

# Enhancing Marine Industry Risk Management Through Semantic Reconciliation of Underwater IoT Data Streams

Gianluca Correndo

IT Innovation Centre, University of  
Southampton

Gamma House, Enterprise Road  
SO16 7NS, Southampton, UK

gc@it-innovation.soton.ac.uk

Simon Crowle

IT Innovation Centre, University of  
Southampton

Gamma House, Enterprise Road  
SO16 7NS, Southampton, UK

sgc@it-innovation.soton.ac.uk

Juri Papay

IT Innovation Centre, University of  
Southampton

Gamma House, Enterprise Road  
SO16 7NS, Southampton, UK

jp@it-innovation.soton.ac.uk

Michael Boniface

IT Innovation Centre, University of  
Southampton

Gamma House, Enterprise Road  
SO16 7NS, Southampton, UK

mjb@it-innovation.soton.ac.uk

## ABSTRACT

The “Rio+20” United Nations Conference on Sustainable Development (UNCSD) focused on the “Green economy” as the main concept to fight poverty and achieve a sustainable way to feed the planet. For coastal countries, this concept translates into “Blue economy”, the sustainable exploitation of marine environments to fulfill humanity needs for resources, energy, and food. This puts a stress on marine industries to better articulate their processes to gain and share knowledge of different marine habitats, and to reevaluate the data value chains established in the past and to support a data fueled market that is going only to in the near future.

The EXPOSURES project is working in conjunction with the SUNRISE project to establish a new marine information ecosystem and demonstrate how the ‘Internet of Things’ (IoT) can be exploited for marine applications. In particular EXPOSURES engaged with the community of stakeholders in order to identify a new data value chain which includes IoT data providers, data analysts, and harbor authorities. Moreover we integrated the key technological assets that couple OGC standards for raster data management and manipulation and semantic technologies to better manage data assets.

This paper presents the identified data value chain along with the use cases for validating it, and the system developed to semantically reconcile and manage such data collections.

## CCS Concepts

H.2.8 [Database Management]: Database Applications – *Spatial databases and GIS, Scientific databases*; E.2 [Data Storage

SAMPLE: Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SEMANTICS’16, September 15–16, 2016, Leipzig, Germany.

Copyright 2016 ACM 1-58113-000-0/00/0010 ...\$15.00.

DOI: <http://dx.doi.org/10.1145/12345.67890>

**Representations**]: Linked representations; H.3.5 [Online Information Services] Web-based services.

## Keywords

Marine industry; data value chain; linked data, IoT.

## 1. INTRODUCTION

From transport, energy and fishing to high growth marine sectors of seabed mining, aquaculture and offshore energy, emerging information ecosystems will transform utilization and sustainability of ocean resources. Oceans are challenging environments to construct and operate industrial assets. Foundations of offshore structures are vulnerable to localized scour whilst busy navigation channels are subject to ongoing sediment transport. Natural hazards are amplified by extreme weather events and the impact of climate change. Maintaining knowledge about the complex interplay between human activity and micro/macro ecological processes is essential for optimization of periodic maintenance, response to crisis or extreme weather events, assessment of environment impact and license compliance.

Sensor observation from IoT infrastructures, information services and decision support are important measures to monitor and predict natural processes for evaluation of socio-economic and environmental risk. Nevertheless, the extraction of valuable knowledge from data streams and transforming them into actionable insights for public and private Enterprise requires the establishment of commercially viable marine information ecosystems based on semantically enriched and web accessible information.

In this paper we present a service platform supporting semantic alignment, geo-spatial fusion and linked-data access to underwater IoT data streams. The platform supports harmonization of temporal and spatial resolution (resampling and interpolation), array based data composition, data reconciliation of environmental properties and map layer visualization. In addition, using an architecture based on open standards (Open Geospatial Consortium and W3C), the platform reduces the time, cost and complexity of developing marine applications, whilst allow for added-value data abstractions to be published and made accessible to wide range of applications.

In the following sections we provide background motivation for our work, a system architecture and results from an application trial conducted at the Leixões harbour in Porto, Portugal, to manage hazard maps for maintenance of the navigation channel.

## 2. BACKGROUND

The Scientific community now faces a situation where access to sensor hardware and management frameworks (e.g. Internet of Things) is easier than ever before; this is leading to the production of an increasingly large quantity of experimental data (such data can quickly reach terabytes in size). This outpouring of data poses many challenges, certainly from a data representation and processing point of view, but also from communities of practice where contributions from many parties must be semantically represented, organised and aligned towards a common ontology or domain relevant nomenclature to ease data discovery and integration.

Scientists generally rely on general purpose data formats to store and share the data produced by their experiments (e.g. Comma Separated Values (CSV) or more optimised binary formats like **NetCDF** [2] or **HDF** [3]). Despite increasing quantities and complexity of data, scientists are reluctant to move to relational database management systems (DBMS) that offer optimised data access and the parallelisation of analytic procedures; this is in large because [4]:

- DBMS are usually business oriented and lack support for scientific data types (arrays, spatial, temporal).
- DBMS do not provide good access to data for analytic tasks; file-based storage together with programmatic APIs.
- There is little support for data visualization and plotting.

The ICT community has also recognised these limitations and proposed solutions for the management of large non-relational data collections in different domains (e.g. multidimensional array data). **MonetDB** [5], **rasdaman** [6] and **SciDB** [9] among others, have been developed to support scientific communities in managing and analysing their experimental data. The earth science community has produced tools to efficiently manage collections of well-known binary formats (e.g. the aforementioned **NetCDF** and **HDF**, but others as well) but the support to query, process and aggregate managed data remains limited.

Dissemination of experimental data is also an important issue. Exchanging experimental data, even using self-descriptive formats like **NetCDF**, does not allow the automated machine processing needed to align the internal data semantics to a shared and formal conceptualization of the physical realm under study. Even a meta data format for **NetCDF** files like **NcML** [10], allows only primitive data type properties, not including the possibility of aligning external terminologies in the format of URIs.

The Open Geospatial Consortium (OGC henceforth) have produced a wealth of standards in support of data interoperability for geospatial data and applications. When multiple data sources are used in conjunction in order to implement an analytical work flow a great deal of work is devoted to homogenize spatial and temporal representations with the minimum loss of information.

OGC has standardised the vocabulary used to describe and query “coverages”, which are described in OGC’s standards as “digital geospatial information representing a space/time varying phenomena”. The concept of coverage, within the context of IoT

data streams, is quite important since every IoT framework collects some phenomena of sort. The standardisation of the services for requesting and harmonising of “coverages” is described in the OGC Web Coverage Service (WCS) suite which is composed of a core set of services that specify the definition of coverage as:

“Feature that acts as a function to return values from its range for any direct position within its spatiotemporal domain” [7].

The OGC WCS suite is then enriched with a set of extensions for handling: transformation of data into a different coordinate reference system (CRS), request for a server-side interpolations mechanism when requesting for a coverage, initiate server-side processing and filtering of data, extract specific fields in the data by range sub setting, and request for specific scaling of a coverage.

In order to ease the data discovery task it is important to harmonise the meta data information of a data set with a common reference model and adopt a domain/task relevant ontology to semantically annotate the meta data.

When dealing with sensor observation we can focus on different, often vendor based, sensor data formats and respective data models, rely on some more open formats, or adopt widely adopted standards. **NetCDF** is a set of software libraries and data formats that support the creation, access, and sharing of array-oriented scientific data. The format defines meta data that provide a description of what the data represents, and the spatial and temporal properties of the data. Initially developed by Unidata, **NetCDF** is an OGC accepted standard since 2011.

The **NetCDF** data model includes four main entities:

- **Dataset:** according to ISO 19101, a dataset is an identifiable collection of data.
- **Attribute:** contains meta data about properties of a variable or of a data set. Variable attributes may be used to specify properties such as units.
- **Dimension:** are used to specify variable shapes and may be associated with coordinate variables to identify coordinate axes.
- **Variable:** holds a, potentially multidimensional, array of values of the same type.

This basic meta data model is general enough to be applied to many domains where it is used to describe data sets profiles which are subsequently aligned towards domain ontologies [8]. The ontological commitment offered by the shared ontology is then used to fuel higher level system functionalities.

The semantic alignment of data sets’ meta data encoded in **NetCDF** is a task that has more resemblances with record reconciliation than schema alignment since the actual schema of reference is provided and the entities handled are used as instances to annotate the data sets. In other domains, in order to support the semantic reconciliation of data items in data sets, tools like OpenRefine identified a specific service called Reconciliation Service.

## 3. MARINE INFORMATION ECOSYSTEM

The marine information ecosystem is made up of stakeholders responsible for the modelling, analysis and planning of marine management activities based on the production and consumption of marine information. Probabilistic and deterministic analytical processes build situational awareness regarding the impact of human activity and natural events, licensing, ecological modelling, ecosystem management and sustainable development. The

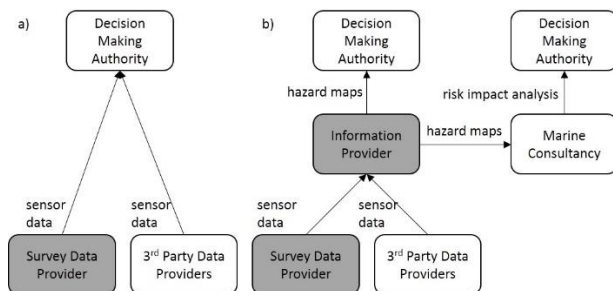
availability of increasing accurate and timely information produced at lower costs increases situational awareness and hence the quality decisions making.



**Figure 1 Marine information ecosystem**

Key technological trends within the ecosystem in respect to autonomous underwater vehicles (AUVs henceforth), advanced data analytics, and linked data are seamlessly connecting surveyors, marine analysts, decision makers and regulators in ways that were not previously possible. Firstly, low-cost AUVs, which are now small enough to be launched by one person from the shore and with trends in cost reduction can now be pre-configuration for specific marine applications such as scour and sediment transport analysis. Surveyors no longer need to operate manned sea going vessels and marine surveying can become a land based activity operated by personnel deployed on-demand. Combine the AUVs with Internet connectivity, geospatial data fusion, and linked data descriptions accessible on the web, and marine analysts can have the information they need for assessing situations rapidly. In addition, ongoing curation of marine information made accessible to teams of analysts through open and closed data policies will increase geospatial knowledge and drive improvements in environmental predictions.

When these technological advances are combined IoT Infrastructure Operators can maximize the potential of data streams through repurposing and reuse of data across multiple applications to ensure positive return on investment associated with the IoT infrastructure costs, whilst Application and Service Providers can collect, aggregate and synthesize large-scale heterogeneous data streams to deliver insights to decision makers. These advances will create opportunities for new information services that support novel data value chains and exploit and valorize advanced IoT infrastructures and added-value higher level data analytics services.



**Figure 2 a) current and b) future data value chains**

Fundamental to viability will be data-centric business models for multi-level information services that publish observations (e.g. seabed geometry, current flows, turbidity, temperature, etc.), semantically enriched and harmonized data (e.g. coverages), and higher level abstractions such as hazard and risk maps. The emergence of information services and brokers supporting non-

linear interactions and values chains will be essential. Evidence for a sustainable business model for information brokering must be designed considering the needs of data provider and data consumers in the market. In the following section we describe the EXPOSURES platform supporting marine information services followed by an application trial exploring a representative data value chain.

## 4. SYSTEM DESCRIPTION

### 1.1 System Architecture

The platform developed in EXPOSURES aims to valorise the in-situ data collected by IoT infrastructures by providing a web enabled platform that implements:

- An open service platform for semantic annotating scientific experimental data collected from heterogeneous underwater sensors and 3<sup>rd</sup> party data sources.
- A linked data interface to the managed data sources in order to ease their discoverability and to enable the ability to semantically querying the meta data assets.
- A portal to query, resample, aggregate and process the managed data in order to generate analyses that can be exported as binary files via standard services.
- Integration with SUNRISE underwater IoT Infrastructure Web Service API to ingest data collected during the experiment conducted within the identified test beds.

The EXPOSURES platform implements a web portal where the users can access the platform services via a REST layer to: upload their data sets (either via files or by connecting with SUNRISE's portal), semantically annotate them, share them with other users, using the domain ontology and the map interface to search for other data sets, and process the experimental data using OGC standards.

The system will maintain an RDF representation of each dataset's meta data whose schema will reflect the one provided by the **NetCDF** standard with a functional interpretation of data set and the subdivisions of variables, dimensions, and attributes.

In order to promote a modular architecture of the platform in a set of services that can be reused and re-purposed, ease the integration with third party software, and promote the platform interoperability with a scientifically oriented community the devised architecture adopts open standards whenever it is possible. The standards adopted in the architecture belong mainly to the OGC family for the description and encoding of the observation data sets and to the W3C family for the data sets' semantic description and dissemination of meta data using linked data. In particular:

- SPARQL (query language and protocol): is used to interface the system with the RDF meta data representation.
- WCS: is used to deliver the coverages to GIS tools (e.g. QGIS) that can process them.
- WCST: is used to ingest into the chosen array database the user's coverages.
- WCPS: is used to issue user's array queries to run on the array database.
- Reconciliation API: as defined by the OpenRefine documentation.
- Linked Data: is used to disseminate meta data elements outside the EXPOSURES portal.

In details, the architectural components belonging to the EXPOSURES Service Layer depicted in Figure 3 are described as follow:

**SUNRISE Interop:** this module is in charge of managing the interoperability between the EXPOSURE platform and the SUNRISE portal to ingest experimental data coming from the IoT infrastructure.

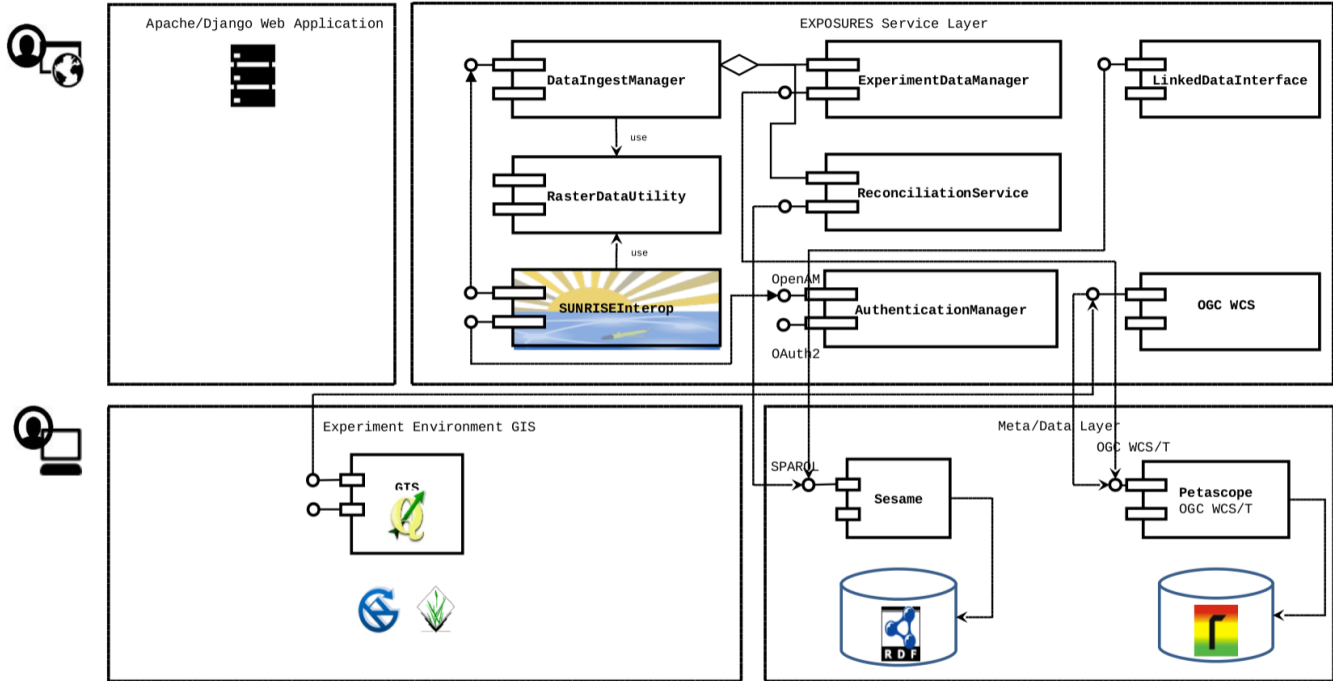


Figure 3: EXPOSURES Architecture

**Authentication Manager:** this module implements users' authentication mechanisms using Google OAuth2 standard to register with the platform, and OpenAM to integrate with SUNRISE's authentication system.

**Data Ingest Manager:** this module is in charge of importing and managing third party data sets and experimental data (i.e. both meta data and data) into the platform. The search API integrate both geospatial and ontology based search criteria.

**Experiment Data Manager:** this module is in charge of storing and managing the observation data (i.e. the numerical portion of a data set) using the OGC compliant services.

**Reconciliation Service:** this module is in charge of the semantic reconciliation of data sets' meta data towards a reference ontology (e.g. at present the NASA SWEET ontology is used but this can be parameterised).

**Raster Data utility:** this modules provides utility functions to extract meta data information from and translate between multiple raster formats (**NetCDF** and **GeoTIFF** supported so far).

**OGC WCS:** this module is in charge of wrapping the existing WCS endpoint provided by the chosen array database (i.e. **rasdaman** with **Petascop**) and extending its functionalities by implementing user oriented features:

- Enforce access control policies to maintained data sets
- Personalise the access to data sets by GIS tools based on user's data queries and a user's personalised data space

**Linked Data Interface:** this module is in charge to access the underlying SPARQL compliant RDF store (e.g. Sesame in this case) and to expose the managed meta data as Linked Data (RDF/XML and Turtle formats supported so far).

**Petascop:** this is the OGC WCS family compliant REST endpoint to the **rasdaman** array database [6] which stores and process multi-dimensional data cubes of environmental data sets.

**Sesame:** this is the chosen RDF store which manages the data sets' meta data and semantic annotations.

## 1.2 Enabled Use Cases

The EXPOSURES platform allows users to share environmental data in general, and IoT data sets in particular, more easily within a community. In fact it supports users, which are always in control of the process, in explicitly semantically annotate meta data entities using a chosen domain ontology and enable data value chains from the data collection done by surveyors, to the data discovery and processing via standard interfaces done by data analysts. Four main use cases have been identified:

1. User data ingestion and reconciliation
2. User data sets' search
3. GIS tool data usage
4. Server data set processing

## 1.3 User data ingestion and reconciliation

In this use case the user, a surveyor, has a raster file which contains the observations of a given property of a region of interest. The user uploads the file using the submission form specifying the collection date and providing a title and a description for the submitted data set. Alternatively, the user can submit a point cloud data to the system containing the measurements of the environmental property associated with a geospatial location. The EXPOSURES platform will automatically generate a raster data from it although many of the meta information will be lost.

A last option for the user is to use the SUNRISE portal to collect the data set, and then provide EXPOSURES with the relative experiment id to ingest the data directly using the SUNRISE APIs. The data from the SUNRISE portal will be transformed from point cloud into a raster format as described earlier, and then ingested as a normal raster data set.

The system detects the particular raster format of the submitted file and use the appropriate function to extract the meta data: CRS used, bounding box, variables and dimensions contained in the data set. Then the data and meta data are stored in the appropriate back ends (i.e. meta data into the RDF store and the data into the array database) and the user is then forwarded to the data set page where he/she can edit the meta data.

A particular icon allows the user to invoke the reconciliation service over the dimensions, variables, and measurement unit entities to semantically annotate them. The reconciliation service propose the user a set of entities from the domain ontology which are syntactically similar to the entity name. If no returned entity is appropriate the user can enter a new name and restart the search. Once the user approve of an alignment, the equivalence is then saved in the triple store.

These semantic annotations will be used to characterise the particular environmental parameter contained in a data set using a reference domain ontology. The alignments to the ontology's domain concepts will then be using during the query phase to support the data discovery.

## 1.4 User data sets' search

In this use case the user, a data analyst, wants to search for managed data sets in a region of interest. In order to do so the data analyst uses the map to visualise the general area and then draws a detailed region of interest in the map as a polygon. Afterwards he/she uses the reconciliation service to query for property of interest (e.g. **depth** for bathymetric data) from the managed ontology.

Once both the region and the property of interest are provided, the system provides the user (see Figure 4) a list of data sets which satisfy the search criteria. The provided data sets represent either the wanted property or a more concrete property collected within the region.

If the user wants, he/she can save some of the returned data sets by dragging and dropping them onto the "Data Space" menu. The "Data space" is merely a collection of data sets associated with the user which can be of use later on.

By clicking on the "Data space" menu the user can always edit or reset his/her own data sets collection.

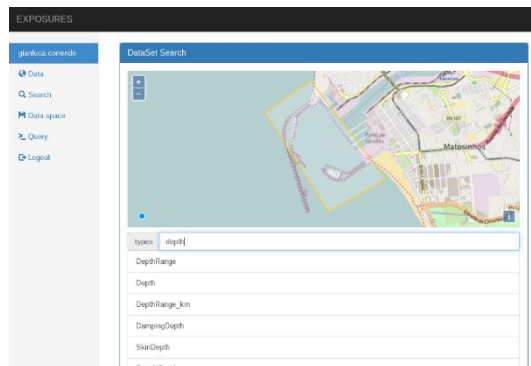


Figure 4 User data search interface

## 1.5 GIS tool data usage

In this use case the user wants to use some of the data sets managed by the EXPOSURES platform within a GIS tool to perform geospatial algorithms or data fusion (see Figure 3, Experiment Environment GIS).

As a precondition to this use case, the user has already searched for the wanted data sets by using the previous use case and has saved them in the personal "Data space". Then, the user starts a GIS tool with the capability of accessing WCS services. In our case QGIS with a WCS plugin has been used. The tool is used to connect with the EXPOSURES platform's WCS endpoint.

The endpoint will use the user authentication credentials to identify the user and will propose as available not all the coverages managed by the system, but instead only those within the user's data space. In this way the user can search for the wanted data sets and reuse the results of previous searches in the GIS tool.

Finally the user specify the usual parameters for downloading the coverages (i.e. coordinate reference system, bounding box, and interpolation method), downloading them as raster files and processing them within the tool.

## 1.6 Server data set processing

In this use case the user wants to process on server side a given coverage in order to delegate part of the processing burden to the array database storing the coverage.

As a precondition to this use case, the user has already searched for a particular data set which he/she wants to process. By accessing the data set description the user store the dataset's coverage id.

By selecting the "Query" menu, the user is provided with a textual interface which allows the user to enter a WCPS query for the given coverage. The results of the query can be encoded using the WCSP language's features into an image layer (e.g. in png or jpg image format).

Once the user click on the "Run" button (see Figure 5) the results, if they are encoded as an image, are displayed in the map overlaid with open street map to see where the data is located.

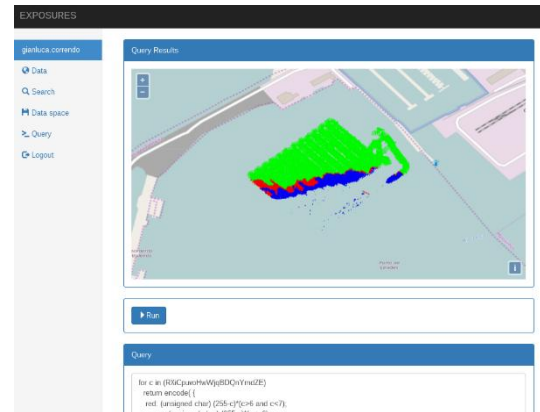


Figure 5 WCPS query processing

## 5. APPLICATION TRIAL

In this section we describe an evaluation of the marine data value chain based on underwater IoT, and shortly introduced in Section 3, and the EXPOSURES platform. Based in Porto, Portugal, the *Porto de Leixões*, is one of the country's major seaports and

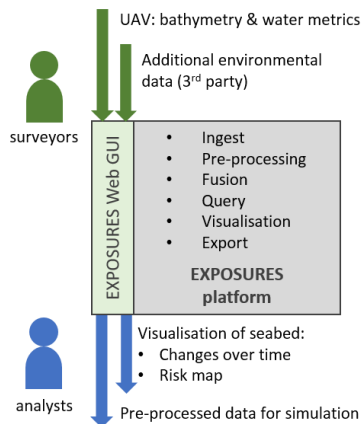


routinely manages high levels of industrial marine traffic. Maintaining the efficient and safe flow of traffic in and out of the port is a core element of the Porto harbor management's concerns. Achieving this goal requires oversight of a wide range of processes including the appraisal of environment factors that can impact the safety of ships in transit. One such factor is the continuously changing seabed profile which over time (or after exceptional environmental events) can present risks to maritime traffic when accumulated to form large, built up bodies of sediment. Key to mitigating against these risks is a strategy that can quickly and efficiently appraise the seabed state and take appropriate action to mitigate risks.

Recent advances in the availability of low-cost, underwater autonomous vehicles (UAVs) capable of carrying an array of sensors that can transmit data wirelessly means that surveys of the sea water volume and its bed can now be carried out much more easily and cheaply than previously possible.

Data captured by the UAVs is important to harbor stakeholders but must be processed and transformed for it to become useful in realizing port management goals. Previously the capture and processing of this data has been arduous due to the often complex, multi-step and sometimes ad-hoc methods required. However, the confluence of data streams generated by new UAV technology with emergence of novel data chains offered by the EXPOSURES platform affords capabilities that significantly increase the accessibility and utility of such survey data. Specifically, the EXPOSURES platform provides a means by which 'raw' seabed survey data can be easily uploaded to an information service and from there transformed into a format that is ready for fusion with other data sets and easier to apply to existing data analysis methods.

Our trial in the Porto harbor is used as a case study to exercise and evaluate the capabilities of the EXPOSURES platform with respect to realizing this new data value chain for the marine industry. The context for the case study is one where harbor management makes a routine evaluation of the current state of the seabed and where after any potential problems found may lead to action based on informed decisions. Here we focus on the data value chain scoped to the capture of 'snapshot' seabed data; its analysis and finally the process leading to evaluation and decision making.



**Figure 6 EXPOSURES platform linking surveyors and analysts in data chain**

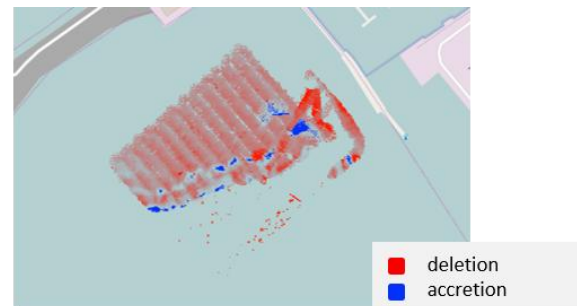
At the start of the data value chain are the surveyors that operate the UAVs which collect the seabed bathymetry, the water current

velocities, and the spectral data related to water turbidity. Once a survey mission specification has been defined, the surveyors deploy the UAV to take measurements within the area of interest (AOI) defined by harbor management. Upon completion of the physical sampling of the AOI, the data is uploaded in its raw form (as point-cloud layers) into the EXPOSURES platform, see Figure 6 above. Data coverages of this sort are typically formed of highly clustered data points that are sparsely distributed over the AOI. EXPOSURES interoperability with standard third GIS tools allows its users to efficiently pre-process this data such that missing or extraneous data points are handled using interpolation and clipping functions respectively. The result of which is a coherent, rasterized plane of environmental values (such as seabed height or water temperature) that can be accessed via query at a later stage.

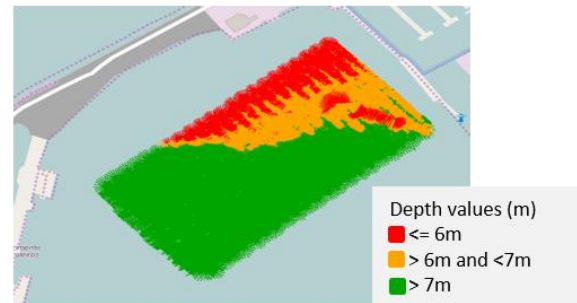
Multiple surveys across the same or proximal AOIs can be ingested in this way and then fused to produce a more complete view. Additional environmental or contextual data can also be added to this data set from third party sources. For example, land boundary geometries can be accessed from OpenStreetMap [11], rasterized and then fused with bathymetry data in order to extract a complete view of the profile of the harbor area that encapsulates the data originating from the sea. Views on the aggregated data sets are accessible via a browse and query interface provided by the EXPOSURES web application.

Analysts may choose to use this view on the data to a) visualize the current state of the harbor seabed for immediate appraisal and b) as inputs for computer simulation where potential future states can be approximated.

By way of example of the former case, by using EXPOSURES users can access bathymetries of the harbor area and create visualizations of the accretion/deletion of the seabed sediment profile. In this way analysts can create a map visualization wherein potentially dangerous build up on sediment that could threaten ship safety are shown (see Figure 7 (a) for a study conducted between 17/01/2013 and 02/08/2013).



**(a) EXPOSURES accretion/deletion of seabed sediment**



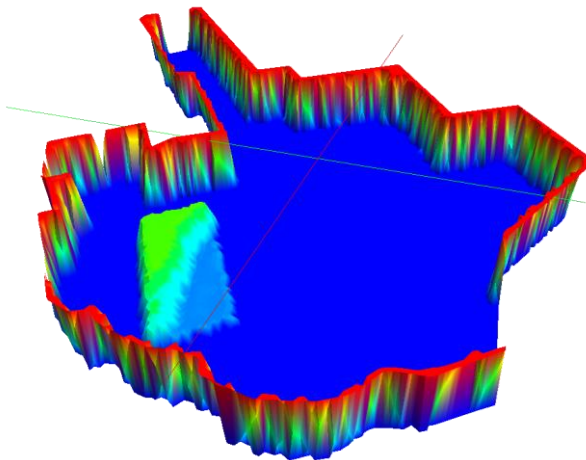
**(b) EXPOSURES: seabed sediment risk map**

**Figure 7 EXPOSURES visualizations**

Accretion and deletion of sediment between surveys provide indicators of the relative differences in seabed profile between two time periods. In this way analysts can spot areas that require special attention from harbor's authorities.

The risk map visualization makes clear those areas directly below the path of vessels where there is potential for unsafe passage. These visualizations provide the analyst with an immediate view on the environment under or near shipping lanes and may suggest that immediate actions (such as dredging) should be taken (see Figure 7 (b) for a sediment risk map computed for Porto's area on 02/08/2013).

However, these visualizations do not offer any insight on the processes that drive these changes in the first place, and provide little insight on future outcomes. To this end, the analyst may use these views as a well prepared data set to input into a computer simulation suite, such as Open Telemac [12], which can simulate the sediment transport. Using the EXPOSURES platform the analyst can significantly reduce the preparation workload usually required to set up the environmental data preparation for a simulation. For example, since the bathymetry data and land mass geometry are accessible as a fused data set within EXPOSURES, it is possible to define the simulation bottom geometry within pre-processing tools (such as Blue Kenue [13]) using data exported from EXPOSURES (see Figure 8), rather than having to follow arduous conversion and construction steps by hand.



**Figure 8 Porto harbor simulation mesh: Blue Kenue visualization of fused harbor hard boundary and sample bathymetry using data exported from EXPOSURES platform**

With additional, expert lead technical configuration of simulation steering parameters (provided by the analyst) simulations of Porto's harbor sea water and bed can be relatively quickly assembled. The results of such simulations can provide a longer term estimation of the dynamic changes to the harbor's natural environment over time. Alternatively, unusual and exceptional events simulation (such as violent storms or significant industrial accidents, such as spills) can be played out and the results used to guide port emergency planning.

Our on-going case study using the Porto harbor study is being used to gather further observations of the sea and seabed data and assess its application in new data chains that exploit visualization and simulation. The utility of the EXPOSURES platform will be evaluated with expert users selected from the marine industry.

## 6. CONCLUSIONS

In this paper, we have presented a service platform supporting semantic alignment, geo-spatial fusion and linked-data access to underwater IoT data streams. By supports harmonization of temporal and spatial resolution (resampling and interpolation), array based data composition, data reconciliation of environmental properties and map layer visualization, the cost and complexity of developing marine applications can be reduced. In addition, through open standards based on OGC and W3C linked data interfaces knowledge of marine environments can be curated, published and shared by communities and commercial teams of analysts responsible for understanding the impact of human and natural events within the blue economy. Future work aims to extend the services to support a wider range of geo-spatial fusion capabilities and to investigate automating semantic annotations of simulation input data within analytical workflows, whilst exploring commercial business models for marine information services with industry stakeholder groups.

## 7. ACKNOWLEDGMENTS

This research has been supported by the SUNRISE project under the FP7 framework, agreement number 611449.

## 8. REFERENCES

- [1] Karpathiotaki, M., Dogani, K., Koubarakis, M., Valentin, B., Mazzetti, P., Santoro, M., & Di Franco, S. (2014). Prod-Trees: Semantic Search for Earth Observation Products. In *The Semantic Web: ESWC 2014 Satellite Events* (pp. 374-378). Springer International Publishing.
- [2] G. Davis, et al, NetCDF. Network Common Data Form. An interface for array-oriented data access and a library that provides an implementation of the interface. <http://www.unidata.ucar.edu/packages/netcdf>, 2012.
- [3] M. Folk, G. Heber, Q. Koziol, E. Pourmal, and D. Robinson. An overview of the HDF5 technology suite and its applications. in *Proceedings of the EDBT/ICDT 2011 Workshop on Array Databases*, 2011, pp. 36-47.
- [4] J. Gray, et al. Scientific Data Management in the Coming Decade. *ACM SIGMOD Rec.*, vol. 34, no. 4, pp. 35-41, Jan. 2005.
- [5] S. Idreos, et al. MonetDB: Two decades of research in column-oriented database architectures. *Bull. IEEE Comput. Soc. Tech. Comm. Data Eng.*, vol. 35, no. 1, pp. 40-45, 2012.
- [6] P. Baumann. rasdaman: Array Databases Boost Spatio-Temporal Analytics. In *Computing for Geospatial Research and Application* (COM. Geo), 2014 Fifth International Conference on, 2014, pp. 54-54.
- [7] P. Baumann, J. Yu, M. Martinez, and S. Meissl, "OGC WCS 2.0 Interface Standard - Core," OGC 09-110r4, 2014.
- [8] Karpathiotaki, M., Dogani, K., Koubarakis, M., Valentin, B., Mazzetti, P., Santoro, M., & Di Franco, S. (2014). Prod-Trees: Semantic Search for Earth Observation Products. In *The Semantic Web: ESWC 2014 Satellite Events* (pp. 374-378). Springer International Publishing.
- [9] M. Stonebraker, et al. SciDB DBMS Research at MIT. In *IEEE Data Eng Bull*, vol. 36, no. 4, pp. 21-30, 2013.
- [10] S. C. Hankin, J. D. Blower, T. Carval, K. S. Casey, C. Donlon, O. Lauret, T. Loubrieu, A. Srinivasan, J. Trinanes, O. Godoy, and others. NetCDF-CF-OPeNDAP: Standards for ocean data interoperability and object lessons for community data standards processes. In *Oceanobs 2009*, 21-25 September 2009, Venice.
- [11] OpenStreetMap (2016) <https://www.openstreetmap.org>

[12] Open Telemac Mascaret (2016) <http://opentelemac.org/>

[13] Blue Kenue™ (2016) [http://www.nrc-cnrc.gc.ca/eng/solutions/advisory/blue\\_kenue\\_index.html](http://www.nrc-cnrc.gc.ca/eng/solutions/advisory/blue_kenue_index.html)