in a certain point in time. The observation results are mapped in the **ObservationResultTemperatureDataMap.**

4.2 Querying Sensor Observations and Background knowledge

Once the data from the distributed and heterogeneous data sources are available in a homogeneous, contextualized and ordered representation, the totality of the streams may be explored to generate new information. For this component of the work, the C-SPARQL [1] query language is used to apply queries against the streaming data.

C-SPARQL provides continuous query capabilities in a SPARQL query language [16]. Both C-SPARQL and SPARQL, represent queries as query graphs, with query evaluation performed through graph pattern matching over graphs formed by the incoming data streams. C-SPARQL queries consider temporal windows of the incoming data in a continuous manner. C-SPARQL also allows processing across streams and static information sources. The constructed RDF graphs may be locally stored or from external sources such as the LOD cloud data sources. This is of high relevance for forecasting the abnormal events since it involves association of information across multiple data streams and static sources.

Furthermore, as the data provided by the devices are geo-located, the system needs to have the ability to ask and answer questions about geographic features and their attributes, and about the relationships between those features. These queries focus more on the physical relationships between data such as finding all the vessels that are near, or in a certain region. To enable C-SPARQL and spatial queries, the Apache Jena SPARQL engine (ARQ) is used for spatial searches with SPARQL [12] . As C-SPARQL is built on top of ARQ, the resulting integration results in the combination of streaming querying with spatial search capabilities.

Figure 3 shows the system's modules and data sets. The data processing module queries one or more stream sources to capture real-time information from the incoming data streams.

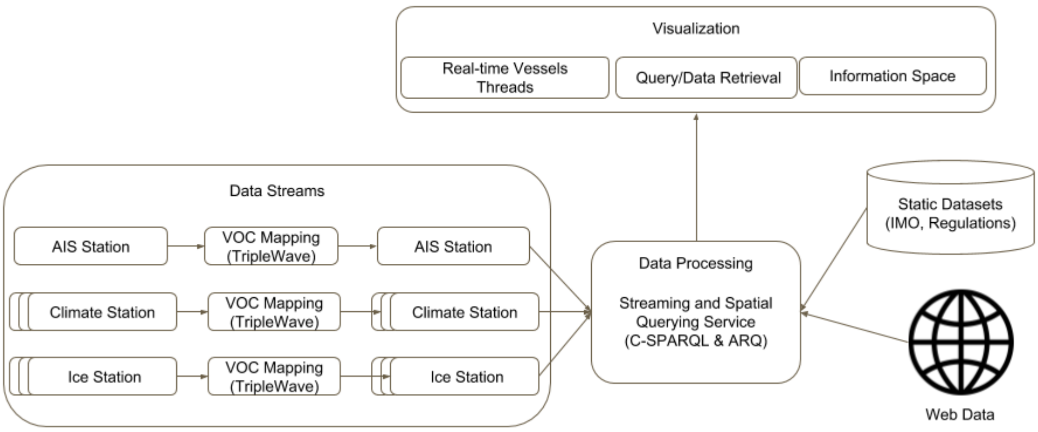

Figure 3: Framework for real-time querying marine events

The IMO data set is used to capture detailed information about the vessels that are being tracked using AIS. Web data is also used to capture more contextualized information, i.e. information about the vessels' country of origin, vessels' type descriptions, and semantic geographical information. The queries results are shown to the user via visualizations. The user can additionally navigate thought the information space, i.e. check the data the system has used so far.

Figure 4 and Figure 5 show C-SPARQL queries used for tracking traffic events as defined in scenario 2. Data streams that capture the current location of the vessels and the current weather conditions are used. The C-SPARQL query in Figure 4 returns, every 9 minutes, the vessels that are currently speeding above 20 knots in a specific area.

The details of the Figure 4 listing are as follows. The prefixes used in the query are declared on lines 1-8. The query runs against the AIS stream. The query considers only the most recent data to be processed which, in this case, are the last 9 minutes of the stream using a sliding window 3 minutes (line 10) in length which cuts the stream in three parts to be processed. The current vessel location and MMSI are captured in lines 13-15 combined with semantic descriptions and vessel characteristics. The FILTER in line 19 assures that the match is only valid for the vessel speeding above 20 knots. Finally, the *spatial:nearby* property is used to capture the vessels which are navigating near the given area (line 20).

```
1: PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#> "+
2: PREFIX prov:<http://www.w3.org/ns/prov#> " +
3: PREFIX sc:<https://schema.org/> " +
4: PREFIX ssn:<http://purl.oclc.org/NET/ssnx/ssn#> "+
5: PREFIX dal:<http://dal.ca/resource/> " +
6: PREFIX vessel:<http://dal.ca/vessel_ontology/> "+
7: PREFIX geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>" +
8: PREFIX spatial: <http://jena.apache.org/spatial#>" +
9: SELECT ?vessel ?long ?lat "+
10: FROM STREAM <http://127.0.0.1:8114/TripleWave-transform/sgraph> [RANGE 9m STEP 3m] " +
11: WHERE { " +
12:      ?vessel ssn:madeobservation ?obs;
13:                  vessel:mmsi ?mmsi;
14:                   geo:long  ?long;
15:                   geo:lat  ?lat.
16:      ?obs ssn:observedProperty dal:Speed_Over_Ground.
17:      ?obs ssn:observationResult ?res.
18:      ?res ssn:hasValue ?speed.
19:       FILTER (?speed > 20.0)
20:      ?vessel  spatial:nearby (1198808.71822814 5136042.06501582 10 'km') }
```

Figure 4: Framework for real-time querying marine events

The C-SPARQL query illustrated in Figure 5 returns, every 10 minutes, the vessels that are currently navigating to areas with severe weather conditions in the last three hours.

The details of the Figure 5 listing are as follows. The prefixes used in the query are declared on lines 1-9. The query runs against two input streams. The first stream is the AIS stream. The query considers only the most recent data to be processed which, in this case, are the last 30 minutes of the stream using a sliding window 10 minutes (line 11) in length which cuts the stream in three parts to be processed. The second stream is a climate stream. Here again the query consider the most recent updates in the last three hours using a sliding one hour window (line 12). The query returns the vessels and their current location, the climate station near them which is observing deteriorating weather conditions, the climate station's temperature and humidity observations, and its locations. Lines 15 – 28 contain the query for all climate stations that are observing low temperature and high humidity. The current location and vessel MMSI are captured in lines 30-33 combined with semantic descriptions and vessel characteristics. Finally, the *spatial:withinCircle* property is used to capture the vessels which are navigating near the climate stations where the weather observations have been made (line 14).

```
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX prov:<http://www.w3.org/ns/prov#>
PREFIX sc:<https://schema.org/>
PREFIX ssn:<http://purl.oclc.org/NET/ssnx/ssn#>
PREFIX dal:<http://dal.ca/resource/>
PREFIX vessel:<http://dal.ca/vessel_ontology/>
PREFIX geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>
PREFIX spatial: <http://jena.apache.org/spatial#>
PREFIX climate: <http://dal.ca/climate_ontology/>
SELECT ?mmsi ?long ?lat ?station ?t ?h ?clong ?clat
FROM STREAM <http://127.0.0.1:8114/TripleWave-transform/sgraph> [RANGE 30m STEP 10m]
FROM STREAM <http://127.0.0.1:8118/TripleWave-transform/sgraph> [RANGE 3h STEP 1h]
WHERE {
  ?station spatial:withinCircle (?lat ?long 100.0 'miles' 3).
    { SELECT ?station (AVG(?temp) AS ?t)  (AVG(?hum) AS ?h)
    WHERE{
           ?station climate:name ?name;
                        geo:long  ?clong;
                      geo:lat  ?clat;
                      ssn:madeobservation ?obs.
          ?obs ssn:observedProperty dal:Temperature;
              ssn:observationResult ?res.
          ?res ssn:hasValue ?temp.
           ?obs1 ssn:observedProperty dal:Huminity;
                 ssn:observationResult ?res1.
            res1 ssn:hasValue ?hum. }
      GROUP BY ?station
      HAVING ((AVG(?temp) < 0.0) && (AVG(?hum) > 50.0))}
      {SELECT ?lat ?long ?mmsi
       WHERE{
              ?vessel vessel:mmsi ?mmsi;
                     geo:long  ?long;
                      geo:lat  ?lat. }}}
```

Figure 5: Framework for real-time querying marine events

5.0 RELATED WORK

In [3], the authors review current research challenges and trends tied to the integration, management, analysis and visualization of vessels addressed in the many fields of computer science. A few Semantic Web approaches have been proposed addressing data representation, integration, and processing. In [9,2,11,18], the authors propose or extend vocabularies, taxonomies, and ontologies to capture the semantics of the low-level data from maritime sensors and thus provide a standard view of the different data sources. By adopting established models, we aim to facilitate integration, linking and reuse across the data sources as well as make data accessible to a wide range of applications.

Instead of proposing a model, in [8] the authors presented an OWL ontology design pattern for oceanographic cruises and showed how this pattern was specified as a combination and reuse of existing patterns: trajectory, event, and information object. Trajectory modeling and semantic enrichment are also proposed by [17,2,13,21]. Comparable to the approach presented here, [17] introduces a system that enables experts in the maritime domain to characterize abnormal ship behavior based on formal semantic properties and rules. The framework presented here goes further by supporting stream querying and processing.

Similar to the work presented here, the work in [2] proposed a model for integration and analysis of data for vessel movement in a RMSA system. The vessel data is heterogeneous and integrated with open party data, (e.g., GeoNames and OpenStreetMap). Also supporting semantic alignment, geospatial fusion and linked-data access to underwater IoT data streams is the work by Correndo et al. [6]. Both approaches support SPARQL to interface the system with the RDF data representation, but no stream-support is provided.

6.0 CONCLUSION

This paper presented a framework for data integration of heterogeneous data streams produced by different kinds of maritime and marine sensors from IoT infrastructures, and for real-time querying and data processing. The approach presented makes use of Semantic Web standards and technologies, and aims to improve knowledge discovery in the maritime domain. Additionally, the approached focuses on the semantic streaming models and engine to process data streams at scale. Three different RMSA forecasting scenarios are presented as a proof of concept. (1) How to detect vessels which are not respecting the speed limit restrictions, (2) how to detect vessels that are approaching heavy weather condition areas, and (3) how to detect vessels that are approaching areas of heavy ice.

Future work includes extending the framework to consider trajectory stream processing and data mining tasks such as trajectory clustering to detect trajectory anomalies (e.g., trajectory appears to be not close to any of the existing cluster) and issue alarms. Another direction for future work includes dealing with noisy, missing, and uncertain information, and consistency checking. This work is part of the MIMIR (Mission-relevant Information Management for Integrated Response) project, which seeks to develop ideas and techniques for seamless integration of data streams of the many IoT sensors infrastructure.

REFERENCES

[1] Barbieri, D.F., Braga, D., Ceri, S., VALLE, E.D. and Grossniklaus, M. (2010), C-SPARQL: a continuous query language for RDF data streams, *International Journal of Semantic Computing*, 4 (01), 3-25.

[2] Brüggemann, S., Bereta, K., Xiao, G. and Koubarakis, M. (2016), Ontology-based data access for maritime security, In *Proceedings of International Semantic Web Conference*, 741-757.

[3] Claramunt, C., Ray, C., Camossi, E., Jousselme, A.-L., Hadzagic, M., Andrienko, G.L., Andrienko, N.V., Theodoridis, Y., Vouros, G.A. and Salmon, L. (2017), Maritime data integration and analysis: recent progress and research challenges, In *Proceedings of EDBT*, 192-197.

[4] Compton, M., Barnaghi, P., Bermudez, L., GarcíA-Castro, R., Corcho, O., Cox, S., Graybeal, J., Hauswirth, M., Henson, C. and Herzog, A. (2012), The SSN ontology of the W3C semantic sensor network incubator group, *Web semantics: science, services and agents on the World Wide Web*, 17, 25-32.

[5] Llaves, A., Corcho, O., Taylor, P. and Taylor, K. (2016), Enabling RDF Stream Processing for Sensor Data Management in the Environmental Domain, *International Journal on Semantic Web and Information Systems (IJSWIS)*, 12 (4), 1-21.

[6] Correndo, G., Crowle, S., Papay, J. and Boniface, M. (2016), Enhancing Marine Industry Risk Management Through Semantic Reconciliation of Underwater IoT Data Streams, In *Proceedings of Proceedings of the 12th International Conference on Semantic Systems*, 161-168.

[7] Jajaga, E. and Ahmedi, L. (2017), C-SWRL: SWRL for Reasoning over Stream Data, In *Proceedings of 2017 IEEE 11th International Conference on Semantic Computing (ICSC),*, 395-400.

[8] Krisnadhi, A.A., Arko, R.A., Carbotte, S., Chandler, C., Cheatham, M., Finin, T.W., Hitzler, P., Janowicz, K., Narock, T.W. and Raymond, L. (2015), Ontology Pattern Modeling for Cross-Repository Data Integration in the Ocean Sciences: The Oceanographic Cruise Example.

[9] Leadbetter, A., Smyth, D., Fuller, R., O'Grady, E. and Shepherd, A. (2016), Where big data meets linked data: applying standard data models to environmental data streams, In *Proceedings of Big Data (Big Data), 2016 IEEE International Conference on*, 2929-2937.

[10] Mauri, A., Calbimonte, J.-P., Dell'Aglio, D., Balduini, M., Brambilla, M., Della Valle, E. and Aberer, K. (2016), Triplewave: Spreading RDF streams on the web, In *Proceedings of International Semantic Web Conference*, 140-149.

[11] Narock, T., Arko, R., Carbotte, S., Krisnadhi, A., Hitzler, P., Cheatham, M., Shepherd, A., Chandler, C., Raymond, L. and Wiebe, P. (2014), The oceanlink project, In *Proceedings of Big Data (Big Data), 2014 IEEE International Conference on*, 14-21.

[12] Okure, A.S. and Pan, J.Z. (2013), Querying EL ontology stream with C-SPARQL, In *Proceedings of Advanced Computer Science Applications and Technologies (ACSAT), 2013 International Conference on*, 337-340.

[13] Parent, C., Spaccapietra, S., Renso, C., Andrienko, G., Andrienko, N., Bogorny, V., Damiani, M.L., Gkoulalas-Divanis, A., Macedo, J. and Pelekis, N. (2013), Semantic trajectories modeling and analysis, *ACM Computing Surveys (CSUR)*, 45 (4), 42.

[14] Le-Phuoc, D., Nguyen-Mau, H.Q., Parreira, J.X. and Hauswirth, M. (2012), A middleware framework for scalable management of linked streams, *Web Semantics: Science, Services and Agents on the World Wide Web*, 16, 42-51.

[15] Rudolph, S., Parsia, B., Hitzler, P., Patel-Schneider, P., and Krötzsch, M. OWL 2 web ontology language primer (second edition). Technical report, W3C, December 2012. http://www.w3.org/TR/2012/REC-owl2- primer-20121211/.

[16] Seaborne, A. and Prud'hommeaux, E. SPARQL query language for RDF. W3C recommendation, W3C, January 2008. http://www.w3.org/TR/2008/REC- rdf-sparql-query-20080115/.

[17] Vandecasteele, A. and Napoli, A. (2012), An enhanced spatial reasoning ontology for maritime anomaly detection, In *Proceedings of System of Systems Engineering (SoSE), 2012 7th*

International Conference on, 1-6.

[18] Xiong, J., Wang, J. and Gao, F. (2013), Research and Application of Ontology-Based Marine Ecology Knowledge Management, In *Informatics and Management Science II*, pp. 465-471, Springer.

[19] Wood, D., Cyganiak, R., and Lanthaler, M. RDF 1.1 Concepts and Abstract Syntax. W3C Recommendation, January 2014. http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/.

[20] Haller, A., Janowicz, K., Cox, S., Le-Phuoc, D., Taylor, K., Lefrancois, M. Semantic Sensor Network Ontology. W3C Recommendation, 19-October-2017. http://www.w3.org/TR/vocab-ssn/

[21] Moreno B., Júnior A.S., Times V., Tedesco P., Matwin S. (2014) Weka-SAT: A Hierarchical Context-Based Inference Engine to Enrich Trajectories with Semantics. In: Sokolova M., van Beek P. (eds) Advances in Artificial Intelligence. AI 2014. Lecture Notes in Computer Science, vol 8436. Springer, Cham

DOCUMENT CONTROL DATA

*Security markings for the title, authors, abstract and keywords must be entered when the document is sensitive

1. ORIGINATOR (Name and address of the organization preparing the document. A DRDC Centre sponsoring a contractor's report, or tasking agency, is entered in Section 8.) IST-160 Specialists' Meeting on Big Data and Artificial Intelligence for Military Decision Making NATO Science and Technology Board, Bordeaux, France	2a. SECURITY MARKING (Overall security marking of the document including special supplemental markings if applicable.) CAN UNCLASSIFIED 2b. CONTROLLED GOODS NON-CONTROLLED GOODS DMC A

3. TITLE (The document title and sub-title as indicated on the title page.)

Semantic Integration of Real-Time Heterogeneous Data Streams for Ocean-related Decision Making

4. AUTHORS (Last name, followed by initials – ranks, titles, etc., not to be used)

Dividino, R.; Soares, A.; Matwin, S.; Isenor, A.W.; Webb, S.; Brousseau, M.

5. DATE OF PUBLICATION (Month and year of publication of document.)	6a. NO. OF PAGES (Total pages, including Annexes, excluding DCD, covering and verso pages.)	6b. NO. OF REFS (Total references cited.)
May 2018	12	21

7. DOCUMENT CATEGORY (e.g., Scientific Report, Contract Report, Scientific Letter.)

External Literature (N)

8. SPONSORING CENTRE (The name and address of the department project office or laboratory sponsoring the research and development.)

DRDC - Atlantic Research Centre
Defence Research and Development Canada
9 Grove Street
P.O. Box 1012
Dartmouth, Nova Scotia B2Y 3Z7
Canada

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.)	9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.)
10a. DRDC PUBLICATION NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document.) DRDC-RDDC-2018-N082	10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.)

11a. FUTURE DISTRIBUTION WITHIN CANADA (Approval for further dissemination of the document. Security classification must also be considered.)

Public release

11b. FUTURE DISTRIBUTION OUTSIDE CANADA (Approval for further dissemination of the document. Security classification must also be considered.)

12. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Use semi-colon as a delimiter.)

Information Collection and Integration; Information Sciences; Information Management; Information Integration

13. ABSTRACT/RÉSUMÉ (When available in the document, the French version of the abstract must be included here.)

Information deluge is a continual issue in today's military environment, creating situations where data is sometimes underutilized or in more extreme cases, not utilized, for the decision-making process. In part, this is due to the continuous volume of incoming data that presently engulf the ashore and afloat operational community. However, better exploitation of these data streams can be realized through information science techniques that focus on the semantics of the incoming stream, to discover information-based alerts that generate knowledge that is only obtainable when considering the totality of the streams. In this paper, we present an agile data architecture for real-time data representation, integration, and querying over a multitude of data streams. These streams, which originate from heterogeneous and spatially distributed sensors from different IoT infrastructures and the public Web, are processed in real-time through the application of Semantic Web Technologies. The approach improves knowledge interoperability, and we apply the framework to the maritime vessel traffic domain to discover real-time traffic alerts by querying and reasoning across the numerous streams. The paper and the provided video demonstrate that the use of standards-based semantic technologies is an effective tool for the maritime big data integration and fusion tasks.