

# Ontology-based Data Access for Maritime Security

Stefan Brüggemann<sup>1</sup>, Konstantina Bereta<sup>2</sup>, Guohui Xiao<sup>3</sup>, and Manolis Koubarakis<sup>2</sup>

<sup>1</sup> Airbus Defence & Space, Germany

<sup>2</sup> National and Kapodistrian University of Athens, Greece

<sup>3</sup> Free University of Bozen-Bolzano, Italy

**Abstract.** The maritime security domain is challenged by a number of data analysis needs focusing on increasing the maritime situation awareness, i.e., detection and analysis of abnormal vessel behaviors and suspicious vessel movements. The need for efficient processing of dynamic and/or static vessel data that come from different heterogeneous sources is emerged. In this paper we describe how we address the challenge of combining and processing real-time and static data from different sources using ontology-based data access techniques, and we explain how the application of semantic web technologies increases the value of data and improves the processing workflow in the maritime domain.

**Keywords:** Maritime Security, EMSec, RMSAS, Geospatial, OBDA, *Ontop*, SPARQL, R2RML, OWL

## 1 Introduction

The maritime security domain is challenged by a number of data analysis needs with a focus on increasing the maritime situation awareness. Vessel movements are of major importance for maritime data analysts and decision makers. Abnormal vessel behaviors and suspicious vessel movements need to be detected and understood to properly increase the maritime domain awareness. The project EMSec<sup>4</sup> (real-time services for the maritime security) has the aim to support the maritime security by improving the availability and accessibility of relevant data and information ashore and offshore. The central data management component of EMSec is the “Real-time Maritime Situation Awareness System” (RMSAS), which is in charge of integrating various types of data from different sources.

This paper focuses on the integration and analysis of data for vessel movements in RMSAS. For the proper analysis a storage capability is needed to properly analyze vessel data and to identify vessel trajectories and their stops and movements for the last days. However, the vessel data integration and management task in RMSAS is challenged by the following requirements: (a) The vessel data is heterogeneous and in particular consists of dynamic position data or static metadata. (b) There is a need for integrating third party data, i.e., open data like GeoNames and OpenStreetMap. (c) The size of the data is large, deriving from the acquisition and processing of large radar and satellite

---

<sup>4</sup> <http://www.emaritime.de/projects/emsec/>

images. (d) The data about vessels are produced in real-time, i.e., approximately 1000 vessel positions are acquired per second.

The motivation of this work was to address the above needs for combining data from heterogeneous sources, such as in-situ data, AIS data, and open data developing an automated solution avoiding manual work as much as possible. We considered that the conceptualized model offered by ontologies would meet the requirements of that purpose, so we used state-of-the-art technologies and tools into this direction. The rationale behind our design choices was to avoid the creation of replicas of the same data in other formats and also to avoid storing natively data that are already available as open data. For the first, we opted for a well-known ontology-based data access (OBDA) system, *Ontop*, instead of using a triple store, to avoid the cost (in disk space and response time) of materializing our frequently updated data to RDF and storing them natively. For the second, we used federation to query data coming from different endpoints (e.g., in-house data and linked data like geonames). In the same direction, we used the web-based tool *Sextant* to visualize geospatial data coming from different SPARQL endpoints (*Ontop* endpoints, triple stores, etc.) creating composite maps, instead of storing everything natively in a geospatial relational database and visualizing them using GIS tools.

The RMSAS system has been implemented on top of these state-of-the-art Semantic Web techniques and tools. The evaluation has shown that RMSAS eases the data analysis by using virtual triples and standardized vocabularies. Next, the integration of several heterogeneous data sources is a benefit for maritime decision makers and the maritime security. Finally, the approach contributes significantly to the detection of routine traffic and abnormal vessel behavior.

The rest of the paper is structured as follows: Section 2 describes the data sources that are available in the German maritime research project EMSec and focus on the data integration requirements. Section 3 defines a maritime domain ontology. Section 4 shows how to properly analyze these data using semantic technologies, focusing on how the concept of ontology based data access can be used to add a level of semantics on these data and how to access and spatially analyze relational data. Section 5 shows how OBDA approach can add a value to the analysis of vessel movements. Section 6 concludes the paper with a summary and discussion.

## 2 The Maritime Context

The maritime security is currently challenged by several influences: More than two-thirds of the overall volume of cargo worldwide is transported seaborne. This massively increases the number of ships traveling on the seas. Next, the continuously increasing number of offshore wind parks has an impact on the security of the citizens. The energy supply must be ensured even without fossil fuels. This turns offshore wind park into assets with a strong demand for protection. Moreover, industrial nations worldwide use the potential of the seas, but are threatened by pirates and terrorists. Finally, besides the danger of criminal actions, disasters and storm floods also challenge the maritime security [4].

In Germany, it is the task of several institutions from the federal government and the federal states to ensure the maritime security and to mitigate risks. For that reason, several

detailed information are needed to effectively gain maritime domain awareness [11] and to analyze maritime emergency situations. Data from different data sources are needed to support these tasks and to create a common overview of the maritime situation.

The German federal ministry of education and research funded project EMSec (real-time services for the maritime security) is aiming at providing a consistent and user-oriented access to data and information from different data sources. These data can be satellite images, aerial images, weather information like wind, rain, drift, or else. All these data shall be displayed in a flexible manner to ensure the situation awareness of the end users, combining data from several sources. A faster and more detailed provision of these data shall enable responsible organizations and decision makers to early recognize and avoid critical situations. In an emergency situation or within a criminal activity, action forces can benefit from accessing detailed information in real time to handle the situation efficiently.

## 2.1 RMSAS: Real-time Maritime Situation Awareness System

As previously mentioned, a system is needed inside EMSec that is capable of integrating data coming from several data sources. Since the concrete information needs of the user are not known and are highly situation-dependent, a flexible system and an agile iterative design approach is needed. A distributed federated system is needed to face the diversity of the maritime players and governmental constraints like laws, IT-security, or else. A system of systems approach has been identified as being able to cope with these challenges in the maritime domain [12]. Consequently, with RMSAS, a real-time maritime situation awareness system is implemented in EMSec as a system of systems to:

- integrate vessel data coming from various sensors,
- enrich these data with data from other sources (e.g. open data),
- harmonize these data using established maritime standards,
- retrieve new information from these integrated data,
- infer knowledge from this information,
- retrieve and deliver this knowledge in near real-time,
- create a maritime domain awareness for the end user, and
- enable maritime decision makers to handle maritime situations more efficiently.

A service oriented architecture using semantic web technologies has been evaluated to support these tasks. The approach presented in [12] will be used and extended to challenge IT-security constraints.

Situation-aware data shall be presented to the user in near real-time. To achieve this, data have to be integrated and consolidated from several different sources. This allows for properly displaying the combination of this data to the user via an application. Providing data faster and in further detail shall allow the involved parties to identify critical situations better and earlier, to avoid these situations, and to manage them efficiently.

Being the central data management component in EMSec, the RMSAS aims at integrating and consolidating data. RMSAS uses the “System of Systems” approach and

implements a federated information system based on separate services (SOA). Data are integrated in RMSAS in near real-time, next they are consolidated based on semantic data models and techniques and provided to the end user as information products. Ontologies are used in the consolidation of these heterogeneous data.

## 2.2 Data Sources

In the following we describe some typical data sources that are used in EMSec. The Automatic Identification System (AIS)<sup>5</sup> is a common data source for the maritime navigation used worldwide. These data serve as a basis for several maritime applications and also serve as reference data in EMSec. AIS is a cooperative system and calls for the active participation of every vessel. However, it has only a limited trustworthiness due to the fact that every vessel owner can manipulate the system or can completely switch it off. Transferring and receiving the data can also be manipulated or hindered. Terrestrial AIS uses coast based receiver with good availability but with limited range and coverage. Satellite based AIS is also available, but with limited good reliability these days. EMSec analyzes the benefits of additional data sources that are based on Earth Observation from satellites and airborne systems. Here it is important to extend the spatial and temporal resolution of maritime data to add a value to the maritime security.

The EMSec partners provide the following data for the integration in RMSAS:

**AIS.** AIS messages are not only used as reference data, they are additionally used for quality inspections and – wherever possible – analyzed to identify certain movement pattern, for instance for ferries. Several AIS types are available: Terrestrial AIS, satellite AIS<sup>6</sup>, and the AIS signal that comes from the Columbus-module of the ISS<sup>7</sup>. In EMSec we receive AIS data about 800-1000 vessels in the German bay every 1-3 seconds.

**Satellite SAR.** TerraSAR-X provides satellite-based synthetic aperture radar (SAR) and creates radar images with a high resolution. Algorithms can be used to detect objects (e.g., vessels) and to link these detected objects with previously collected AIS messages. The radar images can also be analyzed to extract wind and wave information and connect them with conventional secondary weather information.

**Airborne systems.** The EMSec consortium utilizes an airplane that comes with an AIS receiver and a radar system. The AIS messages are used as described before. The radar system provides objects and their movements as plots and tracks. Next, another airborne system provides optical images that are used to detect vessels in these images. RMSAS is capable of providing these object detections together with weather information and geospatial information to the end user applications.

---

<sup>5</sup> [https://de.wikipedia.org/wiki/Automatic\\_Identifier\\_System](https://de.wikipedia.org/wiki/Automatic_Identifier_System)

<sup>6</sup> [http://www.esa.int/Our\\_Activities/Space\\_Engineering\\_Technology/ESA\\_satellite\\_receiver\\_brings\\_worldwide\\_sea\\_traffic\\_tracking\\_within\\_reach](http://www.esa.int/Our_Activities/Space_Engineering_Technology/ESA_satellite_receiver_brings_worldwide_sea_traffic_tracking_within_reach)

<sup>7</sup> [http://www.esa.int/Our\\_Activities/Space\\_Engineering\\_Technology/Space\\_Station\\_keeps\\_watch\\_on\\_world\\_s\\_sea\\_traffic](http://www.esa.int/Our_Activities/Space_Engineering_Technology/Space_Station_keeps_watch_on_world_s_sea_traffic)

### 2.3 Request management in EMSec

This section describes the concept of managing and answering requests in EMSec. Figure 1 describes that these requests may come from a user, a SOA-architecture or else. The main concept is that requests are formulated using the Top level ontology (TLO) and are posed using SPARQL. This enables the end user to use the described high level semantics of the TLO. The semantic data processing component utilizes *Ontop* to translate the queries to SQL queries in order to be evaluated in the underlying RDBMS, e.g., a PostgreSQL database. This paper focuses on relational data sources, so other input data formats such as CSV will not be discussed here.

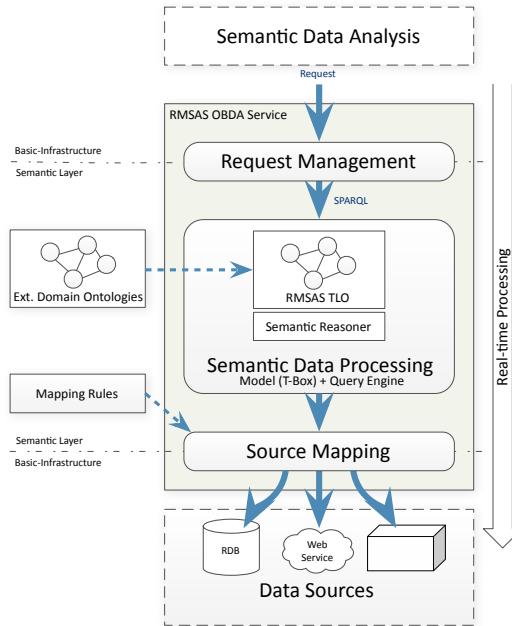


Fig. 1: Request Management within EMSec

### 2.4 Scenarios in EMSec

The validation of the created methods, architectures, algorithms, and concepts will be done in a campaign, where two maritime security scenarios are executed. First, a concrete satellite mission is utilized. Second, both airborne missions are requested and executed. The generated data are transferred to RMSAS in near real time and integrated, analyzed, consolidated and finally transferred to the user. Several maritime regions are deserving protection. Restricted areas can be off-shore platforms, wind parks, or preserved areas.

These call for limited vessel traffic with certain restrictions. *Geographic fences* can be created to analyze the vessel traffic focusing in specific areas of interest. Possible scenarios are to check that the speed over ground is within a limited range in these regions, that certain vessel types like oil-tanker may not pass these regions, or that under certain sea conditions no vessel traffic is allowed.

## 2.5 Categorization of data

RMSAS processes several types of data. These data can be categorized as follows:

**Data streams.** Data streams are data which are created continuously in real-time. These data streams can be AIS-data that continuously report new arriving vessels in the German bay.

**Static data.** Data are static when they are stored in databases, on FTP-servers, or in external systems. These can be metadata about vessels as for example the vessel type, cargo, port of departure, and historical data about previous routes. After transmission to the earth, satellite data are made available as packages.

**Open data.** Open data are data coming from the linked open data cloud or from other external data sources. Using these data can improve certain kinds of analysis. For example, these data sources contain information about real existing harbors (as in GeoNames<sup>8</sup>), or information about certain points of interest (as in DBpedia<sup>9</sup> or OpenStreetMap<sup>10</sup>), or that contain weather data (as in OpenWeatherMap<sup>11</sup>).

GeoNames is a gazetteer that collects both spatial and thematic information for various place names around the world. GeoNames data is available through various Web services but it is also published as linked data. The features in GeoNames are interlinked with each other defining regions that are inside the underlined feature (children), neighboring countries (neighbors) or features that have certain distance with the underlined feature (nearby features).

OpenStreetMap (OSM) maintains a global editable map that depends on users to provide the information needed for its improvement and evolution. OpenStreetMap datasets are available in RDF format from the LinkedGeoData project<sup>12</sup>. However, it was more convenient for us to download the most up-to-date original OpenStreetMap data about Bremen, available as Shapefiles<sup>13</sup>. We imported the Shapefiles into a PostGIS database and created virtual geospatial RDF views on top of them using *Ontop-spatial*, as described at <https://github.com/ConstantB/ontop-spatial/wiki/Shapefiles>.

---

<sup>8</sup> <http://geonames.org>

<sup>9</sup> <http://dbpedia.org>

<sup>10</sup> <http://openstreetmap.org/>

<sup>11</sup> <http://openweathermap.org/>

<sup>12</sup> <http://linkedgeodata.org>

<sup>13</sup> <http://download.geofabrik.de/europe/germany/bremen.html>

### 3 Ontologies for the Maritime Domain

In this section we present the central knowledge representation and data management in maritime component of RMSAS. It includes a model of the maritime domain, the logical data model and the ontology. The last part of the section is devoted to interlinking the RMSAS data with the linked open data (LOD) cloud.

#### 3.1 Modeling the maritime domain

Maritime domain models are results of several research projects, both national and international. The CoopP-project has created the CISE-ontology [1], which is reused in RMSAS and adopted to meet the project's specific requirements.

**Object.** Objects can be any involved parts of the maritime domain. They can be physical elements that are airborne, onshore and offshore, such as vessels, containers, planes, icebergs, or satellites. Vessels are central elements of interest and modeled in greatest detail with a special focus on the information that are available in AIS.

**Geometry.** Geometry is dedicated to deal with information about space and geographical localizations of the maritime objects. The geometries contained in our data are encoded in WKT format, which is an OGC standard for the serialization of geometries. The geometries encountered in our dataset are mainly polygons and points. The geometries of areas, for example, are represented as polygons. These areas can be marked regions ashore, for instance. Dimension describes the specifics of an object like length, width, or height. A location describes places with a geographical name like cities or harbours. They can be identified using a URI which makes it possible to interlink them with external sources like GeoNames or DBpedia. Movements describe the track of an object including its course and speed over ground and optionally its rate of turn. Points describe a dedicated geographical point described using its geographical coordinates and its height. A position then is a point combined with a timestamp.

**Time.** Time is used to describe timestamps that can be used to model positions of objects, to label data during data integration and to support temporal data analysis.

#### 3.2 The RMSAS movement ontology

In order to model our data, we have constructed an ontology that is shown in figure 2. In this paper we focus on the aspects of vessel movements and trajectories.

The movement ontology defines the necessary structures for modeling object movements like vessel, satellites or aircrafts. The ontology allows for enriching native position data with semantics. This allows to model vessel positions as being moves or stops. Any moving object has position data and consists of trajectories that reflect the historic positions of an object. The use of semantics to these positions facilitates the monitoring of the status of the moving object, i.e. whether it has stopped or was moving.

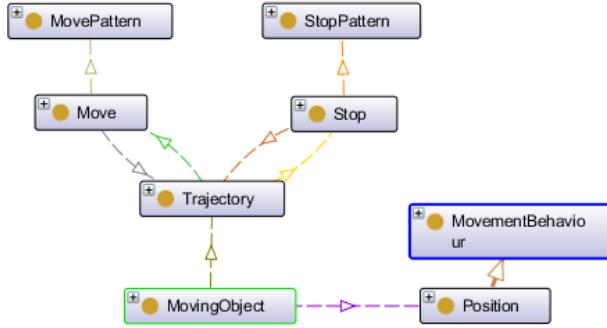


Fig. 2: RMSAS Movement ontology

## 4 Semantic Data Analysis

In this section we describe how RMSAS uses the Semantic Web technologies mentioned in the introduction in order to achieve the following goals:

- Transparent integration of different, geospatial and thematic data sources using ontologies.
- Processing of in-house dynamic and static data, enriching them with information already available on the web (linked open data).
- Avoid replicating the same data as much as possible (e.g., materializing data to RDF, storing data from scratch when a SPARQL endpoint for them is already available) using OBDA techniques and federation.
- Visualization of the data and creation of persistent, web accessible maps, with no need to load the datasets or issue the queries again every time we want to populate the existing databases/endpoints with fresh data.

We illustrate the abstract architecture of RMSAS in Figure 3. RMSAS uses the OBDA system *Ontop* and *Ontop-spatial* to expose the data we need from the relational databases as SPARQL endpoints. For accessing non-relational data sources, RMSAS first wraps these sources into relational ones by Teiid, and then uses *Ontop* [8,5] to access them. For federating third party SPARQL endpoints like GeoNames, Sesame is used for the SPARQL 1.1 federated query answering. Finally, Sextant is used for visualizing the results on temporally-enabled maps combining geospatial and temporal results from different (Geo)-SPARQL endpoints.

### 4.1 Linking RMSAS data to the RMSAS ontology

The relational data in RMSAS can be faithfully mapped to the ontology using the ontology-based data access (OBDA) approach. We use *Ontop* and its extension *Ontop-spatial* for this purpose. As illustrated in Figure 3, *Ontop* allows for querying relational data sources through a conceptual representation of the domain of interest, provided in terms of an ontology, to which the data sources are mapped. *Ontop* answers

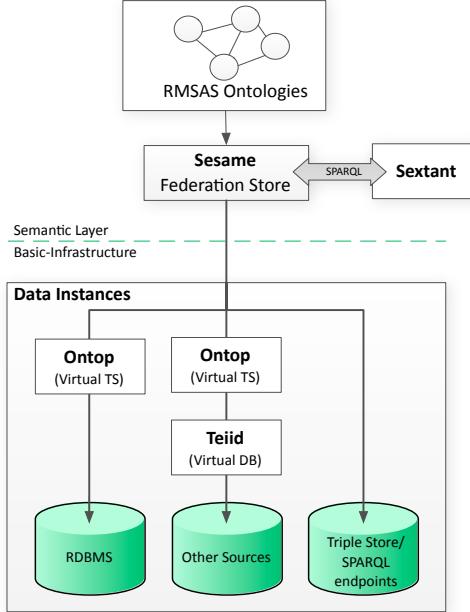


Fig. 3: Abstract Architecture of RMSAS

the SPARQL queries by translating them into SQL queries over the database and avoids materializing triples. *Ontop-spatial* is an extension of *Ontop* with geospatial features.

*Ontop* uses declarative mappings to encode how relational data are mapped to the respective RDF terms. *Ontop* supports W3C R2RML mapping language [6] and its native *Ontop* mapping languages. In this paper, we use the native syntax because it is more compact. An *Ontop* mapping consists of three fields: *mappingId*, *source* and *target*. The *mappingId* is an identifier for mapping; the *source* is an arbitrary SQL query over the database; and the *target* is a triple template written in Turtle syntax that contains placeholders referencing column names mentioned in the source query.

For example, all information about the positions of vessels are stored in a spatially enabled PostGIS database. In Figure 4, we present mappings related to vessels in *Ontop* native syntax. The coordinates of the mappings that are stored in the respective columns named `longitude` and `latitude` in the database in textual form are mapped into RDF literals, as objects of the respective virtual triples as indicated in the mapping assertion with *mappingId* “Position”. The respective geometries that represent the vessels positions are also stored in the well-known binary format (WKB) in a separate column, named `geom`. The mapping assertion `Geometry` indicates how this information is mapped to RDF: The binary geometry of the database is exported as a well-known text literal (WKT), following the OGC GeoSPARQL standard [2].

---

```

mappingId Vessel
target      :Vessel-{v.id} a :Vessel ; :hasName {v.name} .
source      SELECT v.id, v.name FROM Vessel v

mappingId VesselPosition
target      :Vessel-{v.id} :hasLocation :Position-{vp.position_id} .
source      SELECT v.id, vp.position_id FROM Vessel v,
                           Vessel_position vp WHERE v.id=vp.vessel_id

mappingId Position
target      :Position-{id} a :Position ; :hasLatitude {latitude} ;
                           :hasLongitude {longitude} ; :hasDateTime {ts} ;
                           :hasGeometry geos:Geometry-{id} .
source      SELECT id, latitude, longitude, ts FROM position

mappingId Geometry
target      geos:Geometry-{id} a geos:Geometry ;
                           geos:asWKT {geom}^^geos:wktLiteral .
source      SELECT id, geom FROM position

```

---

Fig. 4: Example mappings in RMSAS

#### 4.2 SPARQL queries

In the following we present two example SPARQL queries that we used in order to process our data using OBDA technologies and combine them with other sources.

The query described in Figure 5 retrieves geometries of the locations of vessels (ordered by the timestamps) that are stored in binary (WKB) format in the relational database. Objects of this datatype are internally handled by *Ontop-spatial* and are eventually transformed into RDF literals of WKT datatype, as specified the OGC standard GeoSPARQL and indicated by the mappings that we presented in the previous section. This is the template of the queries we posed to retrieve the locations of ferries to three German islands (Langeoog, Spiekeroog, and Wangerooge). Figure 8(a) presents the visualization of results using Sextant.

---

```

PREFIX : <http://www.rmsas.de/DMARitime#>
PREFIX geos: <http://www.opengis.net/ont/geosparql#>
SELECT DISTINCT ?x ?z ?g ?timestamp
WHERE {
  ?x rdf:type :Vessel. ?x :hasName "Vesselname"^^xsd:string.
  ?x :hasLocation ?z. ?z :hasDateTime ?timestamp .
  ?z geos:asWKT ?g. }
ORDER BY DESC(?timestamp)

```

---

Fig. 5: SPARQL query retrieving positions of a vessel through time

---

```

PREFIX geof: <http://www.opengis.net/def/function/geosparql/>
PREFIX osm: <http://linkedgeodata.org/ontology#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
SELECT DISTINCT ?lu ?geo WHERE {
  ?x osm:landUse lgd:port . ?x geo:asWKT ?geo .
  ?x1 geo:asWKT ?geo1 . ?x1 osm:landUse ?lu .
  FILTER (geof:sfIntersects(?geo,?geo1))}
```

---

Fig. 6: SPARQL query retrieving locations of ports and land use of intersecting areas

---

```

PREFIX : <http://www.rmsas.de/DMARitime#>
PREFIX geos: <http://www.opengis.net/ont/geosparql#>
SELECT ?vessel ?location ?geometry ?wkt ?mmssi ?length ?height
WHERE {
  SERVICE <http://www.rmsas.de/openrdf-sesame/PositionStore> {
    ?vessel rdf:type :Vessel .
    ?vessel :hasLocation ?location. ?vessel :hasName "388328333".
    ?location :hasGeometry ?geometry. ?geometry geos:asWKT ?wkt.
    OPTIONAL {
      SERVICE <http://www.rmsas.de/openrdf-sesame/ObjectStore> {
        ?vessel :hasMMSI ?mmssi ; :hasName "388328333".
        ?vessel :hasLength ?length ; :hasHeight ?height .}}}}
```

---

Fig. 7: SPARQL Federation: finding locations of a vessel and their static metadata

The query described in Figure 6 retrieves the geometries that represent the locations of ports and the land use of areas that they intersect with (e.g., farmyards, commercial/religious areas).

#### 4.3 SPARQL federation

For federating third party SPARQL endpoints like GeoNames, RMSAS relies on the SPARQL 1.1 federated query [10] implemented in Sesame [3]. In the query described in Figure 7, we use “SERVICE” function in order to combine information coming from different endpoints exposed by *Ontop*. The first endpoint (PositionStore) contains dynamic data about the locations of vessels stored in a PostGIS database. The second endpoint (ObjectStore) contains static metadata about vessels, such as dimensions, name, etc. The query retrieves all available information about a specific vessel combining both *Ontop* endpoints in a federated store.

#### 4.4 Visualization of results

For the visualization of the geospatial results presented above, we used the tool Sextant [9]<sup>14</sup>, which is a web based and mobile ready platform for visualizing, exploring and

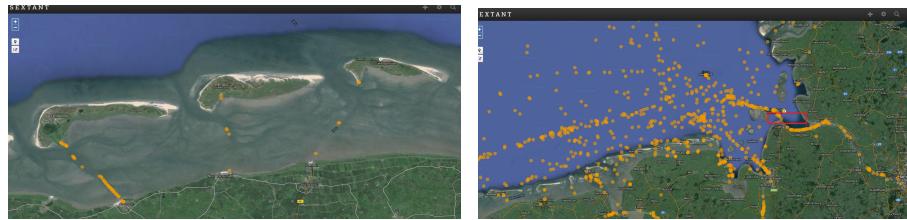
---

<sup>14</sup> <http://sextant.di.uoa.gr/>

interacting with linked geospatial data. Sextant is mainly used to create thematic maps by combining geospatial and temporal information that exists in a number of heterogeneous data sources ranging from standard SPARQL endpoints, to SPARQL endpoints following the OGC standard GeoSPARQL, or well-adopted geospatial file formats, like KML, GML and GeoTIFF.

More specifically, we use the capabilities of Sextant to issue queries to remote GeoSPARQL endpoints and project geometries that are included in the result set on a map. Every layer on that map corresponds to results (i.e., geometries) retrieved from a SPARQL or GeoSPARQL query. By this way, we can combine and visualize different geospatial sources. Another Sextant capability that is useful in this use-case is the *timeline* capability. As the location of vessels is associated with a timestamp, the results of a GeoSPARQL query that retrieves both the geo-location of the vessel and the timestamp can be visualized in both the map and the timeline respectively; as the user scrolls the timeline band of Sextant, they can see where the vessel was at that time. The temporal features of Sextant are described at [9].

As *Ontop* can be used as a standard SPARQL endpoint, thus, *Ontop-spatial* can be used as a GeoSPARQL endpoint, so we used *Ontop-spatial* endpoints as some of the source endpoints of Sextant. Then, we posed geospatial queries like the ones described



(a) Identification of three ferries to three German islands, Langeoog, Spiekeroog, Wangerooge. (b) The red polygon represents a geofence that observes all incoming and leaving vessels via the river Elbe

Fig. 8: Utilization of Sextant to display WKT-data that is made available by Ontop-spatial

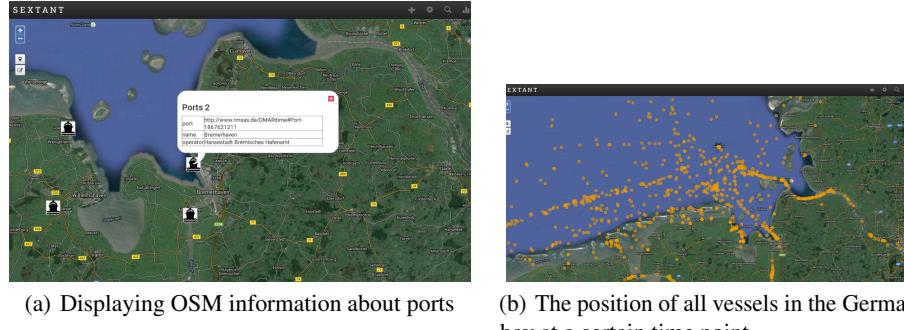


Fig. 9: Ports and vessels visualized in Sextant

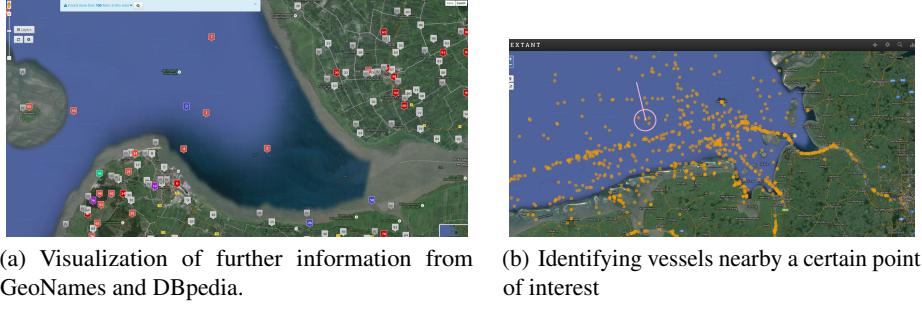


Fig. 10: GeoNames, DBpedia, and GeoSPARQL are used with Sextant

in the previous section and the geometries that were included in the result set of each query were displayed on the map, creating one layer for each one of the geospatial queries posed. Screen shots of the query results visualized using Sextant are provided in Figure 8, 9 and 10.

## 5 Evaluation

In the context of the project EMSec we have developed an approach for integrating data that comes from various data sources in the RMSAS system. The main focus in this work was on analyzing data about vessels. The benefits of the approach that we presented in this paper are explained below.

**Improved data analysis using virtual triples.** The data given in this project mainly exists in databases and data streams and is modeled with respect to different data models. The use of OBDA techniques facilitates the process of data analysis as these data are mapped to the ontology that has been created for RMSAS. This allows decision makers to formulate queries against a standardized ontology instead of articulating different queries in different languages against different data sources like the ones described in Section 2.2.

**Benefits of data integration for maritime decision makers.** Compared to the old workflow with respect to information exchange and integration that was identified in the beginning of the EMSec project, where maritime staff had to exchange data often in very traditional ways like email, USB-sticks, mail, paper, or else, the current workflow is significantly improved. With the presented technologies in place, maritime decision makers have all the desired information at hand in near real-time, integrated from different data sources. This increasing having an overview on the maritime security and having a better maritime situational awareness.

**Detection of routine traffic and abnormal vessel behavior.** In the process on data analysis, SPARQL queries and SWRL rules [7] have been used as a good means (w.r.t. expressivity and efficiency) to detect routine traffic and abnormal vessel behavior. Since we cannot display these rules here for confidentiality reasons, we can state that vessel

movements can be easily classified using the movement ontologies that were introduced in section 3.2 and that vessel behavior can be classified using the introduced movement pattern. Having combined this with the OBDA approach and with the utilization of (Geo-)SPARQL functionalities, this has strong benefits regarding the detection of routine traffic and abnormal vessel behavior.

## 6 Summary, Lessons Learned and Future Work

### 6.1 Summary

In this paper we described challenges for the maritime security and how a German project named EMSSec addresses these challenges by introducing a system called RMSAS. This system has been developed to incorporate data from different data sources and in different formats. An ontology was introduced to model the maritime domain and to provide a common view on the different data sources. The concept of ontology based data access was used to map the original data to this ontology. This allows maritime decision makers to efficiently pose queries against the high level ontology. We have then shown how these queries are articulated by the *Ontop* framework and how they are translated to SPARQL.

We have further introduced the movement ontology as a concrete use case on identifying and analyzing vessel movements and detecting abnormal vessel behavior. Open geospatial data like OpenStreetMap and Geonames have been used to put vessel data and information regarding ports in relation to the publicly available data. This allowed for a comparison of the data acquired about vessels with the data that have been publicly created and reviewed.

### 6.2 Lessons learned

Looking back the design and implementation of RMSAS which heavily replies on OBDA techniques for *virtual* data integration, it is clearly that these new techniques introduced some learning curve for developers. However, we emphasize that this curve would be bigger if we had to use ETL tools for converting and storing RDF data or if we did this without Semantic Web technologies since relational databases do not offer the conceptual model for data integration that RDF/OWL could offer.

Regarding the new languages and tools in RMSAS, at this moment, mapping designers are comfortable with the *Ontop* mapping language since it is easy to understand and write. Meanwhile a more friendly GUI tool for assisting mapping construction would still be helpful for improving productivity. *Ontop* Sesame workbench comes with a basic and simple interface for configuring SPARQL endpoints; however it currently lacks functionalities like version control or more fine-grained configurations.

Another limitation of the Semantic Web technologies that were used in this use case was the lack of geospatial federation support. Although *Ontop-spatial* supports spatial filters in queryless, federated queries that perform spatial joins spanning different geospatial endpoints are not supported in any federated system to the best of our knowledge. Sextant is a user-friendly tool which provides another approach of data integrating by visualization of open linked geospatial data and its comparison with in-situ data. We would also like to be able to pose federated geospatial queries and project their results in Sextant but due to the reasons described above, this is not supported yet.

### 6.3 Future work

The RMSAS system is already scheduled to be deployed this autumn outside of the lab according to the project management and there is no technical barrier for the deployment. Future work will focus on clustering abnormal vessel behavior and the creation of early pre-warning systems. This would allow to concentrate the situational awareness to potentially conspicuous vessels. Another field of future work is to manage uncertainties that are due to missing or inconsistent data and where rules like SWRL rules may fail. We will also carry out an extensive performance evaluation.

**Acknowledgement.** Brüggemann is supported by the German BMBF project EMSSec; Bereta and Xiao are supported by the EU under the large-scale integrating project (IP) Optique (Scalable End-user Access to Big Data), grant agreement n. FP7-318338.

## References

1. Towards the integration of maritime surveillance: A common information sharing environment for the EU maritime domain. Technical report, 2009. Available at <http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=URISERV:pe0011&from=EN>.
2. Open Geospatial Consortium. OGC GeoSPARQL - A geographic query language for RDF data. OGC Candidate Implementation Standard, 02 2012. Available at <http://www.opengeospatial.org/standards/geosparql>.
3. J. Broekstra, A. Kampman, and F. van Harmelen. Sesame: A generic architecture for storing and querying RDF and RDF schema. volume 2342 of *LNCS*, pages 54–68. Springer, 2002.
4. S. Brueggemann and S. Foerster. User requirements for real-time services for the maritime security. *Proceedings of ISIS 2014: International Symposium on Information on Ships*, 2014.
5. D. Calvanese, B. Cogrel, S. Komla-Ebri, R. Kontchakov, D. Lanti, M. Rezk, M. Rodriguez-Muro, and G. Xiao. Ontop: Answering SPARQL queries over relational databases. *Semantic Web Journal*, 2016. (to appear).
6. S. Das, S. Sundara, and R. Cyganiak. R2RML: RDB to RDF mapping language. W3C Recommendation, World Wide Web Consortium, 2012. Available at <http://www.w3.org/TR/r2rml/>.
7. I. Horrocks, P. Patel-Schneider, H. Boley, S. Tabet, B. Grosof, and M. Dean. SWRL: A semantic web rule language combining OWL and RuleML. W3C Member Submission, World Wide Web Consortium, 2004.
8. R. Kontchakov, M. Rezk, M. Rodriguez-Muro, G. Xiao, and M. Zakharyashev. Answering SPARQL queries over databases under OWL 2 QL entailment regime. In *Proc. of the 13th Int. Semantic Web Conf. (ISWC)*, LNCS, pages 552–567. Springer, 2014.
9. C. Nikolaou, K. Dogani, K. Bereta, G. Garbis, M. Karpathiotakis, K. Kyziros, and M. Koubarakis. Sextant: Visualizing time-evolving linked geospatial data. *Web Semantics: Science, Services and Agents on the World Wide Web*, 35, Part 1:35 – 52, 2015. Geospatial Semantics.
10. E. Prud'hommeaux and C. Bui-Aranda. SPARQL 1.1 Federated query. W3C Recommendation, World Wide Web Consortium, Mar. 2013. Available at [www.w3.org/TR/sparql11-federated-query/](http://www.w3.org/TR/sparql11-federated-query/).
11. G. Vance and P. Vicente. Maritime Domain Awareness. *Proceedings of the Marine Safety & Security Council*, 63(3), 2006.
12. N. Willems, W. Robert, V. Hage, and G. D. Vries. An integrated approach for visual analysis of a multi-source moving objects knowledge base. *International Journal of Geographical Information Science*, 24(9):1–16, 2010.