Article

POTENTIALLY POLLUTING MARINE SITES GEODB :

AN S-100 GEOSPATIAL DATABASE AS AN EFFECTIVE CONTRIBUTION TO THE PROTECTION OF THE MARINE ENVIRONMENT

By Giuseppe Masetti, Brian Calder and Lee Alexander
(Center for Coastal and Ocean Mapping & Joint Hydrographic Center)
University of New Hampshire

Abstract

Potentially Polluting Marine Sites (PPMS) are objects on, or areas of, the seabed that may release pollution in the future. A rationale for, and design of, a geospatial database to inventory and manipulate PPMS is presented. Built as an S-100 Product Specification, it is specified through human-readable UML diagrams and implemented through machine-readable GML files, and includes auxiliary information such as pollution-control resources and potentially vulnerable sites in order to support analyses of the core data. The design and some aspects of implementation are presented, along with metadata requirements and structure, and a perspective on potential uses of the database.

Résumé

Les sites marins potentiellement polluants (PPMS) sont des objets situés sur le fond marin, ou des zones du fond marin, qui sont susceptibles dans le futur de relâcher de la pollution. La raison d'être et la conception d'une base de données géospatiales visant à inventorier et à manipuler les PPMS sont présentés. Conçue en tant que spécification de produit de la S-100, elle est définie via des diagrammes UML lisibles par l'homme et mise en œuvre via des fichiers GML lisibles en machine, et elle inclut des renseignements auxiliaires, tels que les ressources anti-pollution et les sites potentiellement vulnérables, aux fins d'appuyer les analyses des données de base. La conception et certains aspects de la mise en œuvre sont présentés, en même temps que les exigences et la structure des métadonnées, et une perspective sur les utilisations potentielles de la base de données.

Resumen

Los sitios marinos potencialmente contaminantes (PPMS) son objetos o zonas de fondos marinos que pueden producir contaminación en el futuro. Se presenta un fundamento para y un diseño de una base de datos geoespacial para hacer un inventario y manipular los PPMS. Creada como una Especificación de Producto de la S-100, se especifica mediante un diagrama UML de fácil lectura y se implementa mediante ficheros GML (de marcaje geográfico) legibles por máquinas, e incluye información auxiliar como recursos para controlar la contaminación y sitios potencialmente vulnerables, para apoyar los análisis de los datos fundamentales. Se presentan el diseño y algunos aspectos de la implementación, junto con los requisitos y la estructura de los metadatos, y una perspectiva sobre los posibles usos de la base de datos.

Introduction

The presence of marine sites that are potentially polluting represents an increasing threat to the marine environment together with ocean acidification, ballast water and introduced marine species.

These marine sites may contain various types of hazards, including fuel oil, hazardous cargo, military weapons or munitions carried by warships or delivered to dumping areas, abandoned wellheads, etc. Even if petroleum-based pollutants represent the main threats to the global marine environment, mercury and other toxic substances also represent hazards since, for instance, they can cause contamination of the food chain. Collectively, these sites can be referred to as Potentially Polluting Marine Sites (PPMS).

Independent of the specific type, each of these PPMS represents a potential source of pollution for the marine environment. Each site may release toxic components in amounts variable with the state of preservation. This state is a function of many factors: the period of submergence, building materials, exposure to wave motion, presence of marine organisms, damage at the time of sinking and any attempt at salvage or demolition, etc. (Macleod, 2002). All of these factors influence the marine corrosion that inexorably corrodes the iron and carbon steel of anthropogenic structures.

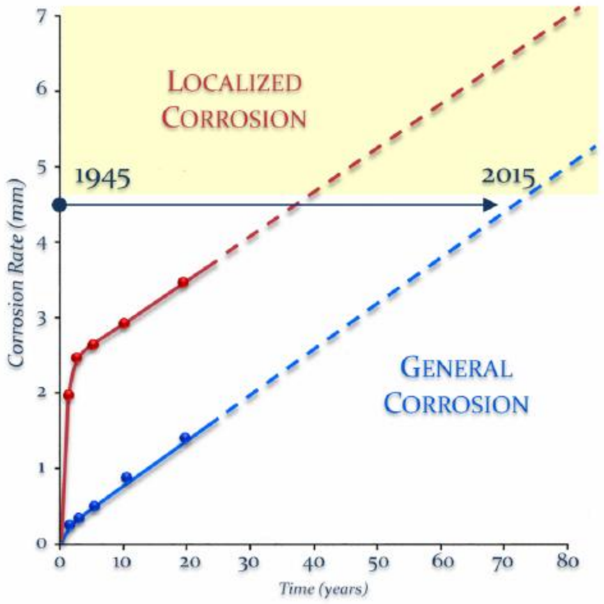

Figure 1 - ***Corrosion rates (adapted from Southwell et al., 1976). Localized corrosion can affect PPMS long before the main hull structure is compromised, and can lead to significant pollution releases (e.g., if an internal fuel pipe corrodes). Wrecks from the period around World War II (1939-1945) have significant potential to be affected by corrosion in the next decade due to both local and general effects.***

A mean value of the general corrosion rate varies from 0.05 to 0.1 mm per year (Macleod, 2010, Schumacher, 1979, Southwell et al., 1976). As a consequence, many shipwrecks from the Second World War (WWII) may start to spill their polluting content during the next two decades (Figure 1). Internal structures of ships are often considerably thinner than the external parts, however, and their collapse can lead to premature release of pollutants even if the main hull remains intact. Localized corrosion can cause perforation of tank walls and damage to internal pipes and valves so that recent shipwrecks may also start to leak their polluting content. Similarly, historic shipwrecks may spill pollutants much earlier than might otherwise be predicted.

Recent pollutant releases from PPMSs have resulted in significant impacts, including loss of marine life, economic impacts to coastal areas, and high costs to mitigate the effects. Events occurring throughout the world have led to an increased focus on the need to look proactively at the risks of oil and other pollutants being released from such submerged sources as shipwrecks, pipelines and dumping areas (Gertler et al., 2009, Michel et al., 2005). Furthermore, these events are related to the density of PPMSs in a particular area. For instance, the Mediterranean contains a high percentage of the world's sunken vessels – about 5% – when compared with its dimension and the intrinsic environmental fragility of a closed basin. Often driven by the occurrence of an environmental disaster, there are around the world many national and regional databases with different structures that are variously related to PPMS. The idea here is to delineate common requirements for a global database that, standardizing the collection of information about these sites, may better monitor and also contribute to reducing these events.

Although International treaties forbid the dumping of toxic wastes and national administrations strictly control their transportation and disposal, the illegal sinking of ships carrying toxic and nuclear wastes is an increasing concern. For instance, there are reports that this is a lucrative activity for various organized crime groups (PAM, 2010).

The cooperation among countries for identifying all the existing PPMSs represents means for better monitoring the presence of new ones. In a resolution adopted in March 2011, the Parliamentary Assembly of the Council of Europe underlined that *"without maps charting these risks, no accurate assessment of the threat can be made"*. The final recommendations of the cited resolution for the member States are, among others, to *"carry out systematic assessments of wrecks to identify any that pose a threat to the environment and keep them updated"*, and to support research in this field (CoE, 2012).

The increasing availability of geospatial marine data provides both an opportunity and a challenge for hydrographic offices and environmental centers to contribute to the identification and risk assessment of various PPMS. To adequately assess the environmental risk of these sites, relevant information must be efficiently collected and stored into a modern geo-database suitable for site inventory and geo-spatial analysis. Improved methods for the analysis and interchange of information on PPMS and threatened marine resources are also needed. Successfully managing information about such sites, and making it available for use and exchange in a uniform manner, is critical to effectively supporting a proactive approach to monitoring and remediation.

In particular, if a solution is to be effective, it must address three fundamental requirements:

- It must be generic enough to handle different types of potential polluters and auxiliary information;
- It must enable easy exchange and re-use of information; and,
- It must be standards-based to allow for ready adoption into available tools.

Shipwrecks are the most obvious, but by no means the only, source of pollution. For example, pipelines or abandoned wellheads can release pollutants, and old munitions or chemical weapons dumping sites are obvious risks to fishermen, divers and the local community. A successful database solution must be generic enough to represent various types of potential polluters, but do so in such a manner to allow specific analyses to be conducted that enable the site to be properly classified.

At the same time, the solution must support integrated thinking about how to plan for and respond to potential polluters. This was recognized by the International Maritime Organization recommendation *"to develop regional co-operation on aerial and satellite surveillance"* for problems (IMO, 2004). Gathering all relevant data in a sufficiently flexible database is one way of supporting this process.

Determining who is responsible for both the activities and cost of remediation after a polluting event if often complex, and may be exacerbated by national and international law. For example, it is generally held that shipwrecks continue to belong to their nation after they are sunk (Aznar-Gomez, 2010, Johnson, 2008), but it is unclear whether the owner is responsible for damages caused by pollution related to these wrecks. The U.S. Navy removed oil from the USS *Mississinewa* after a storm caused leakage of fuel (U.S. Navy, 2004) but asserted that this did not constitute a precedent (Guerin et al., 2010). It is likely that many events or potential events will include more than one actor, therefore, and exchange of information in a uniform manner is essential in timely appraisal and response (Woodward, 2008). Definition and adoption of a state-neutral database is therefore important in supporting the planning and response goals.

As a consequence of the requirement for interchange of information, it is inevitable that data related to PPMS are going to be used by different agencies across multiple software and hardware platforms. Although often dismissed as an implementation problem, it is important to consider requirements for compatibility and standardization when defining the structure of any putative database. In addition, while working within the constraint of a given standard often implies extra effort, this is rewarded by re-use of already available resources (e.g., feature catalogues) and can significantly improve rate of adoption in standard data manipulation packages such as desktop GIS systems. A practical (rather than merely efficient) solution for PPMS must therefore consider the requirement for a standards-based definition.

We propose in this article a model for the implementation of a PPMS geo-spatial database that attempts to satisfy all of these requirements. Drawing on previous example databases that were built parochially for specific purposes, core and extension requirements were extracted for a variety of potential polluters. This is further augmented by auxiliary information such as relevant resources (e.g., availability and location of pollution response equipment) and complementary information (e.g., sensitivities of coastlines to particular pollutants).

To ensure standards compatibility, the database was developed based on the International Hydrographic Organization's S-100 approach (IHO, 2010), while providing generic descriptions of various potential polluters. It is defined through a UML description (to assist in clear documentation) and uses an XML-based schema to provide a GML-structured computer-translatable description of the model. This paper describes the basic structure of the model and its XML implementation, and concludes with the proposal of a possible efficient implementation for the data storage of a PPMS GeoDB.

Adoption of the S-100 Workflow

If a new data structure for managing PPMSs at a global level has to be created, the new IHO S-100 Universal Hydrographic Data Model represents its natural framework (Figure 2).

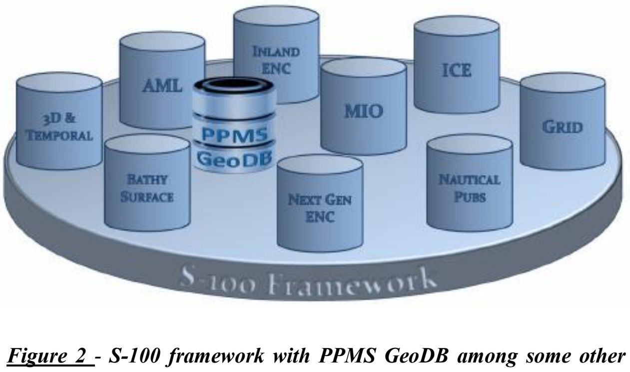

Figure 2 ***- S-100 framework with PPMS GeoDB among some other future S-100 series products. Developing within the S-100 framework allows the GeoDB to adopt already developed resources (simplifying implementation) and present its data in a common framework (simplifying adoption).***

A principal reason for this is the potential to adopt into the developing data structure some of the geographic features already present in the existing S-100 Feature Concept Dictionaries. These features have been created for some of the incoming Product Specifications of the S-100 series, and it is part of S-100 to share structures among different products to promote application interoperability and data reusability. The PPMS GeoDB project integrates the existing IHO data elements with new features and new attributes, derived from different solutions already implemented in existing databases. These new elements will be collected into a dedicated domain of the Supplementary Feature Concept Dictionary, and they will become themselves available for future use by other S-100 Products.

As defined in IHO S-100, a Product Specification (PS) is "a description of all the features, attributes and relationships of a given application and their mapping to a dataset" (IHO 2010). A PS is different but related to metadata: while metadata describes how a dataset actually is, a data PS describes how it should be, focusing on the requirements. The proposed PPMS GeoDB PS conforms to the S-100 requirement to be a precise and human readable technical document that describes a particular geospatial data product for hydrographic requirements (IHO 2010). This includes machine readable files that define the structure (XML Application Schemas), and can be converted to a XML Product Specification.

An S-100 based workflow was used to create the PPMS GeoDB PS. Outputs included:

- Definition of a vector-only product.
- Selection of required features, feature attributes, and enumerates in existing IHO Data Dictionaries.
- Identification of some new features that will be submitted for inclusion in an IHO Supplemental Dictionary.

The defined features and attributes were then described in a Feature Catalogue, and geometry types required in the product were determined. New geometry types will not need to be added to the S-100 framework for the proposed PS.

At this point, it was possible to construct an Application Schema. The creation was conducted in two different but related ways: a Logical model, using a conceptual schema language, and a Physical model using an encoding specific language (XML Schema).

Data Structure

Evaluating the entities required in a PPMS database is complicated by the diversity of objects to be represented. However, some important work was previously conducted with the aim of cataloging shipwrecks by ocean/basin location. This includes the South Pacific Regional Environment Program (SPREP) (Talouli et al., 2009, SPREP, 2002, Monfils et al., 2006) and Barrett Project (Barrett, 2011), the Atlantic, Mediterranean and Indian Ocean (AMIO) database (Monfils, 2005), a Mediterranean area in the Development of European guidelines for Potentially Polluting shipwrecks (DEEPP) project in 2005 (Alcaro et al., 2007), a global International Oil Spill Conference (IOSC) study in 2005 (Michel et al., 2005), among others. Collectively, these have been analyzed in regard to the types of information that are fundamental for a PPMS GeoDB to inform the design outlines here. A similar approach for non-shipwreck PPMSs was more difficult to conduct since there is less in the literature about this type of information in an integrated environmental-risk framework (Overfield, 2005, Aichele, 2010).

The successful collection and integration of PPMS information requires some effort to 'normalize' and standardize the data based on recognized international standards. As recommended in S-100, the Unified Modeling Language (UML) was used to create conceptual models that are implementation-independent. Each UML model class (or attribute) equates to a data dictionary, or an entity (or element). The resulting UML model indicates how the data are logically organized. Some selected UML views, that are portions of the total abstract model, will be discussed in the remaining part of this section.

In the proposed PPMS GeoDB PS, any product has a root element instance of the *Root* class. This root element may be related by composition with three types of composite Feature Collections (***Figure 3***). Thus, each PPMS product may have 3 main types of feature collection:

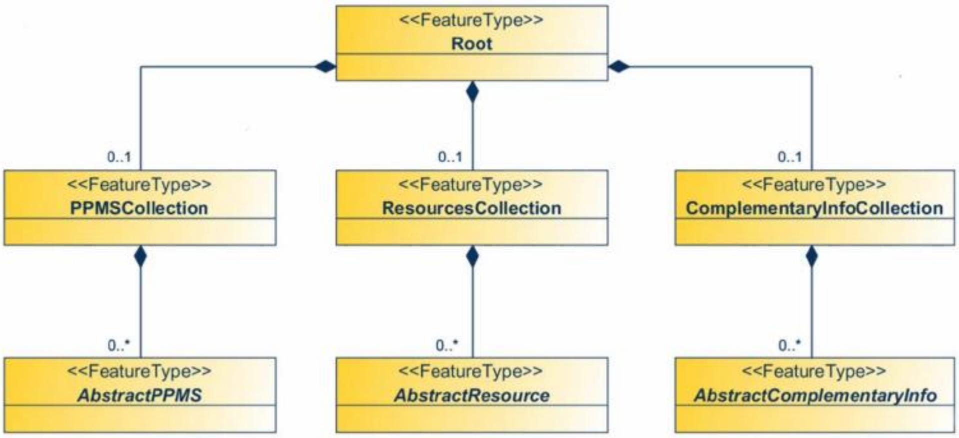

Figure 3 - ***Relationships of Root class. The GeoDB consists of zero or more collections of PPMS, resources and complementary information, as required by the applications for which it will be used. Note that each collection includes an unlimited number of features of common abstract type so that common methods can be applied that are useable on all features within the collection.***

- The Potentially Polluting Marine Sites;
- The Marine Resources threatened by the PPMS; and
- Different types of Complementary Info that represent auxiliary information that may be useful to the different phases of the disaster management cycle (***Figure 4***).

Each of these main feature collections can have infinite instances of different basic feature collections. Further, each collection inherits from an abstract class in which are defined all the shared characteristics between the different features. This allows the definition of shared methods that can be applied to any derived feature type. Finally, each of these composite Feature Collections can have an unbounded number of basic Feature Collections.

This data structure presents a certain level of complexity. For instance, the entities to model the possible types of PPMS are heterogeneous: from submarines sunk during WWII to oil rigs (***Figure 5***). Since some of these entities are already present in a basic "safety-of-navigation" form in the IHO Registry, they are enriched with a series of new attributes and enumerations, mainly on the basis of the content of the existing databases previously reported and the classification proposed by a Regional Marine Pollution Emergency Response Centre for the Mediterranean Sea (REMPEC, 2004). Similar to the feature collection level, the characteristics common to all the feature types have been efficiently grouped in a feature.

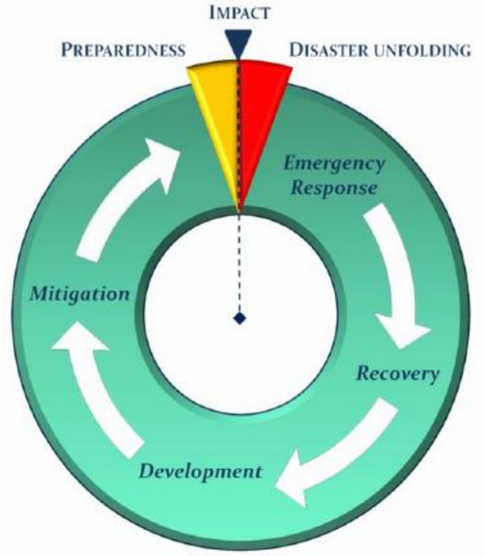

Figure 4 - ***The disaster management cycle. Used correctly, the PPMS GeoDB could provide key information at all stages of the cycle.***

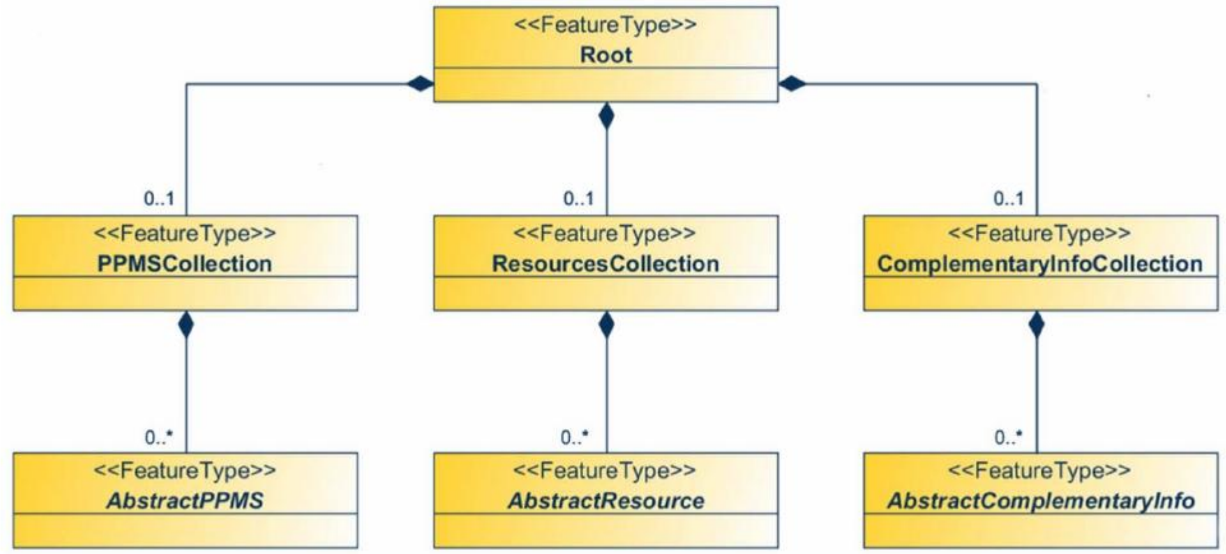

Figure 5 - ***Sub types and relative relationships of the AbstractPPMS class. All of the specializations of a PPMS derive from the Abstract class to allow common methods to be defined, but each specialized PPMS augments the resources maintained to provide information specific to the object being represented.***

As an example, Figure 6 outlines attributes and relationships proposed for one of the PPMS types: the Potentially Polluting Shipwreck (PPSW) class. This class may have different optional attributes. Most are derived from the "hydro" domain - already present in the IHO Registry. The limited number of additional attributes will become part of a specific domain of the IHO Supplementary Dictionary.

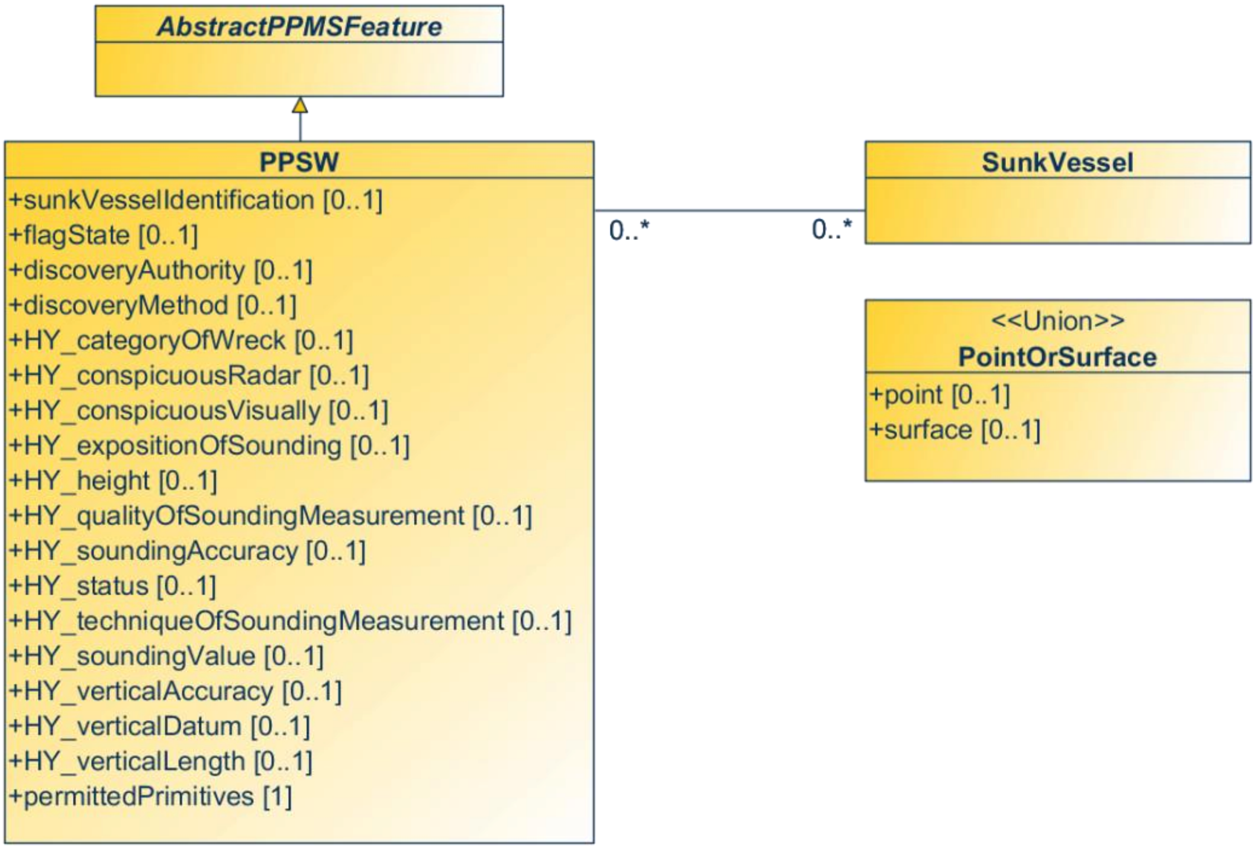

*Figure 6 - **Attributes of the PPSW Class derived from the AbstractPPMSFeature Class. Note particularly the many-to-many relation between the SunkVessel and PPSW, expressing the possibility that any one SunkVessel might be attributed to a number of PPSWs (e.g., the same wreck reported in different locations), and that any one PPSW might be associated with any number of SunkVessel objects (e.g., a wreck of unknown or dubious provenance). This is typical of the complexity of a general representation of uncertain features such as that expressed in the PPMS GeoDB.***

During the modeling process, many problems have been focused and solutions have been provided. For instance, a common problem with a shipwreck database is related to occasionally uncertain identification of the vessel sunk at a wreck site. For example, a wreck site can be associated with more than one vessel sunk in the area (Figure 7, top), or a sunken vessel can be associated to many possible wreck sites (Figure 7, middle). In some cases, a site inspection (e.g., by diver or ROV) is required to resolve uncertain associations (Figure 7, bottom). The many-to-many relationship between Sunk Vessels and PPSW classes is the solution adopted for this particular problem (Figure 6), since it allows for expression of the uncertain association of ships and sites.

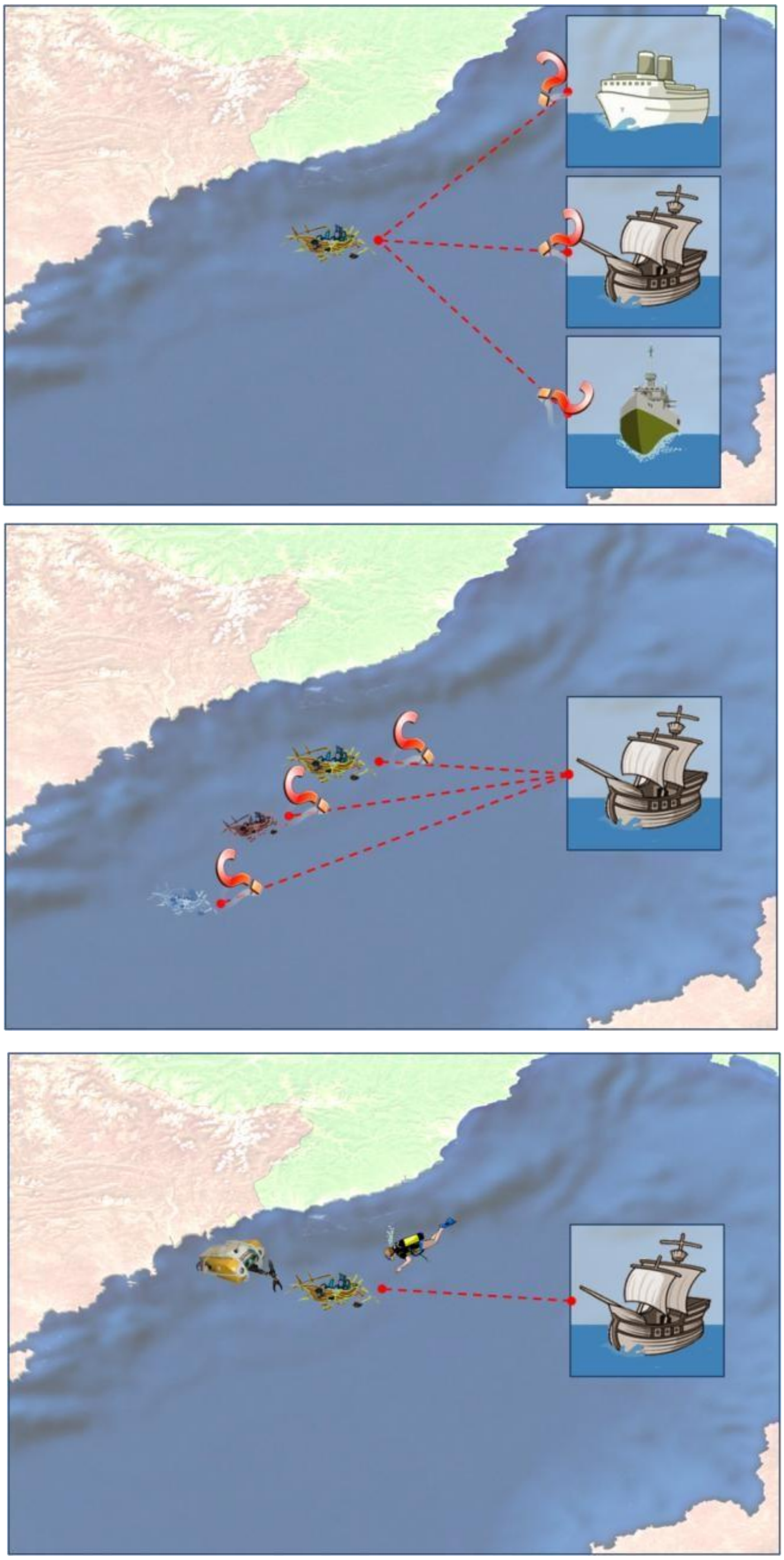

Figure 7 - ***Examples of different possibilities of the many-to-many relationship between PPSW and SunkVessel classes. Many shipwrecks may be associated with one PPSW (top) if the provenance of the wreck is not known, while one shipwreck may be associated with many PPSWs (middle) if its location is uncertain. Typically, a one-to-one relationship (bottom) can only be determined if auxiliary resources are used to investigate the wreck and establish a positive identification. Since this last case is rare, the PPMS GeoDB must support the uncertainty represented by the many-to-many relationship***

Dumping areas are another selected entity due to the large quantities of live ammunition, mines, chemical warfare agents (CWA), and other explosives present in a large number of marine sites (Plunkett, 2003, Sato, 2010, Beddington and Kinloch, 2005). This situation is the result of the past conviction that the dumping of CWA at sea was the best disposal method rather than to store them or incinerate them (Overfield, 2005). Currently, an increasing number of injuries and problems related to these dangerous objects are being reported (Laurin, 1991, Simons, 2003). Although the position of a large proportion of these dumping sites is known, many problems come from the buoyancy of containers used to store the waste materials, and the difficulties for the local authorities to supervise the correct position during dumping operations.

Abandoned and exploratory wells also represent a threat for structural failure over time, and the Deepwater Horizon disaster recently highlighted the dangers related to oil rigs and offshore extraction of hydrocarbons (Orth, 2011). Even if this last event remains in the memory of public opinion, large platform accidents represent only a limited part of marine oil pollution (Fingas and Charles, 2001) when compared to periodic releases of water containing small amounts of oil from offshore oil installations (Espedal and Johannessen, 2000, Farmen et al., 2010). Having these represented in the proposed GeoDB allows for spatial analysis to correlate objects with satellite Synthetic Aperture Radar (SAR) or other remote sensing sensors to distinguish between slicks due to hydrocarbon release and natural phenomena (Brekke and Solberg, 2005).

Some additional data resources are required to enable useful products to be generated from the GeoDB. These include shoreline, archaeological sites, fishing areas/ farms, marine sanctuaries, tourist installations, but are not necessarily 'objects' in the PPMS sense. As such, they are organized in two related groups: *ResourcesCollection* for marine resources directly or indirectly related to PPMSs, and *ComplementaryInfoCollection* for information auxiliary to the previous two entity clusters. Which of these entities have to be implemented is usually correlated to the applications that the database is called to answer. In fact, while for a simple inventorial aim the implementation of these entities may be simply ignored, a specialized application – as, for instance, oriented to risk assessment – will typically require them to be fully populated.

Metadata and Metadata Collections

A key element of the PPMS GeoDB is represented by the wide use of ISO 19100 Series Metadata, and the related S-100 profile currently in development (Figure 8).

In fact, the application schema alone is not always sufficient to grasp the meaning of the underlying data model: for instance, the labels identifying different entities may be ambiguous, and application-specific knowledge and semantic heterogeneities are common sources of misinterpretation (Maue and Schade, 2009). Misunderstanding and incorrectly using geographic data can be usually traced back to missing or unclear descriptions of their intended interpretation (Guarino, 1998).

Figure 8 - ***Sources for the metadata implementation of the PPMS GeoDB. Metadata that supports multiple levels of search and description (e.g., from presence of data to a specific detail of geospatial projection information) must be allowed to make best (and correct) use of the available data.***

A typical activity for a PPMS GeoDB includes the discovery of relevant geospatial data, their pre-processing, the application of appropriate analysis methods, and finally rendering the results on a map. Most potential semantic conflicts during this workflow may appear if source data has not been sufficiently specified at the beginning.

A PPMS GeoDB, as with any geographic data set, is a description of the real world at some level of approximation and simplification. The metadata developed for a PPMS GeoDB fully documents this process, explaining the data limitations and the adopted assumptions. At the same time, metadata permits any potential user to better understand the data, evaluate the applicability for an intended aim and, thereafter, use the data correctly. Furthermore, metadata could be used by the same PPMS GeoDB producer for data management (storage, updating, etc.) and by any user for facilitating data discovery.

The PPMS GeoDB adopts the ISO 19115:2003 core metadata that represent a minimum number of metadata elements required to identify a dataset for catalogue purposes. Their duty is to answer the following four primary questions:

- What: Does a dataset on a specific topic exist?
- Where: For a specific place?
- When: For a specific period?
- Who: Who is a point of contact to learn about/ order a dataset?

In addition to this core metadata, the following ISO 19115:2003 optional entity sets are implemented:

- Discovery Metadata, based on actual web metadata catalogues.
- Quality Metadata, extended for describing the risk assessment process adopted.

Along with these, the ISO 19115:2003 concepts of metadata hierarchy (three different levels of metadata), multilingual support (required for the international profile of the S-100 framework), and support files (to preserve usability) were also adopted. Furthermore, some complementary information collections are represented as collection of metadata (***Figure 9***). This unusual approach should permit an easier integration with other databases (providing a connection gate), and it should also limit wasteful and potentially dangerous data duplication.

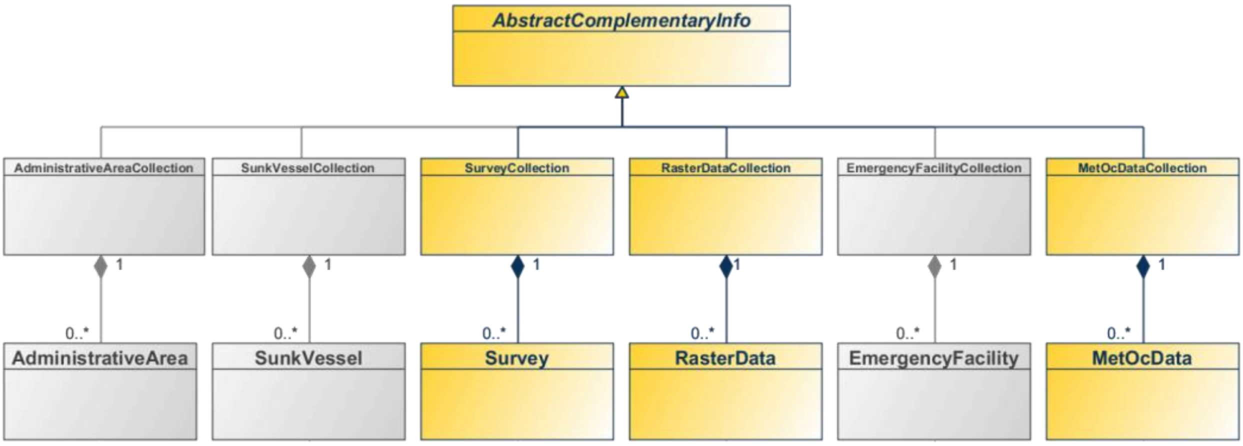

Figure 9 - ***Sources for the metadata implementation of the PPMS GeoDB. Metadata that supports multiple levels of search and description (e.g., from presence of data to a specific detail of geospatial projection information) must be allowed to make best (and correct) use of the available data.***

Physical implementation by Geography Markup Language

One important new feature provided by S-100 is the possibility for Product Specifications to adopt encodings different than the "ENC-traditional" format for information interchange (ISO 8211). In fact, the peculiarities of this latter format (e.g. the updating functionality and the minimal data volume) do not represent the best fit for many products other than ENC. Different encodings are available, and for several reasons the PPMS GeoDB has been defined using the Geography Markup Language (GML).

GML is an XML-encoding tag language defined by the Open Geospatial Consortium (OGC) to describe geographic objects (Lake, 2004).

Being built on the Extensible Mark-up Language (XML), it has some advantages of binary file formats (i.e., easy to understand by a computer, compact, the ability to add metadata), as well as some advantages of text files (i.e., universally interchangeable).

Since it is accepted by most industrial companies and research institutions, GML has become a *de facto* standard in spatial data processing and exchange. In 2007, version 3.2.1 became an ISO standard (ISO 19136). This ISO GML provides "[…] *an open, vendor-neutral framework for the description of geographical application schemas for the transport and storage of geographic information in XML"* (ISO, 2007). GML is one of the S-100 cited encodings, and the creation of a hydrographic community profile for GML has been recently proposed (TSMAD, 2012).

Other reasons for using GML include:

- It is an emerging standard;
- It is not a proprietary format;
- It offers wide interoperability with GIS and web applications; and
- Usability of the developed GML products by existing XML technologies.

A number of steps were followed to create several GML Application Schemas for a Potential Polluting Marine Sites GeoDB:

- Provide the declaration of a target namespace.
- Import the appropriate GML Core Schemas.
- Derive directly or indirectly all objects and object collections from the corresponding GML abstract types.
- Define properties (as global or local elements) for each object's content model.
- Define attributes for all of these objects and properties.
- Define Metadata Schemas as a function of the schema -defined objects.

Since GML is a markup data format (i.e., data without instructions) and not a programming language, the application of any operation to the information stored has to be implemented in an application written in a suitable programming language. Thus, in order to apply some data validation and manipulation on GML document based on the PPMS GeoDB PS, a basic C++ application is being developed.

Commonly, a program working with data stored in an XML format adopts either the Document Object Model (DOM) or Simple API for XML (SAX) method. Both DOM and SAX work on a raw representation of the XML structure (elements, attributes, and text). Thus, the developer needs to write a substantial amount of bridging code to transform information encoded in XML to a representation more suitable for the application. For the PPMS GeoDB application, an alternative approach called XML Data Binding was used. This approach skips the raw representation of XML, and delivers the data in an object-oriented representation generated by a compiler from an XML schema (Surhone et al., 2010, Kolpackov, 2007). XML Data Binding is a more efficient way to handle the GML documents, given the complexity of the PPMS GeoDB Application Schemas.

A possible efficient implementation for data storage and query application

Even if the PPMS GeoDB PS does not mandate any particular data storage, we consider a possible implementation for storing and querying GML since it represents a key element in obtaining the full efficiency from this technology.

A pure XML database does not represent, at the moment, the best choice for the necessary expensive process in its adoption (Ahmad, 2011). It has also been debated whether XML can be effectively used as a database language, since it is best supporting other applications (Schewe, 2005). Thus, a database language for XML is needed, and relational database languages such as SQL represent one possible mature, widely used and scalable solution for storing and querying XML data, if not necessarily the best language.

As a consequence, mapping XML data into relational data represents a crucial step. This operation – called 'shredding' – maps XML data into rows and columns of a relational table. After that, the original queries translated into SQL queries can be applied, and their results are internally translated back to XML. Currently, there is no easy, automated, or free solution for this task. In fact, database vendors are currently building tools to assist in mapping XML documents into relational tables. But, since they are still competing with one another, a standard for the mapping method does not yet exist (Atay et al., 2007).

The mapping process is not an easy operation due to the intrinsic differences between an XML document and a relational database. A relational database stores the data into "flat" tables; while, in a XML document, the information has a hierarchical structure, with elements that may be nested and repeated. Thus, as a first approximation, an XML document can be represented as a tree, where data are the nodes and their relationships are represented by the edges. It is also evident that the structural constraint information represented by the XML Schema may represent a useful element in the creation of the mapping design.

Based on the above considerations, three possible approaches to the mapping were developed. A possible evaluation criterion for these approaches is the number of relation redundancies produced in the relational schema (since they could create anomaly problems).

1. One approach is model-based, and basically traverses the tree, storing the path for every node visited into a table (Bohannon et al., 2002, Qin et al., 2005, Yoshikawa et al., 2001). The main problem is that this splits the data into small pieces that must be joined, increasing the storage size and potentially creating a lot of duplications.
2. In the structural-based approach, the constraint information represented by the XML Schema (or XML DTD) is used as a key element in the creation of the mapping design (Florescu and Kossman, 1999, Lee and Chu, 2001, Shanmugasundaram et al., 1999). In this approach, system generated IDs (that is, "parentID" and "parentCODE") are widely adopted, creating additional data and relation redundancy.

3. Another approach is semantic-based, and potentially without relation redundancies. However, some effort is required to capture the semantics of XML for mapping by keys, foreign keys, and functional dependencies (Liu et al., 2006, Atay et al., 2007, Lv and Yan, 2006).

The proposed PPMS GeoDB storage implementation is based on the third approach, mainly because its correct implementation permits the absence of relation redundancies that are wasteful in large databases. The implementation takes the advantages of the XML Data Bindings to store the PPMS GeoDB information into a dedicated relational database (Figure 10). The implementation of this approach is basically transparent for the user, since all the operation of validation, import, query and export are internally managed by the application interface.

Figure 10 - ***Sources for the metadata implementation of the PPMS GeoDB. Metadata that supports multiple levels of search and description (e.g., from presence of data to a specific detail of geospatial projection information) must be allowed to make best (and correct) use of the available data.***

Since the GML is not stored internally as XML, this structure is commonly called an XML-enabled database. The main reasons for the adoption of relational databases are:

- They are well known.
- They are widely used in the database industry.
- Users are largely familiar with them and with their performances.
- They are largely considered a safe choice by corporations.
- A producer could hesitate to switch suddenly to a new technology.

The above reasons reflect the current situation. But, with the likely development of XML native databases in the future, they could become the best fit for GML and thus also for the PPMS GeoDB.

Current/Future Applications

The PPMS GeoDB, developed in the S-100 framework, is a practical means of providing a geo-referenced picture of hazardous sites and related marine resources. Although the main target of the PPMS GeoDB Application is a PPMS inventory, its implementation can be a tool for each phase of the disaster management cycle: emergency response, recovery, development, mitigation, and preparedness (Figure 4). In addition, a risk index – representing an assessment of the magnitude of risk associated with any site – can be derived to determine the potential impacts of these PPMS using a GeoDB of this type (Masetti et al., 2012).

The impacts of natural or technological disasters can be prevented, or at least bounded, through an integrated approach to environmental risk assessment and safety management to identify the elements of risk and to prioritize actions (Fedra, 1998, Goodchild, 2010). While many studies are present in fields like floods, earthquakes and forest fires, a limited number are centered on the detection, study and analysis of risk from oil spill and other marine pollutants incidents (Castanedo et al., 2009, Kassomenos, 2004, Pincinato et al., 2009, Sofotassios et al., 1997). The information collected by the proposed PPMS GeoDB represents a contribution to this issue at global and sub-national scale; nevertheless the development of some tools and indicators structured on this product is desirable to better manage and monitor the risk of a large number of PPMSs.

The possibility to identify potential risks before the release of pollutants is a key element for a proactive approach. This approach could permit evaluation of each shipwreck site in order to decide on a direct intervention (i.e. the removal of the threat sources), the isolation of the threat, the preparation of a release management plan before the event, or the definition of a monitoring protocol, etc.

At the same time, a PPMS GeoDB permits inventory of possible assets and responders present in the area in case of a release notice. In the case of an unidentified source of oil (or any other pollutant) the PPMS GeoDB could return a list of suspected sites, possibly on the basis of the results from an analysis of oil samples recovered that permits determination of the type and age of the oil.

Because of different types of marine sites potentially dangerous to the marine environment, a PPMS GeoDB represents a better global solution to efficiently manage many PPMS-associated types of information. At the same time, the decision to develop an S-100 compliant Product Specification has the advantage of enabling a wide exchange of PPMS information. Furthermore, the proposed data structure – with the connection gates represented by the collections of metadata combined with the large adoption of existing IHO features and attributes – permits an easy integration with other existing HO's databases.

The adoption of an S-100-compliant GeoDB standard can thus become an important global contribution from the hydrographic community to reduce or at least better manage environmental and economic risks related to Potentially Polluting Marine Sites.

Bibliography

Ahmad, K. (2011). A Comparative Analysis of Managing XML Data in Relational Database

Intelligent Information and Database Systems. *In:* Nguyen, N., Kim, C.-G. & Janiak, A. (eds.). Springer Berlin / Heidelberg.

Aichele, R. O. 2010. *RUST - Creating a new data base for maritime salvage responses* [Online]. Available: http://www.inforworks.com/rust1.htm [Accessed].

Alcaro, L., Amato, E., Cabioch, F. h., Farchi, C. & Gouriou, V. (2007). DEEPP Project DEvelopment of European guidelines for Potentially Polluting shipwrecks. Rome (Italy) / Brest (France): EU D.G. Environment - Civil Protection Unit.

Atay, M., Chebotko, A., Liu, D., Lu, S. & Fotouhi, F. (2007). "Efficient schema-based XML-to-Relational data mapping", **Information Systems**, 32**,** 458-476.

Aznar-Gomez, M. J. (2010). "Treasure Hunters, Sunken State Vessels and the 2001 UNESCO Convention on the Protection of Underwater Cultural Heritage", **The International Journal of Marine and Coastal Law**, 25**,** 209-236.

Barrett, M. J. 2011. *Potentially Polluting Shipwrecks*. Master Project.

Beddington, J. & Kinloch, A. J. (2005), **Munitions Dumped at Sea: A Literature Review**, Imperial College London Consultants, London (UK).

Bohannon, P., Freire, J., Roy, P. & Simeon, J. (2002). From XML Schema to Relations: A Cost-Based Approach to XML Storage. **Proceedings of the 18th International Conference on Data Engineering**, 26 Feb 2002 - 01 Mar 2002 San Jose, CA (USA). 64–74.

Brekke, C. & Solberg, A. H. S. (2005). "Oil Spill detection by satellite remote sensing", **Remote Sensing of Environment**, 95**,** 1-13.

Castanedo, S., Abascal, A. J., Medina, R., Fernandez, F., Liste, M. & Olabarrieta, M. (2009). Development of a GIS-based oil spill risk assessment system. **OCEANS 2009 - EUROPE**, 2009 Bremen (Germany). 1-9.

CoE (2012), **Resolution 1869 - The environmental impact of sunken shipwrecks**, Council of Europe - Parliamentary Assembly.

Espedal, H. A. & Johannessen, O. M. (2000). "Detection of oil spills near offshore installations using synthetic aperture radar (SAR)", **International Journal of Remote Sensing**, 21**,** 2141-2144.

Farmen, E., Harman, C., Hylland, K. & Tollefsen, K. E. (2010). "Produced water extracts from North Sea oil production platforms result in cellular oxidative stress in a raibow trout in vitro bioassay", **Marine Pollution Bulletin**, 60**,** 1092-1098.

Fedra, K. (1998). "Integrated risk assessment and management: overview and state of the art", **Journal of Hazardous Materials**, 61**,** 5-22.

Fingas, M. F. & Charles, J. (2001), **The basics of oil spill cleanup**, 2nd, Lewis Publishers, Boca Raton, Fla.

Florescu, D. & Kossman, D. (1999). A Performance Evaluation of Alternative Mapping Schemes for Storing XML Data in A Relational Database. **VLDB 1999**, 1999 Edinburgh, Scotland (UK).

Gertler, C., Yakimov, M. M., Malpass, M. C. & Golyshin, P. N. (2009). Shipping-related Accidental and Deliberate Release into the Environment. *In:* Timmis, K. N. (ed.) **Microbiology of Hydrocarbons, Oils, Lipids**. New York: Ed. Springer.

Goodchild, M. F. (2010). Whose hand on the tiller ? Revisiting “Spatial Statistical Analysis and GIS”. *In:* Anselin, L. & Rey, S. J. (eds.) **Perspectives on Spatial Data Analysis**. Berlin (Germany): Springer.

Guarino, N. (1998). Formal ontology and information systems. *In:* Guarino, N. (ed.) **Proceedings of the International Conference on Formal Ontology in Information Systems (FOIS 1998)**. Trento (Italy): IOS Press.

Guerin, U., Egger, B. & Penalva, V. (2010), **Underwater Cultural Heritage in Oceania**, UNESCO.

IHO (2010). S-100 - Universal Hydrographic Data Model. 1.0.0 ed. Monaco (Montecarlo): International Hydrographic Bureau.

IMO (2004). Implementation of the OPRC Convention and the OPRC-HNS Protocol and Relevant OPRC Conference Resolutions. Marine Environment Protection Committee.

ISO (2007). ISO 19136:2007 Geographic Information - Geography Markup Language.

Johnson, E. 2008. Japan Asked to Avert Disaster From Pacific World War II Wrecks *Bloomberg* [Online].

Kassomenos, P. A. (2004). "Risk analysis for environmental hazards: the case of oil spills, in Crete", **Global Nest: the International Journal**, 6**,** 39-51.

Kolpackov, B. 2007. *An Introduction to XML Data Binding in C++* [Online]. Available: http://www.artima.com/cppsource/xml_data_binding.html [Accessed].

Lake, R. (2004), **Geography mark-up language (GML). Foundation for the GeoWeb**, John Wiley & Sons, Chichester (UK).

Laurin, F. (1991). "Scandinavia's underwater time bomb", **The Bulletin of the Atomic Scientists**, 47**,** 10-15.

Lee, D. & Chu, W. W. (2001). "CPI: Constraints-Preserving Inlining algorithm for mapping XML DTD to relational schema", **DataKnowledge Engineering**, 39**,** 3-25.

Liu, C., Vincent, M. & Liu, J. (2006). "Constraint Preserving Transformation from Relational Schema to XML Schema", **World Wide Web**, 9**,** 93-110.

Lv, T. & Yan, P. (2006). "Mapping DTDs to relational schemas with semantic constraints", **Information and Software Technology**, 48**,** 245-252.

Macleod, I. D. (2002). In Situ Corrosion Measurements and Management of Shipwreck Sites. *In:* Ruppe, C. V. & Barstad, J. F. (eds.) **International Handbook of Underwater Archaeology**. New York (USA): Kluwer Academic & Plenum Publishers.

Macleod, I. D. (2010). Japanese Ships and Aircraft. Corrosion Rates. **International Corrosion Workshop**, 2010 Newport News (USA).

Masetti, G., Calder, B. & Alexander, L. (2012). Developing a GIS-Database and Risk Index for Potentially Polluting Marine Sites. *Canadian Hydrographic Conference 2012.* Niagara Falls, Ontario (Canada).

Maue, P. & Schade, S. (2009). Semantic Annotations in OGC Standards. *In:* Consortium, O. G. (ed.) *Discussion Paper No. 08-167r1.* Wayland, MA (USA).

Michel, J., Gilbert, T., Etkin, D. S. & Urban, R. (2005). Potentially polluting wrecks in marine waters. *International Oil Spill Conference.* Miami (USA).

Monfils, R. (2005). The global risk of marine pollution from WWII shipwrecks: Examples from the seven seas. *International Oil Spill Conference.* Miami Beach (USA).

Monfils, R., Gilbert, T. & Nawadra, S. (2006). "Sunken World War II shipwrecks of the Pacific and East Asian Regions. The need for regional collaboration to address the potential pollution threat", **Ocean & Coastal Management**, 49**,** 779-788.

Navy, U. S. (2004). USS Mississinewa Oil Removal Operations. Washington (USA): Naval Sea Systems Command.

Orth, D. (2011). "Administering America's Offshore Oil Fields: How Fewer, Performance-Based Regulations Can Produce Better Results", **Journal of Environmental Law & Litigation**, 26**,** 509-534.

Overfield, M. L. (2005). Resources and Under Sea Threats (RUST) database: identifying and evaluating submerged hazards within the National Marine Sanctuaries. *MTS/IEEE Oceans Conference.* Washington (USA).

PAM (2010), **Dumping of toxic and radioactive waste and Human Trafficking in the Mediterranean**, Parliamentary Assembly of the Mediterranean, Rabat (Morocco).

Pincinato, F. L., Riedel, P. S. & Milanelli, J. C. C. (2009). "Modelling an expert GIS system based on knowledge to evaluate oil spill environmental sensitivity", **Ocean & Coastal Management**, 52**,** 479-486.

Plunkett, G. (2003). Chemical Warfare Agent (CWA), Sea Dumping off Australia. Australian Government Department of Defence.

Qin, J., Zhao, S.-M., Yang, S.-Q. & Dou, W.-H. (2005). XPEV: A Storage Model for Well-Formed XML Documents

Fuzzy Systems and Knowledge Discovery. *In:* Wang, L. & Jin, Y. (eds.). Springer Berlin / Heidelberg.

REMPEC (2004). Regional Information System. Part C: Databanks and Information Services. Section 2: List of Alerts and Accidents in the Mediterranean. *In:* IMO/UNEP (ed.).

Sato, R. (2010), **Sea-Dumped Chemical Weapons: Japan**, Global Green USA, Washington (USA).

Schewe, K. (2005). Redundancy, Dependencies and Normal Forms for XML Databases. *Sixteenth Australasian Database Conference.* Australian Computer Society, Inc.

Schumacher, M. (1979), **Seawater corrosion handbook**, Noyes Data Corp., Park Ridge, N.J.

Shanmugasundaram, J., Tufte, K., He, G., Zhang, C., DeWitt, D. & Naughton, J. (1999). Relational Databases for Querying XML Documents: Limitations and Opportunities. **25th VLDB Conference**, 1999 San Francisco, CA (USA). Morgan Kaufmann Publishers Inc., 302-314.

Simons, M. 2003. Discarded War Munitions Leach Poisons Into the Baltic. *New York Times*, June 2003.

Sofotassios, D., Tzimas, G. & Tsakalidis, A. (1997). A GIS-based Framework for oil-spill emergency management in the Mediterranean Sea. **JEC-GI '97**, 1997 Vienna (Austria). 412-422.

Southwell, C. R., Bultman, J. D. & Alexander, A. L. (1976). Corrosion of Metals in Tropical Environments – Final report of 16-Year exposures. Materials Performance.

SPREP (2002). A regional strategy to adress marine pollution from World War II wrecks. Majuro (Marshall Islands): 13th South Pacific Regional Environment Program.

Surhone, L. M., Timpledon, M. T. & Marseken, S. F. (2010), **XML Data Binding**, VDM Verlag Dr. Mueller e.K.

Talouli, A., Gilbert, T. & Monfils Gilbert, R. (2009), **Strategic Environmental Assessment and potential future shoreline impacts of the oil spill from WWII shipwreck Hoyo Maru Chuuk Lagoon - Federated States of Micronesia**, SPREP.

TSMAD (2012), **GML Encoding in S-100. Paper for Consideration by TSMAD. Submitted by UK.**, TSMAD 24 /DIPWG4-11-4A.

Woodward, T. (2008). Pacific World War II Wrecks Pose Risk of Toxic Leaks. *National Geographic Magazine.*

Yoshikawa, M., Amagasa, T. & Shimura, T. (2001). "XRel: a path-based approach to storage and retrieval of XML documents using relational databases", **ACM Trans. Internet Technol.**, 1, 110-141.

Biography of the authors

Giuseppe Masetti has served with Italian Navy since 1999 and, in 2008, received a Master in Marine Geomatics (University of Genoa, Italy).

After a period onboard the hydrographic vessel ITN Aretusa as Operation Officer, he achieved the FIG/IHO Category A certification.
In addition, he graduated (with honors) in Political Sciences in 2003 from the University of Pisa (Italy), and in Diplomatic and International Sciences in 2004 from the University of Trieste (Italy).

In 2010 he started a Ph.D. program in System Monitoring and Environmental Risk Managemente at University of Genoa.

At the Center for Coastal and Ocean Mapping he is studying Ocean Engineering mainly focusing on various issues related to the implementation of the new S-100 data model and the management of potentially polluting marine sites.

Brian Calder is an Associate Research Professor and Associate Director at the Center for Coastal and Ocean Mapping at the University of New Hampshire.

He has a Ph.D in Electrical and Electronic Engineering, completing his thesis on Bayesian methods in Sidescan Sonar processing in 1997. Since then he has worked on a number of signal processing problems, including real-time grain size analysis,seismic processing, and wave-field modeling for shallow seismic applications.

His research interests include methods for error modeling, propagation and visualization, and adaptive sonar backscatter modeling.

Dr. Calder is currently focusing on statistically robust automated data processing approaches and tracing uncertainty in hydrographic data.

Lee Alexander is a Research Associate Professor at the Center for Coastal and Ocean Mapping and Joint Hydrographic Center at the University of New Hampshire, and an Adjunct Professor at the University of New Brunswick (Canada).

Previously a Research Scientist with the U.S. Coast Guard, he was also a Visiting Scientist with the Canadian Hydrographic Service.

His area of expertise is applied RDT&E on electronic charting, international standards development, and the use of electronic charts for safety-of-navigation and marine environmental protection. He serves on several international committees

and working groups dealing with electronic charting, including the International Hydrographic Organization, the International Electrotechnical Commission, and the International Association of Lighthouse Authorities.

In addition, he chairs the Harmonization Group on Marine Information Overlays (MIOs).

Dr. Alexander received his M.S. from the University of New Hampshire, and Ph.D. from Yale University in Natural Resource Management and is also a captain (now retired) in the U.S. Navy Reserve.