A GEO-DATABASE FOR POTENTIALLY POLLUTING MARINE SITES AND ASSOCIATED RISK INDEX

BY

GIUSEPPE MASETTI

Master Degree in Marine Geomatics, University of Genoa, 2008

THESIS

Submitted to the University of New Hampshire

in Partial Fulfillment of

the Requirements for the Degree of

Master of Science

In

Ocean Engineering

December, 2012

This thesis has been examined and approved.

Thesis Director, Dr. Brian Calder
Research Associate Professor, UNH

Dr. Lee Alexander
Research Associate Professor, UNH

Dr. Jonathan Beaudoin
Research Assistant Professor, UNH

30 NOVEMBER 2012
Date

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

ACID	Atomicity, Consistency, Isolation and Durability
AEC	Atomic Energy Commission
AMIO	Atlantic, Mediterranean and Indian Ocean database
API	Application Programming Interface
AtoN	Aids-to-Navigation
AWOIS	Automated Wreck and Obstruction Information System
CCOM	Center for Coastal and Ocean Mapping
CGI	Common Gateway Interface
COTS	Commercial Off-The-Shelf
CRS	Coordinate Reference System
CSL	Conceptual Schema Language
CSS	Cascading Style Sheets
CTI	Coral Triangle Initiative
CWA	Chemical Warfare Agent
DBA	Data Base Administrator
DEEPP	Development of European guidelines for Potentially Polluting shipwrecks
DEM	Digital Elevation Model
DOM	Document Object Model
DQWG	Data Quality Working Group
DWT	Deadweight Tonnage
ECF	Expert Correction Factor
ECMWF	European Centre for Medium-Range Weather Forecasts
ERC	Environmental Research Consulting
ETOPO	Earth Topography Digital Dataset

GCxGC	Two-dimensional Gas Chromatography
GDAL	Geospatial Data Abstraction Library
GeoDB	Geospatial Data Base
GIS	Geographic Information System
GML	Geography Markup Language
GNU	GNU's Not Unix!
GPL	General Public License
GQL	GML Query Language
GRT	Gross Registered Tonnage
GT	Gross Tonnage
GUI	Graphical User Interface
HFO	Heavy Fuel Oil
HTML	Hyper Text Markup Language
IEC	International Electrotechnical Commission
IFO	Intermediate Fuel Oil
IHO	International Hydrographic Organization
IMO	International Maritime Organization
IOSC	International Oil Spill Conference
ISO	International Organization for Standardization
JHC	Joint Hydrographic Center
KML	Keyhole Markup Language
LIDAR	Light Detection and Ranging
MarPol	International Convention for the Prevention of Pollution from Ships
MaSiRI	Marine Site Risk Index
MetOc	Meteorological and Oceanographic
MPA	Marine Protected Area
MT	Metric Ton
OGC	Open Geospatial Consortium
O-O	Object Oriented
OPRC	Oil Pollution Preparedness, Response and Cooperation Technical Group

ORDB	Object-relational Data Base
OR&R	Office of Response and Restoration
PACPOL	Pacific Ocean Pollution Prevention Programme
PAH	Polycyclic Aromatic Hydrocarbon
PPMS	Potentially Polluting Marine Site
PPSW	Potentially Polluting Shipwreck
PS	Product Specification
RDB	Relational Data Base
RDBMS	Relation Data Base Management System
REMPEC	Marine Pollution Emergency Response Centre for the Mediterranean Sea
RINA	Italian Naval Register
RLC	Risk Level of Confidence
ROV	Remotely Operated Vehicle
RULET	Remediation of Underwater Legacy Environmental Threats
RUST	Resources and Undersea Threats
SAR	Synthetic Aperture Radar
SAX	Simple API for XML
SDK	Software Development Kit
SHIELDS	Sanctuaries Hazardous Incident Emergency Logistics Database System
SIDS	Small Islands Developing States
SOAP	Simple Object Access Protocol
SPREP	Secretariat of the Pacific Regional Environment Programme
SQL	Structured Query Language
SRB	Sulfate Reduction Bacteria
SSF	Site Scale Factor
TC	Technical Committee
TSMAD	Transfer Standard Maintenance and Application Development Working Group
UML	Unified Modeling Language
UNCLOS	United Nations Convention on the Law of the Sea
UNDP	United Nations Development Programme

UNH	University of New Hampshire
URI	Universal Resource Identifier
URL	Universal Resource Locator
UTF	Unicode Transformation Format
VGI	Volunteered Geographic Information
VLCC	Very Large Crude Carrier
WGS	World Geodetic System
WMS	Web Map Server
WWII	Second World War
XML	Extensible Markup Language
XSD	W3C XML Schema Definition
XSLT	Extensible Stylesheet Language
W3C	World Wide Web Consortium

ABSTRACT

A GEO-DATABASE FOR POTENTIALLY POLLUTING MARINE SITES

AND ASSOCIATED RISK INDEX

by

Giuseppe Masetti

University of New Hampshire, December, 2012

The increasing availability of geospatial marine data provides an opportunity for hydrographic offices to contribute to the identification of "Potentially Polluting Marine Sites" (PPMS).

To adequately manage these sites, a PPMS Geospatial Database (GeoDB) application was developed to collect and store relevant information suitable for site inventory and geo-spatial analysis. The benefits of structuring the data to conform to the Universal Hydrographic Data Model (IHO S-100) and to use the Geographic Mark-Up Language (GML) for encoding are presented. A storage solution is proposed using a GML-enabled spatial relational database management system (RDBMS). In addition, an example of a risk index methodology is provided based on the defined data structure. The implementation of this example was performed using scripts containing SQL statements.

These procedures were implemented using a cross-platform C++ application based on open-source libraries and called PPMS GeoDB Manager.

CHAPTER 1

INTRODUCTION

1.1 Background

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), and there is a need to adopt a standardized and efficient approach for monitoring them.

In fact, 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 (Macleod, 2002). All of these factors influence the marine corrosion that inexorably corrodes the iron and carbon steel of anthropogenic structures.

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 (in the period 2013-2030), as shown in 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. Even if desirable, an in-depth assessment of each PPMS in order to determine extent of corrosion, together with the amount of pollutants, present on the site is often too difficult and expensive.

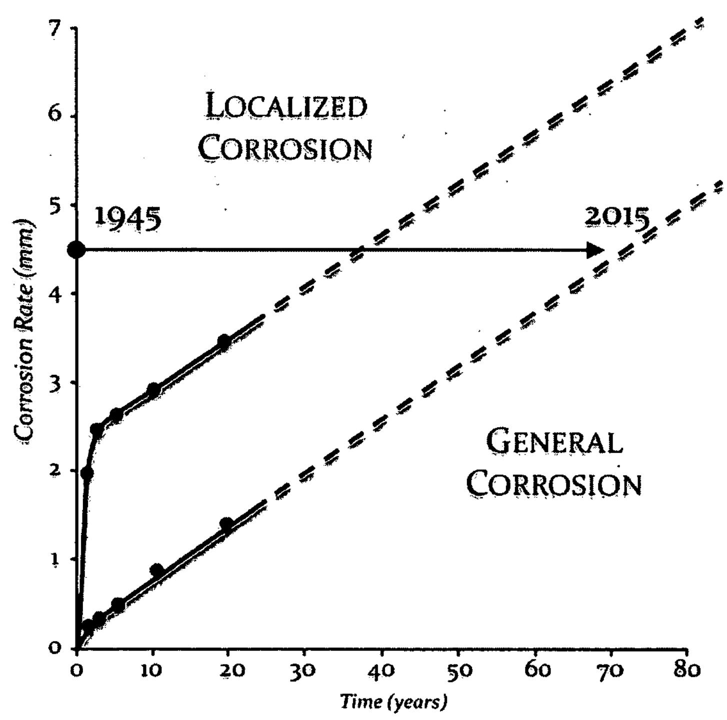


Figure 1 - Corrosion rates (adapted from Southwell et al., 1976) and area (in yellow) of potential leakage due to general corrosion based on mean hull thickness. Localized corrosion can affect PPMS long before the main hull structure is compromised, and can lead to significant pollution releases (e.g., 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.

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. Conversely, pollution incidents in the Pacific Ocean may have serious implications on large areas since this ocean is largely characterized by a lack of major land-barriers and a complex pattern of oceanic currents.

The relevance of PPMSs inventory and monitoring is also significant even when there are no visible traces of oil on the shorelines. In fact, this alone does not mean that the PPMSs in the area are not leaking. Oil odors and light shimmer on the sea surface are clear indication to a sub-surface release of an unknown quantity of pollutant. A part of this pollutant will eventually be assimilated by the marine environment. The rate that this occurs depends on its chemical and physical properties, the amount, the prevailing climatic and sea conditions, and whether the pollutant remains at sea or is washed on shore.

The significant number of vessels sunk during World War II (WWII) and in the many regional and local conflicts since (see, e.g., Brown, 2005), particularly those that may have been damaged by enemy action before they sank, mean that there is significantly increased risk of major polluting events in the near future. Improved methods for the collection, analysis, and interchange of information on wrecks are needed.

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"* (CoE, 2012). 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).

1.2 Problem Definition

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 should address three fundamental requirements:

1 Be generic enough to handle different types of potential polluters and auxiliary information;

2 Enable easy exchange and re-use of information; and,

3 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 as 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 are 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. As such, an exchange of information in a uniform manner is essential for timely appraisal and response (Woodward, 2008). The 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.

This thesis proposes 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 was 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 was defined through a Unified Modeling Language (UML) description (to assist in clear documentation) and used an XML-based schema to provide a GML-structured computer-translatable description of the model.

This thesis describes the basic structure of the model and its Extensible Markup Language (XML) implementation, with the proposal of a possible efficient implementation for the data storage of a PPMS GeoDB. It concludes with an application built on the outlined specifications with the aim to evaluate and prioritize the environmental risks related to the management of a large number of PPMSs.

1.3 Previous Research

During the past few years, much research activity has been conducted on XML technology. New markup and query languages have been developed, and new native XML and XML-enabled databases have been introduced. Even if much of this research is directed towards XML, many developments also represent important opportunities for GML since it is based on XML encoding.

The IHO S-100 data model has been recently introduced. Since some products of the S-10X series continue to be in an "advanced" draft status, there are no officially released Product Specifications. In the meantime, some working groups are beginning to focus on developing this new geospatial standard for marine data and information.

Although 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. These will be used as starting point to develop a tool to monitor the risk and prioritize the intervention with a large number of PPMSs.

1.4 Objectives

The primary objective of this study is to develop an efficient data structure for managing Potentially Polluting Marine Sites and all the geographic information that are variously related to them. To achieve this objective, two main steps are followed:

1 Define Product Specifications compliant with the Universal Hydrographic Data Model (IHO S-100) recently adopted by the International Hydrographic Organization (IHO).

2 Select a highly interoperable and widely-adopted encoding: the Geography Mark-Up Language (GML).

After that, there is a need to efficiently store such GML-encoded information in some sort of XML-capable database. Thus, a relevant secondary objective of this thesis is the analysis of the two main types of XML-capable databases (respectively, XML-native and XML-enabled) and the proposal of a possible solution, even if not part of the proposed Product Specification.

Furthermore, the potential to enlarge the aims of the PPMS GeoDB behind a simple inventory of potentially polluting marine sites has to be validated through the development of some sort of application. Since a common requirement for these types of sites is to be sorted following some meaningful criteria, another objective is to develop a risk index with an associated level of confidence. Ideally, this risk index can be used to prioritize the monitoring and intervention costs related to the large number of PPMSs that exist in various ocean areas.

1.5 Research Questions

Based on the above objectives, the following open questions were used as a guideline for the research efforts:

- What is the data model that best fits the requirements for a geospatial database that can manage Potentially Polluting Marine Sites and related geographic features? How can this data model be developed within the IHO S-100 framework?
- Among the many existing encodings what is the best choice for the implementation in terms of interoperability, availability and standardization? Among the many different storage options, which could be a reliable and ready-to-use solution?
- How could one take advantage of the collected information to prioritize the intervention in case of a large number of PPMSs?

1.6 Method of Research

The following steps have been followed in order to find answers for the defined research questions:

1 Definition by UML of an efficient data model to manage different type of PPMS.

2 Implementation of the outlined abstract data model using the Geography Markup Language (GML) as an encoding format.

3 Creation of an XML-capable database based on a widely adopted and open source product. Cost for the adoption and product popularity are two of the more important criteria used for the selection. Based on them, SQLite was selected since it is widely used in many fields, and its library is in the public domain.

4 Population and test of the database using an existing data set. For this step, the data collected for the European Union's DEEPP Project was used, integrated with a large number of other available sources.

5 Development and application of a risk index. The driving concept was the definition of a criterion with a level of confidence in order to prioritize the risk of any PPMS and thus the sequence of related mitigation operations.

6 Publish a website in order to have a GML Application Schema repository, an example of a web application based on the developed product specifications, and at the same time an easily reachable source of information about the conducted research.

1.7 Development Process

The creation of an application to inventory and manage data related to PPMS required a multi-disciplinary and structured development approach:

- Analyze different types of PPMS;
- Understand their specificities and the common requirements;
- Design a possible solution;
- Translate that solution into an application;
- Deploy the executable version into hardware;
- Make sure components all work together correctly.

This development approach went from general to specific, using a series of shifts in level of abstraction. It began with a conceptual understanding of the PPMS domain and high-level functionalities, moved into use cases and model refinements, and then finished with the design, the development and the deployment of a pilot PPMS GeoDB Application including an example of built-in risk index.

Feedback from a continuing interplay among the development parts aimed to increase the final quality of the solution. In fact, as problem understanding grew, new ideas were incorporated to develop a stronger system.

1.7.1 Requirements Gathering

The aim of this stage was to understand the essential of the PPMSs' domain, and to identify which problems a developed application could solve for the clients (the hydrographic offices, the environmental marine centers and, more generally, the public decision makers).

The products were a set of activity diagrams, capturing the main steps and decision points in the PPMS inventory and management process, and a high-level class diagram. Since the developed application did not emerge in a vacuum, this stage involved the evaluation of existing systems the new application will interface with (e.g., other products of the future S-100 Series, existing geographic standards).

1.7.2 Analysis

This stage increased the understanding of the problem using the results from the requirements gathering. A high-level use case analysis was developed to discover the actors initiating each use case, and the actors who benefit from it. After that, the sequence of steps in each use case was analyzed, and the class diagram was refined. These products show how objects interact internally and with the cooperating systems (e.g., databases to access and their architecture).

1.7.3 Design

The results of the analysis stage were then used to design a possible solution. The process between analysis and design was iterative until the design was complete.

1.7.4 Development

This stage was mainly achieved by coding. The analysis and the design made this stage and the related testing part relatively quick and smooth. Furthermore, the documentation related to the application was written during this stage.

1.7.5 Deployment

When the development process was completed, the system was deployed on the appropriate hardware and integrated with the cooperating systems.

1.8 Thesis Structure

The thesis is organized into three main chapters.

Chapter 1, "Elements for a PPMS GeoDB" describes what has been already done on the topic by different researchers. Considering what has already been attempted in management of some specific types of dangerous sites, the primary purpose of this part is to provide the requirements for a geospatial database that manages marine sites that are potentially polluting.

Chapter 2, "Geodatabase design and implementation" deals with the core component of this thesis: the database structure. This chapter can be divided in two parts:

- The first part explores the abstract design of the data model using a top-down approach, with some refinements that underline part of the model that has been judged of particular relevance or atypical;
- The second part is focused on the adopted specific encoding, briefly explaining the physical implementation using GML and evaluating a possible data storage solution by an XML-enabled relational database.

Chapter 3, "An example application: the Marine Site Risk Index (MaSiRI)" starts by explaining the reasons that motivated the decision to develop a risk assessment application. Subsequently, the chapter runs through each MaSiRI component, explaining the logic of such structures as the filtering criteria or the uncertainty estimation, and then shows the results of its application on an existing database.

The concluding chapter briefly summarizes the main parts of this thesis, focusing on the implications and proposing the adoptions of the design by the IHO as an S-100 series Product Specification.

Finally, several appendices are included that provide additional details about certain topics including:

- Corrosion in seawater;
- Legal aspects;
- Oil characteristic and related spill consequences;
- Metadata;
- The Unified Modeling Language (UML) and the IHO S-100 profile;
- The Extensible Markup Language (XML);
- The Geography Mark-Up Language;
- Building a GML application for a PPMS GeoDB;
- XML Data Binding;
- SQLite and the extension Spatialite.

CHAPTER 2

ELEMENTS FOR A PPMS GEODB

The definition of uniform criteria for data collection and risk assessment of PPMS is a crucial concept, and it represents one of the primary aspects of this study.

Ideally, the process of database standardization has to be performed at a global level in order to allow access to and the comparison of data from different sources. In addition, a common data structure can be useful both for exchanging information, and for integration of data. This has the advantage of permitting the common classification of inventoried wrecks, and to use established data indicators that enable comparable prioritization.

As such, this chapter provides a description of what has been done, with consideration about which elements or criteria are relevant to the PPMS GeoDB data model. In addition, key elements related to risk assessment and the legal framework are reviewed since they influence what fields should be included into the data structure.

The high costs of underwater activities require the definition of a scale of importance to shipwreck interventions. This prioritization activity is usually based on the localization of these sites within closed basins and highly trafficked areas and high environmental risk. In many cases when the risk mitigation operations are not judged necessary, a monitoring program should be organized, e.g., through the adoption of remote sensing techniques for raising an alarm. The availability of a series of periodic hydrographic survey data on the sites can permit a useful estimate of the preservation status of the site. However, before any risk minimization activity is performed, a survey of the wreck is often conducted to enhance the chances of success. For instance, the survey metadata contained in a PPMS GeoDB can aid in assessing the bathymetry and the backscatter mosaic of the area associated with the presence of a PPMS.

Risk assessment can influence the choice of what to include in the data structure since it usually takes into account several factors such as the type of pollutant, the amount, the corrosion rate, many environmental characteristics of the impacted area (e.g. sensitivity, biodiversity), etc. Hopefully, new and more specific future studies will support a better understanding of the relationships between the chemical and physical properties of some pollutants and their weathering processes in the marine environment. The definition of a common set of criteria for assessing the potential risk of pollution is useful to define which fields to include in the PPMS GeoDB Product Specifications.

From the point of view of International law, the impact of a future widespread adoption and application of the Nairobi International Convention on the Removal of Wrecks is widely recognized. Since this Convention applies only to current and future wrecks, the inventory of existing sites in a common data structure is desirable, and also useful to coordinate national legal frameworks.

2.1 Review of Related Works

Evaluating the entities required in a PPMS database is complicated by the diversity of objects to be represented. However, some important work has been conducted recently with the aim of cataloging shipwrecks by ocean/basin location. This includes the Secretariat of the Pacific Regional Environment Programme (SPREP) (Monfils et al., 2006; SPREP, 2002; Talouli et al., 2009) 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), the NOAA RUST database (Overfield, 2005), a global International Oil Spill Conference (IOSC) study in 2005 (Michel et al., 2005), etc.

Collectively, these have been analyzed in regard to the types of information that are fundamental for a PPMS GeoDB. 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 (Aichele, 2010; Overfield, 2005).

2.1.1 The SPREP Regional WWII Shipwreck Strategy and the AMIO Database

The SPREP is a large regional organization with the role of assisting Pacific island member countries to address environmental issues. It has developed a comprehensive program to address marine pollution from ship-based sources called the Pacific Ocean Pollution Prevention Programme (PACPOL), recently revised for the period 2010-2014. Based on the request of the members of the 12th SPREP meeting in 2001, SPREP developed a regional strategy to address the pollution risk posed by the many World War II shipwrecks across the Pacific (SPREP, 2002).

The leaking of the USS *Mississinewa*, in Ulithi Lagoon (Micronesia), focused the problem (Nawadra and Gilbert, 2002). It was caused by a tropical storm that passed in July, 2001 through the site causing an oil spill from the wreck that was not contained using a temporary plug until late August, 2001.

The first target of this strategy was to collect and collate a large amount of data related to the military and merchant vessels sunk or scuttled in the Pacific during WWII. The key activities for this part were (SPREP, 2002):

- Preliminary Site and Hazard Identification.
- Environmental Impact Assessment.

For the identification part, the steps followed in the data collection were:

- Collation of existing historical data of Pacific WWII wrecks.
- Identification of the ownership (sovereignty) of individual wrecks and in which jurisdictional responsibility the vessel is located.

- Confirmation of locations of sunken vessels in the region and production of maps of vessels in a GIS system.
- Confirmation of the identity of vessels and cargo quantities and types.
- Determination of whether any reports of previous oil releases have occurred from the vessels.

Although the SPREP database initially focused on the Pacific Island countries, it was decided that all information on Pacific–East Asian wrecks would be collated in order to make the database more complete (Monfils et al., 2006). A long process permitted expansion of the SPREP database to over 3,800 vessels lost in WWII (Gilbert et al., 2003; Monfils, 2005). The sovereignty of the largest number of WWII shipwrecks present in the SPREP database belong to the Japanese Government (with over 80% of vessels) (Monfils et al., 2006).

The PACPOL also published a generalized oil pollution risk assessment methodology that is designed to (SPREP, 2002):

- Determine what exists at each site.
- Determine what pollution incidents are likely to occur at each site.
- Determine what the likely impacts are.
- Formulate recommendations on mitigation and response measures.

It was evaluated that a better understanding of the consequences of any spill event can be derived by the location of the incident/vessel, how the oil/material behaves and weathers, the prevailing sea and weather conditions, and the sensitivities of the environmental resources it impacts upon. In particular, there is a need to undertake shoreline assessments of the areas under threat and determine the resources at risk and the nature of the regional environment.

In 2003 the SPREP member countries decided that the secretariat would take no further action on the strategy, and that any further developments had to be undertaken bilaterally between the flag state and the coastal state with SPREP only offering technical assistance. Essentially any remedial actions (e.g. removal of oil or environmental damages from WWII shipwrecks) would need to be carried out on a bi-lateral basis between the affected country and the shipwreck's flag state. According to protocol, this organization may only respond to specific calls for assistance from member states, i.e. the case of the *Hoyo Maru* shipwreck in Chuuk Lagoon, Federated States of Micronesia (Talouli et al., 2009).

The creation of the SPREP database spurred the creation of the Atlantic, Mediterranean and Indian Ocean (AMIO) WWII shipwreck database.

The aim of the AMIO database is to contain the WWII shipwrecks present in these areas. The Battle of the Atlantic represents a significant source of sunken merchant vessels with three thousand shipwrecks in the North Atlantic alone. The current database only contains information on vessels that were over 1000 tons. Submarines, however, were included as they also contain significant quantities of hazardous chemicals such as lead, acids and mercury.

The database details the location and ownership of over 3950 vessels, over 1000 tons, of which 529 are liquid product tankers that were carrying a variety of products including various crude oils, heavy fuel oil and/or refined fuel products; some were even carrying molasses. The oil tankers pose a higher risk due to the large quantities of oil carried on board at the time of sinking. Although 105 of them were known to be in ballast at the time of sinking and therefore pose little oil pollution risk except for that of its oil used as fuel, 181 had no information available (on whether in ballast and what cargoes they were carrying). Approximately 243 oil tankers have enough accurate cargo information available to make an estimate of the amount of oil on board (Monfils, 2005). The sovereignty of the largest number of WWII shipwrecks present in the AMIO database belong to the United Kingdom and the United States of America (Monfils, 2005). The SPREP and the AMIO databases combined have about 8,000 WWII shipwrecks worldwide (Figure 2).

Summarizing, the approach followed in the SPREP and AMIO databases are similar since both are focused on the need to acquire some information to efficiently manage potentially polluting shipwrecks:

- Vessel and Cargo Identification: the identity of the wreck represents the key to determine the flag state, the vessel type, the potential type of cargo and all the information available on the wreckage.
- Spatial component: the absolute site position can be used, e.g., to define the jurisdictional responsibility and the distance from marine resources at risk.
- Information on previous releases of pollutants: they are important indicators of the conservation status of the wreck.
- Sensitivity of environmental resources: e.g., shoreline assessment.
- Meteorological and oceanographic information: e.g., prevailing winds, water temperature.

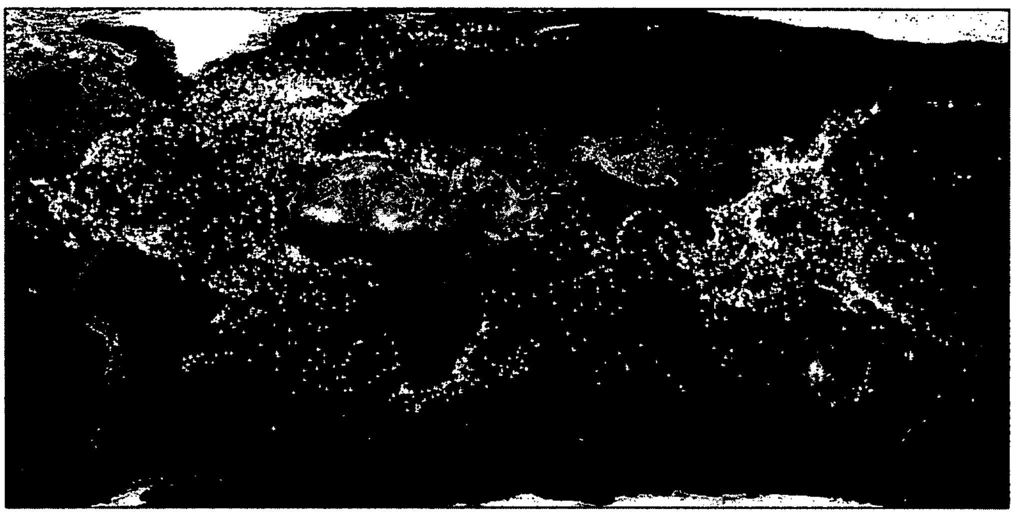


Figure 2 - World War II sunken vessels combining AMIO and SPREP databases for a total of about 8,000 records. The sovereignty of the largest number of them belong to the United Kingdom, the United States of America and Japan (Monfils, 2005).

2.1.2 NOAA's SHIELDS, RUST Database and RULET

The number of lost vessels in US waters is over 150,000 units. Even if the majority of them do not present a risk to the marine environment and to human safety, many vessels carrying hazardous cargo since the late 1800s represent a threat to the marine environment (Zelo et al., 2005).

The NOAA National Marine Sanctuary Program developed the Sanctuaries Hazardous Incident Emergency Logistics Database System (SHIELDS) with the aim of responding to incidents which have potentially significant impacts on a system of 14 underwater parks (Overfield, 2005). SHIELDS is a web-based tool to plan and respond to incidents in the sanctuaries using all relevant local resources.

The Resources and Undersea Threats (RUST) database is incorporated into SHIELDS that was originally established in 2003. The aims are to inventory and determine through analysis the scope of potential marine threats. The scope extends from the coastline of the United States to the outer continental shelf. The main idea of RUST came from the successfully identification of the SS *Jacob Luckenbach* as the source of beach and bird oiling on the central California coast: the identification was possible by the creation of a list of potential sources, from multiple databases, and through oil 'fingerprint' analysis.

RUST is focused mainly on locating and identifying all the underwater threats that may become sources of pollution such as abandoned wellheads, pipelines, platforms, sunken tankers, explosive ordnance, Atomic Energy Commission (AEC) and chemical weapon dumping areas, and also sites of historical and archaeological relevance (Aichele, 2010; Overfield, 2005, 2005).

The initial process was to create a national database through a collation of existing historical databases, identifying data gaps (location, cargo contents, vessel types) and the likely impacts of each site to the surrounding environment. In addition, historically or archaeologically significant sites were also identified to evaluate sensitivity to external threats.

RUST is a relational database with over 150 fields populated for each record. Some examples of this information are:

- The site position (known or reported position with position quality / accuracy information);
- The vessel type;
- The date of wreckage;
- The type of cargo;
- The site proximity to sanctuaries;
- The bottom type, etc.

Other than these information fields, each record has the potential to store images and documents (i.e., Hazardous Material Data Sheets). These data comes from federal, state, and private source databases, after a process of comparison and elimination of duplications. The maintenance of the unique attributes from the source database allows access to the information not stored in RUST.

The position of the shipwrecks supports display in GIS and visual interpretation, evaluation and spatial analysis for vector and raster formats. In particular, the site position allows a determination of the proximity to population centers, jurisdictional borders, and any other geographic point of interest. However, low positioning quality for some sites requires improvement with remote sensing devices (i.e., multibeam echosounders, magnetometers, etc.).

The RUST database is held behind a secure firewall and only designated users may have access to specific tables and fields, after the signature of a non-disclosure agreement. This is due to the sensitive information contained within the RUST database.

The RUST database also has a Risk Assessment query to provide a baseline risk value for each record. This evaluation is based on some selected fields (i.e., structural remains, position accuracy, fuel remaining, cargo type, potential threat to pollution, navigation and human safety) that return a numerical value by a pre-determined algorithm. Although the data entered are often subjective, the result is a base risk-level assessment for each record.

In 2010, NOAA received funds to develop a list of the most ecologically and economically significant potentially polluting wrecks in US waters. This project - called Remediation of Underwater Legacy Environmental Threats (RULET) - supports the U.S. Coast Guard both in prioritizing threats to coastal resources while assessing the historical and cultural significance of these resources.

Initial evaluations of shipwrecks (based on the vessel's age, type and size) found that approximately 570 wrecks of the total RUST data set could pose a substantial pollution threat. These include:

- Vessels sunk after 1891 (when U.S. vessels began being converted to use fuel oil);

- Vessels built of steel or other durable material;
- Cargo vessels over 1000 gross MT (smaller vessels would have limited cargo or bunker capacity);
- Any tank vessel.

Additional research narrowed the number of wrecks to 107 vessels due to the violent nature in which some ships sank and the structural reduction and demolition of those that were navigational hazards. In order to further screen and prioritize these sites, risk factors and scores are applied based on the likelihood of fuel and oil cargo remaining onboard. After this further evaluation, 23 wrecks are currently considered high priority based on their size and pollution potential, and 10 are known to periodically generate oil sheens (Helton et al., 2012).

RULET is based on two group of data fields (Vessel Particulars and Incident Information) reported in Table 1 (Helton and Symons, 2012).

Table 1 – RULET Data Fields.

Vessel Particulars	Incident Information
Name	Date lost
Owner and Flag State	Cause of Casualty
Vessel Type and Class	Vessel contents
Vessel Dimension and Tonnages	Wreck location
Builder	Distance to Shore
Year Built	Water Depth and Bottom Type
Hull Material	Resources at risk
Construction Diagrams	Historical Significance
Cargo Type and Capacity	Wreck Condition/Salvage
Bunker Type and Capacity	Dangerous Cargo/Munitions

The RUST database contains about 300,000 records which identify more than 7,000 sunken vessels scattered around the Atlantic, Pacific, and Gulf coasts (Figure 3), potentially waiting to release oil (Aichele, 2010; McGrath, 2011). The RULET project attempted to reduce this large number of sunken vessels to several tens and to prioritize them mainly based on the size and the pollution potential.

In summary, the NOAA RUST is a national database that collects some information similar to the SPREP and AMIO databases (e.g. vessel and cargo identification, site position) adding new interesting fields such as the proximity to marine sanctuaries or the bottom type. Furthermore, this database enlarges the threat types, e.g., abandoned wellheads and pipelines. Additionally, the RUST provides an attempt at risk assessment.

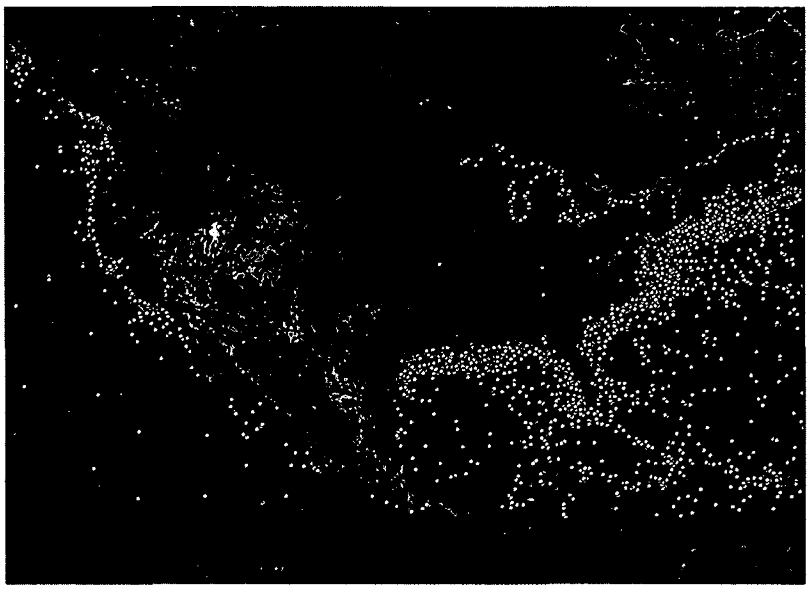


Figure 3 - Records in RUST Database (Symons, 2010).The RULET project attempted to reduce this large number of sunken vessels to several tens and to prioritize them mainly based on the size and the pollution potential.

2.1.3 The 2004 IOSC Study

The 2004 International Oil Spill Conference (IOSC) study was executed by an international group of expert with aims both to create a global database with the distribution of potentially polluting wrecks and to classify these wrecks on the basis of the environmental risk.

The first step of this study was to compile existing data into the first-ever worldwide database of potentially polluting wrecks called the Environmental Research Consulting (ERC)

International Marine Shipwreck Database. This database incorporates information from both the SPREP database and the NOAA RUST as well as other sources (Michel et al., 2005):

- The NOAA Automated Wreck and Obstruction Information System (AWOIS);
- The ERC's Oil Spill and Marine Casualty Databases;
- Northern Maritime Wreck Database;
- German World War II Maritime Shipwreck List (Schiffwrackliste);
- International Registry of Sunken Ships (accessible at *http://www.shipwreckregistry.com*);
- etc.

The data fields reported for each record are reported in Table 2.

Table 2 – Data fields present in the ERC International Marine Shipwreck Database.

Data Field	Description
Date	Date of sinking (or its best estimate)
Year	Year of sinking
Vessel Name	Name of the vessel
Vessel Type	Type of the vessel: passenger ship, tanker, fishing vessel, military vessel, etc.
Vessel Category	Tank (that is, vessel carrying oil) or Non-tank (oil as fuel/bunkers and for operations only)
GRT	Gross registered tonnage of vessel (same as GT or gross tonnage)
DWT	Deadweight tonnage as measure of the vessel's carrying capacity
Vessel length	Reported length of the vessel (used for estimation of vessel size when GRT or DWT are not available)
Vessel flag	State under whose flag the vessel was navigating at the time of the incident
Location	Description of the location of sunken vessel
Nation	Nearest nation to wreck site (not necessarily within the Exclusive Economic Zone)
Sea	Name of the relevant sea, ocean part, or larger waterbody
Latitude/Longitude	Best information about the position
Marsden Square	Location based on Marsden Square (a system that divides the world into grid cells of 10° latitude by 10° longitude, each with a unique, numeric identifier)
Cargo	Petroleum-oil cargo, if known or reported, aboard tanker
Cargo amount / unit	Amount of cargo and the type of unit reported
Cause of sinking	Reported cause of sinking
Depth of sinking	Reported depth of wreck

With the intent to consider those wrecks that may represent a significant pollution risk, the database only includes:

- Non-tank vessels of at least 400 GT holding petroleum-based oil as fuel/bunker.
- Tank vessels of at least 150 GT holding petroleum-based oil as cargo and fuel/bunker.

An interesting element of this database is the estimation method of the likely volume of oil remaining trapped in the wrecks since the volume of oil onboard is unknown for a large number of them. The followed analytical approach is based on the results of several existing works related to this topic (Etkin, 1999, 2002; Michel and Winslow, 2000; Rawson et al., 1998). Specifically, two estimates were calculated:

- High estimate: calculated assuming that a tank vessel had at least 80 % of its cargo capacity on board, while bunkers were assumed to be 70 % full.
- Low estimate: based on this two assumptions:
  - At the time of the wreckage, half of the vessels would have been 80 % full and half would have been 20 % full.
  - The 80 % of the oil would have either spilled at the time of the incident or seeped out in the following years.

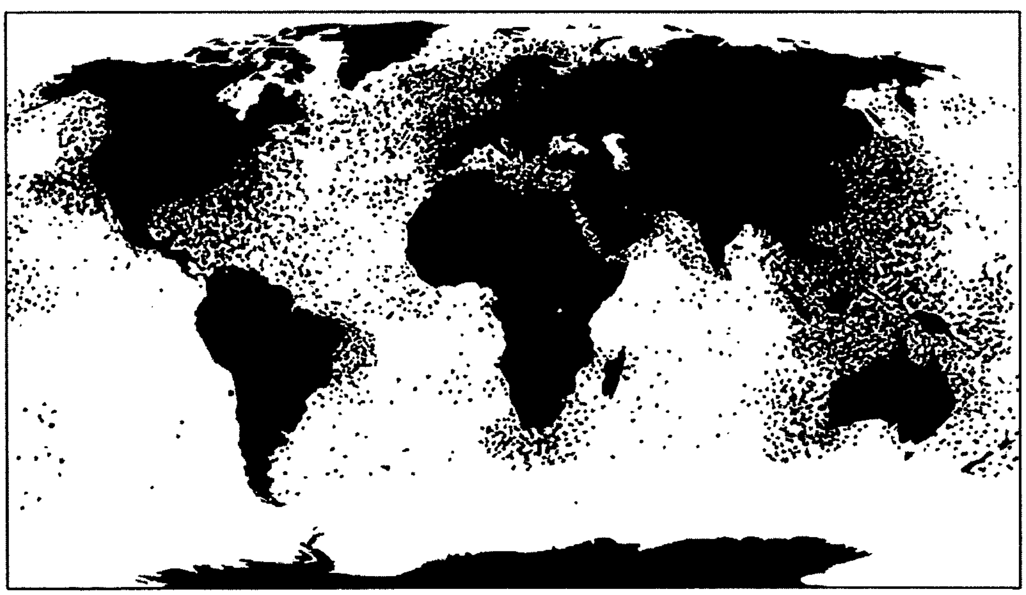


Figure 4 - The ERC International Marine Shipwreck Database includes more than 8,000 vessels sunk in the period 1890-2004 (Michel et al., 2005).

The resulting database includes more than 8,000 vessels (1,583 of them are tank vessels) sunk in the period 1890-2004 (Figure 4). Furthermore, a rough estimate of the total amount of oil onboard based on the cited criteria is between 2.5 and 20.4 million tonnes.

The ERC study focused attention on the vessels sunk during WWII. In fact, they are the largest group of potentially polluting shipwrecks (69 % of tanker incidents and 75 % of non-tanker incidents), and they are approaching the age where corrosion could create leaks or even structural collapse (Figure 5).

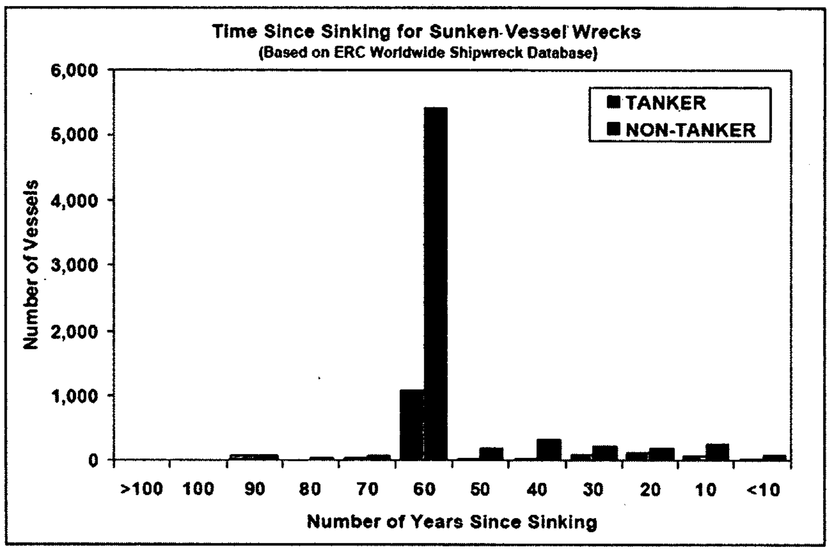


Figure 5 - Time since sinking for sunken vessel wrecks present in the ERC Worldwide Shipwreck Database (Michel et al., 2005). Since the reference year for the calculation is 2004, the peak is related to WWII.

The limitation of the ERC database is that it aggregates information from a large variety of sources used to record information for different purposes. The type of information in the source data sets was not necessarily collected to determine future pollution threats from the wreck. However, in order to deal with the lack of accurate information related to the amount of oil on board, authors adopted a meaningful analytical approach to estimate and extrapolate a range of possible values (Michel et al., 2005).

Another limitation is the lack of representation of non-tank vessels of at least 400 GRT and tankers of at least 150 GRT. In fact, vessels smaller than the ones selected may represent significant environmental risks when close to a tourist coastline or to a marine protected area.

Furthermore, there is the possibility of duplicate records when the incidents are referred to by different vessel names (Michel et al., 2005). A common example is the case of renamed sunken vessels that are sold to a new owner and then salvaged or scrapped. In this case, the vessel under the original name may remain in the database as an existing shipwreck.

2.1.4 The DEEPP Project and the Related Pilot Database

The objective of the DEEPP (DEvelopment of European guidelines for Potentially Polluting shipwrecks) project, co-financed by the European Commission, was to provide European coastal states national administrations with guidelines and criteria to face the environmental threats arising from potentially polluting shipwrecks (ICRAMM-CEDRE, 2007).

The first phase of this project was the creation of a pilot database, produced jointly by an Italian and a French organization (ICRAM and CEDRE), reporting the wrecks present since 1940 within the Mediterranean Cetacean Sanctuary "Pelagos", an area of about 90,000 km^2 in the north-western Mediterranean, habitat for breeding and feeding of many cetacean species but also characterized by intensive maritime traffic (Alcaro et al., 2007).

The main sources of information used to create this database are the Italian and French Hydrographic Institutes (Istituto Idrografico della Marina and Service Hydrographique et Océanographique de la Marine); a private agency, ARI (Agenzia Relitti Italia); the Italian Coast Guard; shipwreck books and dedicated websites. Additional information about the shipwrecks was collected consulting the Lloyd's Register books and the Italian Naval Register (RINA).

All the data were inserted in a Microsoft Access database using 50 fields divided into five categories: General, Characteristics, Location/Position, Tank-Cargo Risks, and More information (Table 3). Some of these fields are based on the classification proposed by a Regional Marine Pollution Emergency Response Centre for the Mediterranean Sea (REMPEC) publication (REMPEC, 2004). Various wreck records have missing information, most commonly related to the cargo.

Table 3 – DEEPP Pilot Database fields.

General	Characteristics	Location/Position	Tanks-Cargo / Risks	More information
Name	Type of wreck	Localization	Bunker Nature and Bunker Volume	Circumstance of shipwreck
Sea Area	Description	Distance to the coast	Bunker Volume left (low estimate)	Discovery organism
Wreckage Date	Building date	Distance from a sensitive area	Classification (low estimate)	Discovery method
Wreckage Year	Gross tonnage	Longitude and latitude	Bunker Volume left (high estimate)	Hydrodynamic conditions
Source	Length	Depth	Classification (high estimate)	Wreck status of conservation
Diffusion Rights	Width	Nature of sea bottom	Cargo Nature and Cargo Volume	
	Flag state	Localization accuracy	Cargo Class	
	Shipyard	Voice economic interests	Risk Factor (low/high estimate)	
	Route		Risk Assessment (low/high estimate)	
	Number of holds		Exacerbating factor	
	Capacity of holds		Evaluation risk	
			Risk description	
			Spill description	

Considerable effort was required to combine information from different sources into a reliable resulting dataset, especially in the case of conflicting data. As such, the result was 436 records (Figure 6), but the classification of 56% of them as 'unknown' permits the possibility that some close records could be duplicated. At the same time, the presence of about 70% of these records on seafloor shallower than 100 m, close to the coast, could be related to the ease of identifying these shipwrecks, both by acoustic surveys for safety of navigation executed by Hydrographic Offices and during fishing and recreational diving activities, and the presence of many traffic routes located near the coast or choke-points. The presence of more than 10% of bulk carriers and military ships, respectively, in the resulting dataset deserves particular attention as well as the probable advanced corrosion of the 50% of wreckage dating from the decade 1940 – 1950 (Alcaro et al., 2007).

Common problems reported in the DEEPP project about the database creation were:

- Need to combine all the available source of information, avoiding duplication but also producing the most comprehensive datasets;
- Various wreck information is difficult to recover, in particular data about the cargo such as the type and the quantity are of overall importance for the risk assessment.

The estimation of the bunker volume based on the gross tonnage represents one good solution in the absence of more detailed information, while the evaluation of the bunker volume remaining follows the IOSC methodology based on two percentages: 10% of the volume as the low estimate, 70% as the high estimate.

More generally, the selection of 50 fields used in the DEEPP pilot database contains more or less all the information present in different forms in the other existing databases.

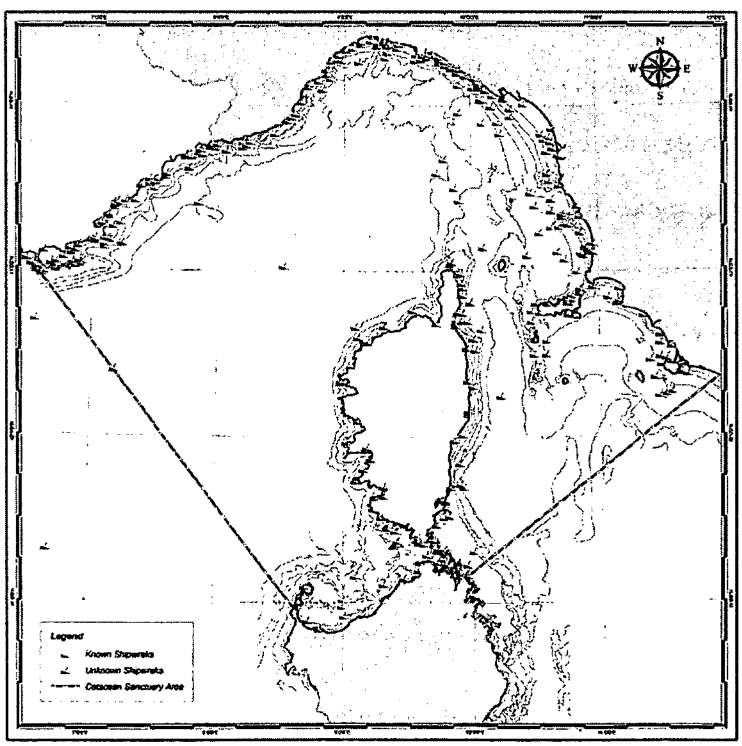


Figure 6 - Distribution of the shipwrecks present in the DEEPP pilot database with known shipwreck in red and unknown ones in black (Alcaro et al., 2007).

2.1.5 The Barrett Project

This project is based on a geospatial analysis of thousands of shipwrecks sunk during WWII in the Pacific theater (Barrett, 2011). The records come mainly from two sources that merge different data from the United States and Imperial Japanese Navy: the "Japanese Naval and Merchant Shipping Losses During World War II by All Causes" of the Joint Army-Navy Assessment Committee (JANAC, 1947), and "The Official Chronology of the U.S. Navy in World War II" (Cressman, 1999).

Oil tankers carrying large amounts of oil are important targets for naval warfare. In fact, the sinking of a single tanker could potentially disable hundreds of military vehicles. The Barrett database contains 3,100 wreck locations across the Pacific region, from 24-ton patrol vessels to the more than 60,000 MT of *Yamato* and *Musashi* battleships. Of all the records, 250 are oil tankers (about 10 %). The spatial and data analyses have been performed in ESRI GIS to address issues of environmental and geopolitical concerns.

Wreck locations are quite varied with some wreck sites close to the coast, while other vessels sank in remote parts of the Pacific. The actual conservation status of these wrecks is based on the amount of damage they received during battle as well as accumulated environmental exposure. With decades of corrosion weakening the ship's structural integrity, some important natural events like tsunamis, earthquakes and volcanoes - common in the "Ring of Fire" - may cause the structure to start to leak.

The main interest of the present research on the Barrett database is related to the use of a biodiversity index to evaluate the environmental sensitivity of a marine area.

This project underlines how certain areas like the Western Pacific are home to a great diversity of marine ecosystems (Roberts et al., 2002). For instance, coastal mangrove forests are particularly sensitive to marine oil spills, and loss of mangrove habitat may create a cascading decline in broader marine systems with decades-scale chronic effects. The Barrett project also defines a basic risk index to give priorities among the large number of records, taking into account the characteristic of the ship and the surrounding environment.

2.1.6 Integrated overall analysis

The analysis of the databases described in the previous sections highlighted:

- A common set of information that are largely present in a similar form within each of them;
- Peculiar attributes present only in one or in a few of them.

A common set of information is related to the PPMS (i.e., PPSW) characteristics such as, for instance, some constructive characteristics (e.g., overall length, cargo/tanker capabilities, date of wreckage) (Alcaro et al., 2007; Overfield, 2005). In this last specific case of PPSW, this could provide weaknesses in the data structure since the association between a site and a sunk vessel is often uncertain (Alcaro et al., 2007; Michel et al., 2005). Based on this evaluation, the solution adopted should have a distinct class representing a sunk vessel, associated (many-to-many relationship) to a shipwreck site. However, the presence of this information underlines that they are required in the development of a common standard for a database of this type. Information about the remaining amount of pollutant is also a key element for whatever risk assessment methodology is adopted (this point will be further discussed in the last chapter of this thesis). Since this volume is usually just an estimation, information about the constructive dimensions also becomes relevant from this point of view (Monfils, 2005). As a direct consequence, since the remaining pollutant volume is unknown for the majority of PPMS records, a valid method for risk assessment should provide a mechanism to represent not only a single value, but also a 'range of risk'. The estimation of pollutant volume based on vessel size also introduces the issue to provide a way to define a 'range of hazard', that is a low and a high hazard estimate (Etkin, 1999, 2002; Michel and Winslow, 2000; Rawson et al., 1998).

Generalizing and enlarging the reasoning behind the choice to separate PPSWs and sunk vessels, the adopted solution will cluster the required information as Potentially Polluting Marine Sites, marine resources threatened by PPMS, and complementary information (in some way related to the previous two groups and useful for future analysis on data) in order to better classify the type of information and to permit the application of common functions for data sharing similar characteristics. This solution represents also a way to make the data structure modular, so that a partial adoption is also possible without lack of consistency. This will permit more flexibility and avoid the implementation of a data structure that is for the most part not used.

A common problem in all the studied databases is represented by duplicate records (e.g., the incidents are referred to by different vessel names) (Michel et al., 2005). Any attempt to merge different databases in a common data structure may represent a dangerous opportunity for new duplications. There is not a global solution to this, at least, binary problem (that is, merging two specific databases). However, the availability of a common structure that uses uniform attributes (e.g., metric unit of measures) will reduce the risk of feature duplication or, at least, highlight the case of possible existing duplications.

Another type of information commonly present is related to the characteristics of the area surrounding the site, both in meteorological and oceanographic senses (ICRAMM-CEDRE, 2007). This information is less homogeneously represented than the constructive (site-specific) attributes. In fact, the related data are often spatially specific (e.g., based on the presence of an existing study in a region or at local scale). Any attempt at global proxies usually has the disadvantage of losing spatial resolution. At the same time, implementations based on specific hydrologic or meteorologic models could provide barriers to wide adoption of the developed database guidelines. The chosen solution is to provide a sufficiently generic class that represents a pathway for any existing model that may be available to the data collector. A similar approach is followed for the shoreline characteristics and for available geophysical surveys on the site's area.

Finally, the SPREP database provides the idea of a collection of information on previous releases of pollutants as important indicators of the conservation status of the wreck (SPREP, 2002; Talouli et al., 2009). This approach is adopted and developed as a goal of the target database. In fact, the presence of organized information in a time coherent framework may offer important statistic indicators (e.g., the conservation status) for wrecks and other types of PPMS.

2.2 The Marine Corrosive Environment

The decay of metals in contact with water (fresh, brackish and sea water) is due to an electrochemical process consisting of two partial reactions:

- The anodic reaction related to metal dissolution; and
- The cathodic reaction due to reduction of an element or compound present in the environment.

In seawater, the reaction which determines the rate of the overall process is the cathodic one. Furthermore, for iron and carbon steels, the corrosion products accumulate on their surface hindering the diffusion of oxygen towards the cathodic areas. As a consequence the global corrosion process gradually slows down and finally stops (Melchers, 2005).

Nevertheless, corrosion may continue if the SO_4^{2-} ion is reduced to HS^- through the intervention of sulfate reducing bacteria (SRB) (Melchers, 1999; Melchers and Ahammed, 1994). The action of SRB is possible only if the following conditions are fulfilled:

- Absence of oxygen at the interface between corrosion products and metal surface;
- Continuous supply of SO_4^{2-} ions; and
- Availability of organic nutrients.

Two different types of corrosions are mainly considered:

- General corrosion, affecting the stability of the structure of shipwrecks; and
- Localized corrosion that can cause the perforation of the walls of crude or fuel oil tanks as well as internal pipes and valves, with consequent leakage of these liquids into the sea.

The rate of localized corrosion was observed as higher than that of general corrosion, at least during the initial period. In fact, both rates tend to become equal after 1-2 years (Macleod, 2002).

On the basis of observation studies, a mean value of general corrosion rate of 0.05 - 0.10 millimeter per year may be assumed (Figure 1), but can significantly increase due to the effect of wave energy at the surface (Macleod, 2010; Schumacher, 1979; Southwell et al., 1976). Since the thickness of side, bottom and deck plates of ships built more than 40-50 years ago ranges from 1 to 4 cm, it is possible to forecast that a complete consumption of the plates will take place within a period variable between 100-800 years (Stephenson, 2005). These values are largely theoretical since the structure of a shipwreck will experience important weakness long before. In fact, it is not necessary that the structure needs to be completely corroded to collapse due to the effects of corrosion of local structures combined with the weight of the rest of the vessel.

Furthermore, local corrosion may take place some years before the complete collapse of the entire wreck, causing a continuous release of pollutants. While the hull plate thickness for most modern vessels is approximately 25 millimeters, internal pipes and tanks are substantially thinner and may collapse long before the main hull (Macleod, 2010; Schumacher, 1979). A catastrophic event – like a tsunami or earthquake – can also represent a causative factor for polluting release after years of corrosion. For this reason it is possible that the wrecks sunk during WWII could spill their pollutants content (bunker, cargo) during the next few decades (in the period 2013-2030).

Summarizing, the probability of a polluting release from PPMS tends to increase with time since the ocean is a corrosive environment for anthropogenic objects. Other significant processes include shifting sea-bottom sediments, destructive storms, marine currents, and the presence of marine bacteria, etc. Important considerations associated with shipwreck sites include damage and explosions suffered at the moment of the wreckage due to grounding, collision or severe weather. Further, military shipwrecks may have additional damage related to the explosion of ammunition (e.g., depth charges). For vessels that sank during World War II (WWII), several studies have focused on the state of preservation or the marine corrosion that has occurred during the past 70 years. Some vessels were damaged by fires and explosions prior to sinking, while others were deformed by water pressure, or weakened by partial burial into seabed sediments.

The type of construction, the length of immersion, the extent of burial and other chemical, physical and biological factors related to the anthropogenic object and its site influence the deterioration rate. Tropical shallow waters are often a rapid destructive environment due to the oxygenated waters, microbial attack, exposure to storms, etc. (Macleod, 2002). Conversely, deep waters have a tendency to decrease the destructive process due to the reduced presence of oxygen and the low temperature. Even if some characteristics of the marine environment slow down or accelerate the process, the moment of release of fuel, lubricants or any other pollutants will eventually occur unless there is human intervention.

It is evident that the time since sinking represents an important factor in the evaluation of spill risk in a PPMS. The sinking date, as well as the building date, represent key elements that have to be present to estimate the possible corrosion effects. Furthermore, the presence of other auxiliary oceanographic information, e.g., the water temperature, can be used to better calibrate the estimate. However, all of this information cannot ever be better than *in-situ* measurements of corrosion rate.

A more detailed discussion about marine corrosion is provided in Appendix A.

2.3 Sovereign Immunity and Environmental Responsibilities

When a polluting event happens, the responsibilities at the international level are still a subject of debate, particularly if related to WWII shipwrecks since a large part of these vessels can be classified as a 'state owned vessel'. In this case, the sovereignty of the shipwreck and all the property, cargo or content on board resides with the Government (Flag State) that controlled the vessel at the time of sinking.

The concept of sovereign immunity is present in Article 95 of the United Nations Convention on the Law of the Sea (UNCLOS), generally recognized from customary law, and is valid until the owner government expressly conducts an act of abandonment (Aznar-Gomez, 2010). Several nations, including Japan, the United States and the United Kingdom, have published policies on the protection, sovereignty and responsibility for WWII shipwrecks. For instance, Japan has consistently claimed sovereignty over its own wrecks to avoid any private salvage activities and has recognized the issue related to environmental risks (Johnson, 2008). At the same time, these sunken vessels may be restricted sites since they represent war graves (Masetti and Orsini, 2011). Further details about the general legal framework and a focus on sunken warships are provided in Appendix B.

It is less obvious if a Flag State has responsibilities for the damages caused by marine pollution related to the shipwreck. Currently, no nation has adopted a proactive approach to monitor potentially polluting shipwrecks outside their territorial waters. The recent events related to the *USS Mississinewa* shipwreck is an example of this approach. Following a cyclone that occurred in 2001, the shipwreck started leaking fuel in the Ulithi Lagoon (Micronesia). The U.S. Navy responded by conducting an environmental assessment and performing a complete oil removal. These operations were completed in 2003 (US Navy, 2004). However, the United States declared that this intervention did not set a precedent concerning its responsibility (Guerin et al., 2010).

The request for assistance on a case-by-case basis is – at best – a rather conservative approach. The lack of international agreement on how to deal with environmental risks related to PPMSs outside national waters allows these sites to corrode, and release their polluting contents into the environment. When a polluting event happens and the nationality of the wreck is known, the Coastal State usually requests assistance from the Flag State when it is often too late (Woodward, 2008). The adoption of a standard for PPMS inventory could help to highlight in time PPMS-related concerns, plan possible joint intervention and facilitate data exchange between the many international actors often involved.

More generally, only a few maritime nations have started to investigate the presence of PPMSs and the risks related to their presence in their territorial waters. For instance, the United States was able to identify the spillage from the *S.S. Jacob Luckenbach* as the *"largest killer of sea birds in North America"* only after several years of research and collection of oil samples, and this event was the impetus for the creation of the RUST database *"to find the next Luckenbach"* (Basta, 2010). Other countries, such as Japan, the United Kingdom and France, have also begun cataloguing their WWII wrecks.

The problem is not only related to WWII vessels. In fact, each single marine wreckage or local conflict can become a polluting disaster that may not occur until many years after the event. For instance, in 2004 the United Nations Development Programme (UNDP) reported the presence of more than 280 shipwrecks in Iraqi territorial waters. These were the result of three wars: the Iran–Iraq War (1980-1988), the 1991 Gulf War, and *Operation Iraqi Freedom* in 2003. Initially, this was considered an economic problem rather than an environmental risk, since dredging could not proceed safely, nor commerce resume, until these wrecks were cleared (Brown, 2005).

Many countries have adopted a passive approach to PPMS: they only face the problem after a shipwreck or any other significant polluting event occurs. This approach has resulted in increased risk for large coastal areas. Two examples include the US oil tanker *Montebello* off the coast of California (Hunter, 2002) and the ammunition ship *USS Richard Montgomery* in British waters (Johnson, 2004). These results are often not due to a technology limitation, but a failure on the part of the responsible parties to accept that these vessels could possibly pose a threat to the environment (Monfils, 2005).

At the same time, some countries are challenged by limited resources and expertise. In fact, the cost of oil removal depends mainly upon the amount of oil, the depth of the sunken ship, and the availability of equipment (Barrett, 2011). For instance, in the Pacific Ocean, small island developing states (SIDS) continue to suffer the risky presence of vessels sunk during WWII, more than a half-century later. In particular, SIDS are challenged by limited resources that reflect their economies. Their economies and cultures are largely dependent upon healthy seas for survival. Currently, there are two large development and conservation initiatives at work in the Pacific region: the already cited SPREP and the Coral Triangle Initiative (CTI) (Alino, 2012). About 1000 vessels (Figure 7) were found in the Barrett project within the Exclusive Economic Zones of CTI member countries (Barrett, 2011).

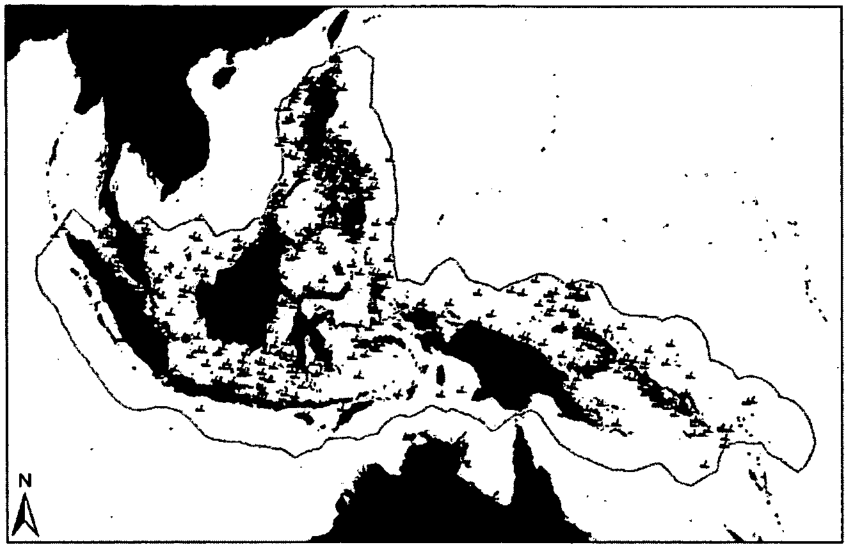


Figure 7 - Shipwreck records of the Barrett Project in the CTI states' EEZ (Barrett, 2011).

In March 2004, the Oil Pollution Preparedness, Response and Cooperation (OPRC) Technical Group of the International Maritime Organization (IMO) encouraged Regional Centers and Secretariats both *"to assess the situation regarding World War II wrecks that may cause oil pollution on their respective sea areas"*, and *"to develop regional co-operation on aerial and satellite surveillance"* (IMO, 2004). While a useful step, a more decisive precautionary action is required in order to avert any potential marine polluting event. In fact, dangerous events related to PPMS require rapid decisions and response by public authorities. Further, there is a need to develop a coordinated, multi-hazard, contingent plan in order to safeguard commercial, historical and cultural marine resources.

In this frame, the PPMS GeoDB may represent an efficient solution to standardize the data collection and, at the same time, to store information related to hazardous events in a homogeneous way. Furthermore, this brief description highlights the importance of the information related to:

- The definition of flag state;
- The eventual location of the PPMS in the national waters of a coastal state;

- The inventory of first-response facilities that, for instance, could be provided to neighbor states in case of incidents.

2.4 Database Entities

The phenomena of a 'universe of discourse' (that is, the set of entities that a model is based on) are perceived in the context of a geographic application which classifies phenomena into feature types (OGC, 2003). A feature type allows the definition of a multiplicity of feature instances. The process of creating feature instances from the universe of discourse is called data capture.

In the context of a PPMS GeoDB, all the subjects of data capture were clustered into three main groups that will be briefly described in the next sections:

- Potentially Polluting Marine Sites (PPMS);
- Marine resources that may be threatened by one or more PPMSs; and
- Complementary Information which represent auxiliary information useful to the management of PPMSs.

2.4.1 Potentially Polluting Marine Sites

Oil spills are an almost ineluctable consequence of producing and transporting petroleum and derived refined products. Chemicals are an important part of manufacturing valuable economic goods which can create dangerous situations in case of accidental release. Oil and chemical spills along coastal waters can harm the environment with potential widespread economic impacts. When facing oil or chemical spills, the responders' team must have a good knowledge about the expected fate and behavior of the spilled pollutant. The expertise required to properly manage these spills spans many fields - such as oceanography, meteorology and biology - in order to assess risks to coastal areas and habitats, estimate oil and chemical trajectories, analyze chemical hazards, etc.

A PPMS includes both shipwrecks of modern or contemporary vessels as well as other underwater sources of environmental risk such as pipelines, abandoned well heads, live ammunition and chemical weapon dumpsites, etc. Like shipwrecks, these types of sites deteriorate over time and may introduce pollutants into the marine environment. Very often the release of pollution from these sources is difficult to identify and attributed to passing vessels. At the same time, the occurrence of live ammunition and explosives can be life-threatening to fishermen, divers, and local residents. Thus, the entities to model the possible types of PPMS are heterogeneous: from submarines sunk during WWII to oil rigs. At the same time, the locations of these PPMSs are quite varied: from very shallow water close to important coastal areas to remote deep water parts of the oceans.

Since some of these entities are already present in a basic "safety-of-navigation" form in the IHO Registry, they will be enriched in this PPMS GeoDB with a series of new attributes and enumerations, mainly on the basis of the content of the existing databases reported in the previous sections.

The tanker vessels sunk during WWII are among the more dangerous since they were carrying crude oils, heavy fuel oil, refined fuel, molasses, etc. In particular, heavy fuel oil is one of the most environmentally significant threats due to the very slow biodegradation rates. A modern example of this issue was the oil spillage from the *Prestige* that was carrying high viscosity heavy fuel oil. On a statistical basis, the amount of heavy fuel oil still available for accidental release on board is 11 megaliters of oil per vessel (Monfils, 2005). The majority of WWII ships carried oils, such as Bunker C, whilst smaller vessels and submarines used diesel fuel (US Navy, 1946).

Furthermore, the different type of oil on the sea surface or present on the shoreline can be used as a proxy to identify the source of an oil spill (if this information is available in a database) or, at least, to exclude some PPMSs. The oil's chemical signature can be determined using comprehensive two-dimensional gas chromatography (GCxGC) (Nelson et al., 2006).

After a spillage, oil is submitted to different physical, chemical, and biological processes as a function of its characteristics (viscosity, density, and solubility) and related to the marine and coastal conditions (i.e. sea and air temperatures) (NRC, 2003). Bunker C easily loses the lighter compounds and its heavier residuals may travel far away by current and wave action (Mervin, 2000). It is also possible to have the formation of large areas of seafloor, called oil patch areas, characterized by the presence of these residuals. These areas represent not only hazards of themselves, but may periodically pollute the shoreline following, for instance, a storm (Parthiot et al., 2004; Pfeifer et al., 2008; Short et al., 2004). Thus, they will also be a feature type in the PPMS GeoDB data structure. A more detailed description of oil characteristics with a focus on the weathering process is provided in Appendix C.

The submarines of WWII also represent high risk vessels due to the significant quantities of hazardous chemicals usually carried (i.e., lead, acids, mercury). For instance, the shipwreck of German U-boat U-864 has released several kilograms of mercury a year into an important Norwegian fishing area (Olsvik et al., 2011).

Another source of risk are the large quantities of live ammunition, mines, chemical warfare agents (CWA), and other explosives present in a large number of WWII military shipwrecks. It is difficult to detect these materials since they are usually within a limited volume inside variable complex matrices (Valkovic, 2009). Just to have an idea of the problem scale, the amount of these CWA – mostly dumped deliberately – is over 300,000 tonnes in European seas, 4,900 tonnes in Japanese waters (on the basis of the indications received by US occupation forces), and 21,000 tonnes in Australia (Kaffka, 1996; Plunkett, 2003, 2003; Sato, 2010). Indications of large quantity of CWA are also present in the Baltic, Barents, North and Irish Sea (Aa et al., 2002; Beddington and Kinloch, 2005; Glasby, 1997; Leewis, 1991; Missiaen and Henriet, 2002).

This situation is the result of the past conviction that the dumping of CWA at sea was the best disposal method rather that to store them or incinerate them. For years, on the basis of the false assumption that the sea has unlimited absorptive capacities, old vessels were loaded with CWA and scuttled, or ammunitions were thrown overboard in designated areas. Many chemical/conventional weapons and nuclear waste dumping sites are the legacy of this policy (Overfield, 2005). Currently, an increasing number of injuries and problems related to these dangerous objects are being reported. In particular, the capacity of modern trawlers to reach depths of 1,500 meters has caused many contacts with dumped wasted material: artillery shells, phosphorous flares, cluster bombs, mustard and nerve gas, and phosgene-charged rockets. For instance, Denmark has reported more than 400 events of fishermen catching toxic materials, while Norway is still looking to locate some of the ships used to dump more than 168,000 tonnes of Nazi ammunition (Laurin, 1991; Simons, 2003). As a consequence of these dangers, many Baltic fishermen carry chemical protective equipment and decontamination gear during fishing activities.

Although the position of a large part 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 ensure the correct position during dumping operations (that permitted cost savings in time and fuel for the disposal companies).

The Deepwater Horizon disaster highlighted the dangers related to offshore extraction of hydrocarbons, and the loss of life and the environmental destruction will long live in the memory of the public opinion (Orth, 2011). However, periodic releases of water containing small amounts of oil from offshore oil installations (Espedal and Johannessen, 2000; Farmen et al., 2010) represent one of the major causes of marine pollution, together with large oil tanker or platform accidents (Fingas and Charles, 2001) and operationally from ships in transit since oil spill correlates very well with major shipping routes (Lu, 2003; Schwehr and McGillivary, 2007).

In addition to international regulations, there is also the need to monitor if regulations are followed. Although ship observations and aircraft surveillance are useful, the large coverage and the relatively low operational cost favor the use of satellite Synthetic Aperture Radar (SAR) for oil spill detection systems. This technology is largely independent of daylight and cloud conditions. The SAR principle is based on capillary and short gravity waves, which with moderate weather conditions are dampened by oil spill producing a dark slick in a SAR image. Several natural phenomena may also appear as dark slicks, thus aircraft are usually used to verify the presence and the extent of pollution, and eventually identify the polluter (Brekke and Solberg, 2005).

Abandoned and exploratory wells could also represent a threat for structural failure over time. As each year passes, the environment acts to further deteriorate them, increasing the risk of a significant release of oil and other chemicals.

To support these observed types of PPMS objects, the selected types for the PPMS GeoDB are:

- Potentially Polluting Shipwrecks (PPSWs);
- Pipelines;
- Oil Rigs;
- Oil Patch Areas;
- Dumping Areas;
- Wellheads.

It is evident that this list does not contain all the possible types of PPMSs, but it was evaluated sufficiently diverse and exemplificative. Other types of PPMSs may be added to the proposed data structure in the future.

2.4.2 Marine Resources

The management of a database of only PPMSs is enough if the only aim of a PPMS GeoDB was the inventory of sites that are potentially polluting. For a larger approach that permits management of the PPMSs from the point of view of the environmental risk, it is necessary to add several feature types that represent the marine resources that are threatened by the PPMSs.

An illustrative list of these types of objects containing the following feature types is implemented in the abstract data model described within the PPMS GeoDB Product Specification:

- Shoreline;
- Marine Protected Areas;
- Historic sites;
- Tourist facilities and accommodations;

- Fishing areas;
- Fish farms.

The shoreline represents the more generic feature type used as a marine resource. The main objective is documenting shoreline types in an area is to be ready in the event of a future release of oil or other pollutant. The shoreline is usually heavily used for subsistence fishing by the local population. Any negative effect on the coast can cause a disruption of the fishery, which in turn would affect the coastal state since a food source is disrupted. For instance, when the fishery was disrupted in Ulithi lagoon due to the oil leaking out of the *USS Mississinewa* the locals needed to find alternative food sources and many were plundering turtle egg nests thus having a far reaching ecological effect (Gilbert et al., 2003).

Information on shoreline types should be documented and photographs taken to illustrate the shoreline characteristics and possible response issues that would arise in the event of an oil spill. In order to assist with their planning and oil spill risk assessment, several stochastic spill modeling are available to simulate hypothetical scenarios for spills coming from PPMS (e.g. ASA OILMAP). These stochastic models give the probability of each shoreline being polluted, as well as the time to reach that location, in the event of a spill. This allows for prediction of where the oil spill response efforts should be focused.

But the shoreline in itself is not the only resource at risk in case of a pollutant spill. For instance, even when the coast is relatively far from the PPMS, the presence of a Marine Protected Area or a historic site in the surroundings may increase the risk related to a marine site that is potentially polluting. Furthermore, the presence of tourist installations that are always near to Marine Protected Areas can increase the relevance of a potential hazard in a coastal area.

An MPA can be generically defined as an area of sea – even if it may include land, seabed and subsoil under the sea – established by a national law mainly for the protection and maintenance of natural and cultural resources (JNCC, 2010). Thus the term encompasses a wide variety of areas with different approaches to marine conservation:

- Marine sanctuaries;

- Estuarine research reserves;
- National park, etc.

An MPA may address or contribute to addressing all of these objectives and may contain dedicated management zones permitting multiple uses, a combination of use and reserve zones, or reserve zones only (Ecology Centre, 2009). In fact, a MPA has usually the functions to protect ecosystems and to permit activities – like fishing, diving, or boating – allowing residents and other users to enjoy the area. In particular, potentially harmful activities are often managed through specific gear restrictions or through zoning/temporal schemes. More rarely, a MPA is a no-take area (usually called 'marine reserves') used to protect spawning or nursery grounds, as well as to protect ecologically important deep-water habitats. Furthermore, a MPA usually complements other types of management measures as fishery regulations and pollution controls. MPAs are garnering strong notice internationally. In 2003, it was estimated that worldwide there were 4,116 MPAs containing coastal and marine elements (Gubbay, 2004).

The link between marine environmental disasters and the fishing industry is clear. For instance, Deepwater Horizon caused significant economic harm to the Gulf fishing industry because of fishery closures and consumer concerns related to the safety of Gulf seafood (Upton, 2011). It was demonstrated that the rapid uptake of oil and polycyclic aromatic hydrocarbons (PAHs) by exposed fish and shellfish, resulting from large oil spillages in coastal waters, poses a potential threat to human consumers of fish and shellfish and also affects the marketability of catches (Law and Hellou, 1999). The presence of fishing areas and fish farms was judged as a good proxy for capture of this information in the environmental risk management.

2.4.3 Complementary Information

The presence of PPMS and related marine resources is often not enough to correctly evaluate the environmental risks. For instance, the corrosion rate, as described in Appendix A, is variously related to many parameters such as the seawater temperature or the depth at which the PPMS lies. At the same time, the spatial analysis to define in which marine area a PPMS is located represents fundamental information (for all the legal aspects outlined in Appendix B), as well as the knowledge of all the emergency response facilities available within the area surrounding a PPMS. Furthermore, the recent development of the hydrographic sciences allow the collection of much more information from an acoustic survey than just the bathymetry (e.g., backscatter mosaic, water column imagery, etc.) which may be useful in evaluation of risk, or mitigation exercises. Finally, the importance of meteorological and oceanographic data sources and spill modeling to determine the movement and fate of oil for spill response, during maritime emergencies and also for shoreline risk assessment has been widely published (Daniel et al., 2008; Gilbert, 1998). In fact, stochastic models are largely based on winds and current data to simulate an oil spill or other pollutant type from PPMS. When a site is close to the coast, the use of a tidal model is often necessary to obtain valid results.

For these reasons, a series of feature types called Complementary Information (i.e., necessary information that are not a PPMS or directly a marine resource) are added to the PPMS GeoDB data structure:

- Administrative areas (both national and international);
- Sunk vessels;
- Emergency facilities;
- Geophysical surveys;
- Raster geographic data;
- MetOc data.

2.5 Spatial Data Quality

A PPMS GeoDB product, as well as any spatial data, will be affected by the presence of imperfections (e.g., topologic inconsistencies, missing or incorrect attributes). These imperfections have a direct influence on the reliability of outputs coming from spatial analysis, and their removal by software may have the effect of masking the variability inherent in data capture procedures (Devillers et al., 2010). Current commercial GIS have only limited tools related to this issue. User awareness of this variability is possible in the PPMS GeoDB through improvements in the communication of the quality information of the provided spatial data.

Generally, the evaluation of spatial data quality is accomplished statistically. A classical example is the calculation of error ellipses and by error modeling based on collection of point measurements in overabundance. This approach can only partially solve the problem when, for instance, non-spatial information are coupled with the spatial feature. The arrival of GIS in the early 1980s focused on the integration of spatial and non-spatial data(Goodchild, 2010).

National mapping agencies made significant efforts to document spatial data quality by several quality elements:

- Lineage;
- Positional accuracy;
- Attribute accuracy;
- Logical consistency;
- Completeness;
- Semantic accuracy;
- Temporal accuracy.

For each element, several methods were proposed to evaluate the specific data quality of vector, raster, and Digital Elevation Model (DEM) data. Some approaches were developed to determine positional error for points, lines, and polygons, but models were also proposed to understand attributes and temporal uncertainties and their propagation (Devillers et al., 2007).

Users of geospatial information often mistakenly believe that data are perfect or they do not think about data quality (Goodchild, 2008). A big issue is related to how to communicate the data quality information from data producers to data users, often because lack of quality is perceived as a partial failure. The compilation of key metadata is a common way, although other possibilities are sometimes used:

- Different graphic visualization;

- Presence of warnings;
- Restrictions of responsibilities in licenses, etc.

Since the spatial data quality is mainly communicated through metadata, the metadata embedded in a PPMS GeoDB product represent a key to improve communication to end-users and to provide a succinct description of the content of a data set.

Even if the term "'metadata" is comparatively new, the "need to provide data about data" was always present in the scientific community. It is also largely recognized that metadata have the potential to exceed the data volume, and it is also not unreasonable to expect that as much effort be spent documenting data as in compiling them (Goodchild, 2008), even if this is rarely the case in practice. The idea of an automated update of metadata during PPMS GeoDB manipulation is clearly desirable, but this may be a largely elusive target for the data quality because of both the difficulties associated with processing different metadata implementations and gaps in knowledge of error propagation (Heuvelink, 1998; Lanter, 1994).

The generation of good metadata is generally time-consuming, but it can assist PPMS GeoDB users to understand the characteristics and the possible uses of a data set. An example may be the implementation of keywords and of a metadata search engine based on themes, spatial and temporal extents.

Many difficulties are related to this target for the PPMS GeoDB metadata. In fact, when the metadata provide a number (e.g. depth accuracy), the user can only make some assumptions on the process adopted to obtain this value (e.g., comparison with ground-truth data, and level of confidence). For these reasons, the ISO 19115 metadata standard presents as optional all the data quality statements (ISO, 2003).

A key concept in the metadata should be the definition of all the information users could need to combine data for a feature class from different PPMS GeoDB products (Devillers et al., 2010). The IHO Data Quality Working Group (DQWG) projects are taking a similar approach, providing to the users all the components of the quality metrics to let them use it as it suits them (Devillers et al., 2002; Dorst and Howlett, 2012; Harper, 2010).

Different studies raised several limitations related to the use of metadata. Metadata is only rarely consulted by users, and at other times only partially (Dassonville et al., 2002; Frank, 1998; Qiu and Hunter, 2002). This is due to the consideration common to many users that metadata are a technical description of a given spatial dataset, rather than understandable user-oriented information (Walford, 2002).

A PPMS GeoDB product is a description of the real world at some level of approximation and simplification. The PPMS GeoDB Metadata should fully document this process, explaining the data limitations and the adopted assumptions. Thus, metadata should permit any potential user to better understand the data, by evaluating the applicability for an intended aim and, afterwards, enabling its appropriate use. Furthermore, metadata could be used, by the same producer, for data management (e.g., storage, updating) and, by any user, for facilitating data discovery. A general discussion about metadata and issues related to the PPMS GeoDB's implementation is provided in Appendix D.

A target for the PPMS GeoDB metadata and, more generally, for an IHO S-100 Product Specifications is to provide the possibility to merge data collected by different organizations with slightly different methods to describe the features being modeled. In fact, the quality of semantics has large impact on any object, its categorization, its attributes and qualitative values, temporal and spatial properties. The lack of a common accepted terminology usually has the effect to name with the same term different concepts, or to have several names coming from different scientific communities to label the same idea. Some standardization bodies, such as the International Standard Organization (ISO) – Technical committee (TC) 211 and the Open Geospatial Consortium (OGC), have helped to reach a good level on general agreement about spatial data taxonomy. For instance, the ISO 19113 and 19114 standards focus on measurement and evaluation procedures of spatial data quality, and its documentation (ISO, 2002, 2003).

At the same time, it is central to the concept of fitness for use that it is necessary to create datasets and related metadata in such a way that permits avoidance of users' misuses. In fact, although the most technically perfect dataset may be created, it becomes useless, or even dangerous, when it does not fit the users' need (Devillers et al., 2010).

An important issue is represented by the existing distance between academic research and the day-to-day use of spatial data by users. In fact, a large part of the scientific knowledge in this field is still only embedded in scientific publications (Goodchild, 2008). The main commercial GIS software vendors only partially translate these findings into the functioning of software. This is also often due to the high level of complexity of the proposed statistics and modeling (Brown and Heuvelink, 2007; Levesque et al., 2007; Zargar, 2009).

The determination of the data quality becomes much more difficult in the case of a 'data mashup' (combining data from several sources) or integrated with volunteered geographic information (VGI) provided by web users. In fact, the Web is a largely interactive environment in which the accumulation of information from individuals is as important as the distribution of information to individuals (e.g., sites like *Wikipedia* or *OpenStreetMap*). Comments from previous users are readily accessible, and can be searched by potential new users. Instead of the highly structured, formalized and producer-centric approach of the 'classic' existing standards, the Web 2.0 is focused mainly on almost unstructured and informal commentaries. Concepts like 'meta-uncertainty' (i.e., uncertainty about uncertainty) and 'uncertainty absorption' were introduced to describe the financial risks associated with using this type of spatial data (Bedard, 1986). Even if this thesis does not take into account this type of information, it is evident that a PPMS GeoDB that was publicly available could receive useful comments from web users (e.g., recreational divers).

Decision makers are mainly interested in solving problems using spatial data as the information contained in a PPMS GeoDB product, but only the availability of data quality information allows assessment of the reliability of the decisions they are taking. Spatial data do not have to be perfect; they have to be good enough to be used to answer some questions. The ease of use permits direct interaction with the data without having to learn a query language (e.g., SQL) or to request the support of a GIS specialist. Incorrect deductions or interpretations of datasets may lead to wrong decisions and adverse consequences.

Several approaches have been proposed to help end-users assessing the fitness for use of a dataset, and to reduce the risk of spatial data misuse. They can be divided in *a priori* (before they arrive to the end-users), and *a posteriori*. The former requires identifying the appropriate uses of a given spatial dataset and restricting the usage to these. The latter looks at improving GIS software capabilities to inform users about the spatial data quality information: for instance, by different color visualization, by opacity or particular texture for objects with poor quality, an additional data 'dimension' for quality, end-users warnings in case of illogical operations, etc. (Beard, 1997; Devillers and Beard, 2006; Hunter and Reinke, 2000; McGranaghan, 1993). A different point of view is represented by the development of a GIS capable of offering functionalities for management, update, assessment and communication of data quality. Different names have been adopted for describe them: error-sensitive GIS (Unwin, 1995), error-aware GIS (Duckham and McCreadie, 2002), or quality-aware GIS (Devillers et al., 2007).

The PPMS GeoDB adopts an *a priori* approach based on some legal considerations described in the next paragraph, leaving the door open to future different solutions.

2.5.1 Legal Considerations

Any spatial data producer has the role and responsibility for the prevention of potential misuse, showing due care and diligence. The main duty is properly informing users, to communicate information included in the metadata or any other description of the data useful to understand the characteristics of a given dataset. Such a duty is strongly related to the context, the nature of the product, and mainly the knowledge of the user. In particular, in a context of mass consumption, the duty for the spatial data producer to properly inform users is particularly high due to their possible lack of expertise in spatial referencing.

The duty to inform is usually accomplished by advising and warning. If informing does not necessarily require directing user decisions, the duty of advising is characterized by the need to give a judgment on the content of the dataset (Levesque et al., 2007). The duty of warning is related to the evaluation of the presence of potential dangers in a dataset. International standards have been developed on warnings, even if they do not fit perfectly the geomatics community needs (ISO, 1999, 2004).

The ISO 3864-2 standard states that safety signs and safety labels should communicate the following four elements:

- The overall danger level of the risk (characterized by an alert word like, for instance, 'Danger', 'Warning', or 'Caution');
- The nature of the risk (often illustrated by a symbol);
- The consequence of interacting with the risk; and
- Indications on how to avoid this risk.

The duties of advising and warning also implies the duty of identifying and revealing potential risks related to the use of a given spatial dataset (Gervais, 2004). These duties are particular relevant for the PPMS GeoDB products due to potential environmental damages related to their misuse.

2.5.2 Risks Related to the Creation and the Use of a PPMS GeoDB

ISO/IEC Guide 51 provides guidelines for the inclusion of safety aspects in standards. In this guide, the risk is defined as the *"combination of the probability of occurrence of harm and the severity of that harm"* (ISO, 1999). For spatial data quality aspects, it was proposed to consider the term *"harm"* as a data misuse or misinterpretation (Agumya and Hunter, 1999; Levesque et al., 2007).

A risk-based approach based on ISO/IEC Guide 51 was proposed for the development cycle of a spatial multidimensional database (Levesque et al., 2007). A similar approach could be adopted for a PPMS GeoDB. The approach provides a continuous and iterative process for risk assessment and reduction that goes from need analysis to implementation and 'feeding' of the PPMS GeoDB (Figure 8).

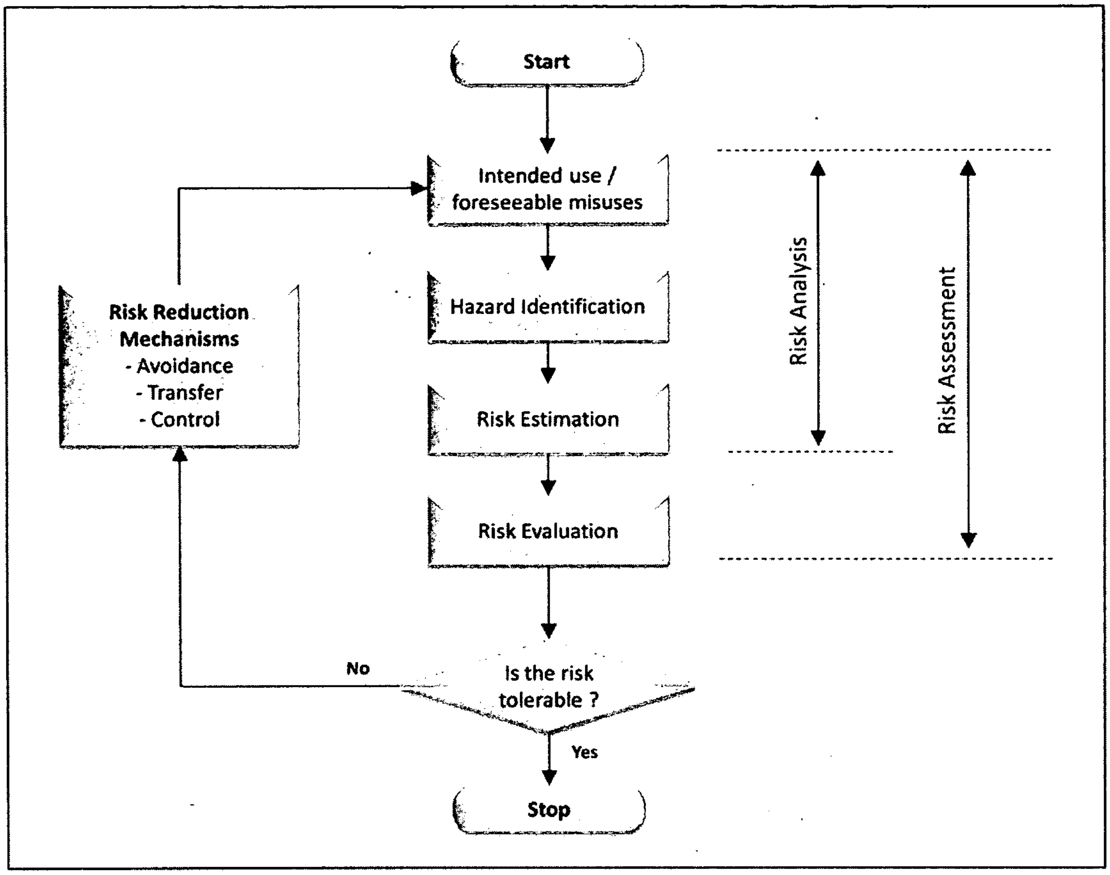


Figure 8 - Iterative process of risk assessment and risk reduction (adapted from (ISO 1999)). Possible risk reduction mechanisms are: avoidance of some groups of users, risk transfer to another party, control by preventive actions.

In this process, the first two steps are related to the risk identification that is the research and the inventory of any possible misuse of a given PPMS GeoDB during the analysis and the interpretation of its spatial dataset. After that, the risk estimation process is accomplished by estimating a probability of occurrence and severity of consequences for each misuse or misinterpretation. Independently to the techniques used, these two criteria should be clustered in a limited number of levels (for instance, three levels such as low, moderate and high). The assignment at each level is not a simple task, and it is usually obtained using experience, judgment, and intuition, taking into account the skills of possible end-users.

The outputs from the risk analysis can be combined to obtain the overall level of danger related to the risk (risk evaluation), usually using a hierarchization matrix based on the probability of occurrence of the risk and the severity of the consequences for the PPMS GeoDB user.

The diagonal of the matrix can be used as a reference for moderate risk (Table 4) or, in case of a more careful behavior, for high risk.

Table 4 – Example of hierarchization matrix. Adapted from (Kerzner, 2009).

		Probability of occurrence		
		Low	Moderate	High
Severity level	High	Moderate	High	High
	Moderate	Low	Moderate	High
	Low	Low	Low	Moderate

When each risk related to any identified misuse or misinterpretation of a PPMS GeoDB product is evaluated, the data producer can decide if it is acceptable (and avoid any effort to control it) or unacceptable. In this last case, the producer is required to apply some mechanisms of risk reduction:

- Avoidance: some category of end-users are refused;
- Transfer: a data producer transfers the risk to another party (e.g., a geomatics professional, an insurance company, etc.);
- Control: a data producer take preventive actions in order to reduce the risk (e.g., implementation of some integrity constraints in the database, introduction of warnings to communicate a risky query to the end-users).

Alternatively, a part of the risk evaluation process can be shared with the end-users in such a way they can decide to use the data only if they judge the risk acceptable. This last procedure is based on the assumption that the user plays an active role in the risk reduction procedure, after the database design from the data producer (Figure 9), but it will always present some residual risk (Bedard, 1988).

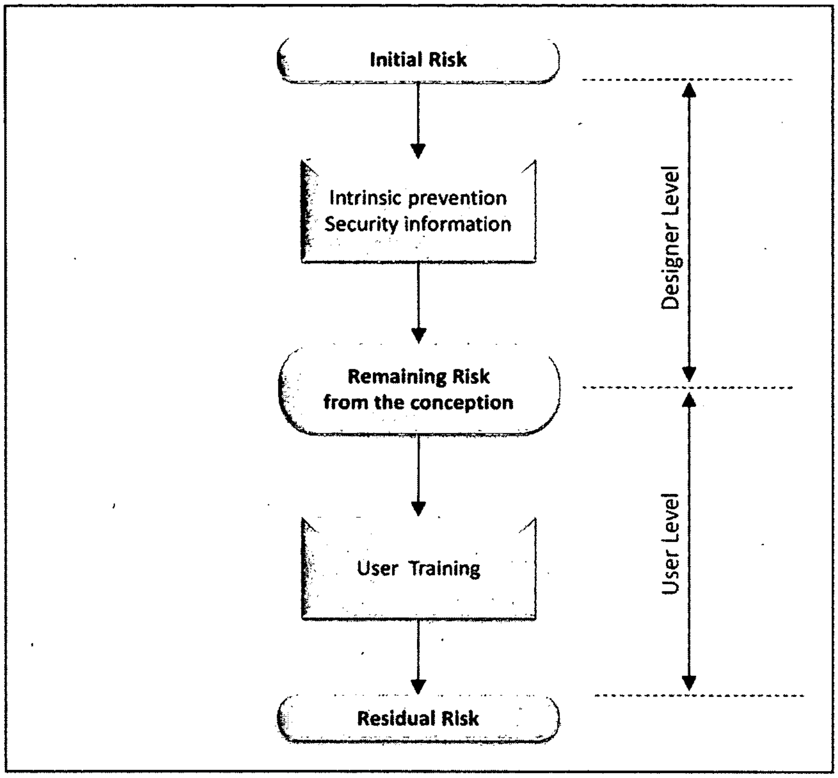


Figure 9 - Risk reduction, adapted from (Levesque et al., 2007). Also when the user plays an active role in the risk reduction, a residual risk is always present.

A possible additional step in the risk reduction process is represented by a risk audit. It mainly consists of rigorously documenting all the previous steps, focusing on how outputs from a given datasets are presented to the end-users in order to have an important source of information for developing future datasets, but also to prove – from a legal standpoint – they complied with their legal duties.

The choice of which approach to follow to define the risk related to the use of a PPMS GeoDB product is left to the product maker.

CHAPTER 3

GEODATABASE DESIGN AND IMPLEMENTATION

Now that the elements required for a PPMS GeoDB are described, this chapter focuses on the design of the data structure both as a conceptual model and a physical implementation.

To ensure standards compatibility, the database was developed based on the International Hydrographic Organization's S-100 approach (IHO, 2010), while also providing generic descriptions of various potential polluters. The conceptual model was developed using the Unified Modeling Language (UML) as required by the IHO S-100 framework. In Appendix E, a general overview of UML and a detailed description of the IHO S-100 profile adopted in the development of the conceptual model are provided.

Since the conceptual model is platform- and encoding- independent, the second part of this Chapter starts with the reasoning for the choice of the Geography Markup Language (GML) as encoding format. It then describes the related physical implementation of the PPMS GeoDB abstract model. In Appendices F and G, brief theoretical presentations of the Extensible Markup Language (XML) and GML - that is a 'geographic dialect' of XML - are provided. However, the selection of the GML format for a possible future adoption by any user is only suggested, not mandatory.

The last part of the Chapter addresses the practical problem of efficiently storing the information collected and exchanged using the PPMS GeoDB Product Specification. Since the XML database technology was judged still to be immature, a solution based on a GML-enabled relational database is presented. This solution is based on the combination of the XML Data

Binding technique and a spatial extension of the popular relational database management system (RDBMS) engine called SQLite.

3.1 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 10).

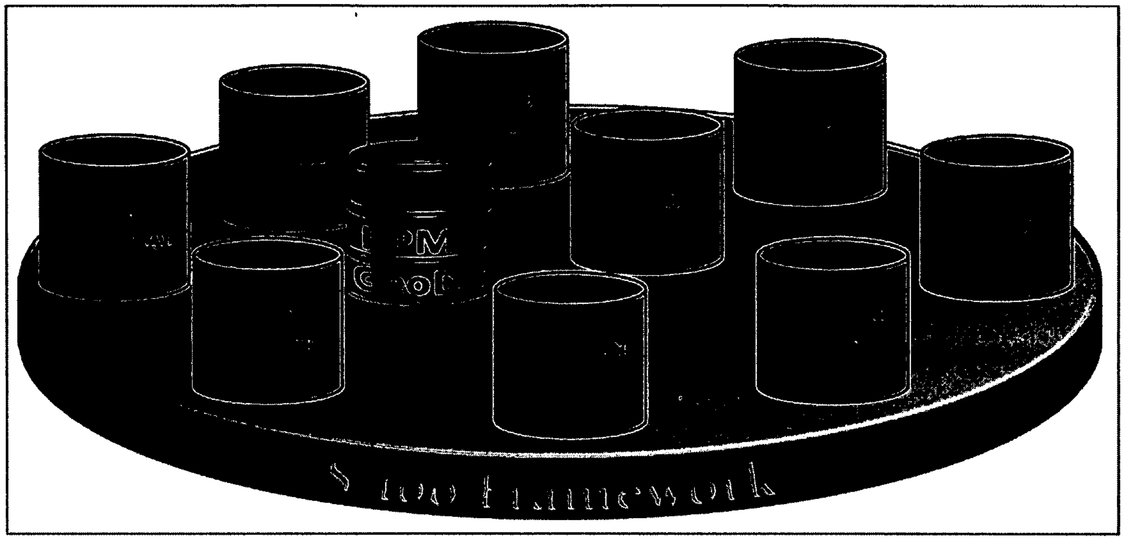


Figure 10 - S-100 framework with PPMS GeoDB among some other future S-100 series products. Developing within the S-100 framework allows the GeoDB to adopt already developed resources (simplifying implementation) and present its data in a common framework (simplifying adoption).

A principal reason for this is the potential to adopt some of the geographic features already present in the existing S-100 Feature Concept Dictionaries into the developing data structure. These features have been created for some of the upcoming 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 an XML Product Specification. The complete PS is provided as an Annex to this thesis as well as being available on the website related to the present project.

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

- Definition of a vector-only product.
- Selection of required features, feature attributes, and enumerates contained 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 are then described in a Feature Catalogue, and geometry types required in the product are determined. New geometry types will not need to be added to the S-100 framework for the proposed PS.

At this point, it is possible to construct an information model required by the developing application. This model is commonly called an Application Schema (ISO, 2005). The creation is 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).

3.2 The Conceptual Model

In the proposed PPMS GeoDB Product Specification, any product has a root element instance of the `Root` class. This root element is used to carry information that is shared among all the descending data. Furthermore, this element contains a link to the highest level of metadata information.

This root element may be related by composition with three types of composite Feature Collections (Figure 11). Further, each collection inherits from an abstract class in which all the shared characteristics between the different features of that collection are defined. 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.

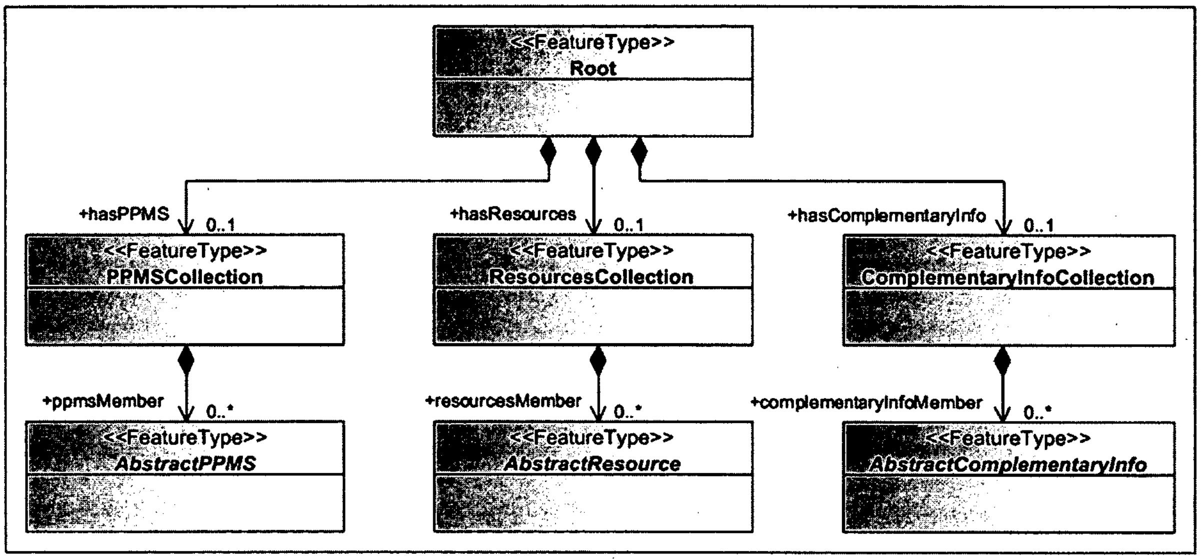


Figure 11 - Relationships of *Root* class. The GeoDB consists from zero to three collections (PPMS, resources and complementary information), as required by the applications for which it will be used. Note that each collection includes an unlimited number of features of common abstract type so that common methods can be applied that are useable on all features within the collection.

In the next paragraphs, the content of each Feature Collection is described from a general point of view. More details are present in the PPMS GeoDB Product Specification both as UML diagrams and context tables provided as an Annex, and available from the project URL: `http://ppms.ccom.unh.edu`.

3.2.1 The PPMS Entities

The entities to model the possible types of PPMS are heterogeneous: from submarines sunk during WWII to oil rigs (Figure 12). Since some of these entities are already present in a basic 'safety-of-navigation' form in the IHO Registry, they are just enriched here 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 the Regional Marine Pollution Emergency Response Centre for the Mediterranean Sea (REMPEC, 2004). The attributes included from the existing IHO Registry are easily recognizable in the PS since they have a different namespace (e.g., `hydro`)

The `AbstractPPMS` class is defined by the following six concrete Feature Collections, each of them grouping by type an unbounded number entities:

- `PPSWCollection` class;
- `DumpingAreaCollection` class;
- `WellHeadCollection` class;
- `PipelineCollection` class;
- `OilPatchAreaCollection` class;
- `OilRigCollection` class.

Any member of the above Feature Collections has as common super type the `AbstractPPMSFeature` class.

The `AbstractPPMSFeature` class has a series of attributes inherited by all the PPMS feature implementation classes and associations with some feature classes related to the `ComplementaryInfoCollection` class (Figure 13). These last associations are useful to provide such relevant geographic information as:

- The location in a specific International area (the importance of which has been extensively documented in the first chapter and in Appendix B);
- The distance from emergency facilities;

- The availability of such supplementary information as bathymetric surveys on the site or records of meteorological and oceanographic data.

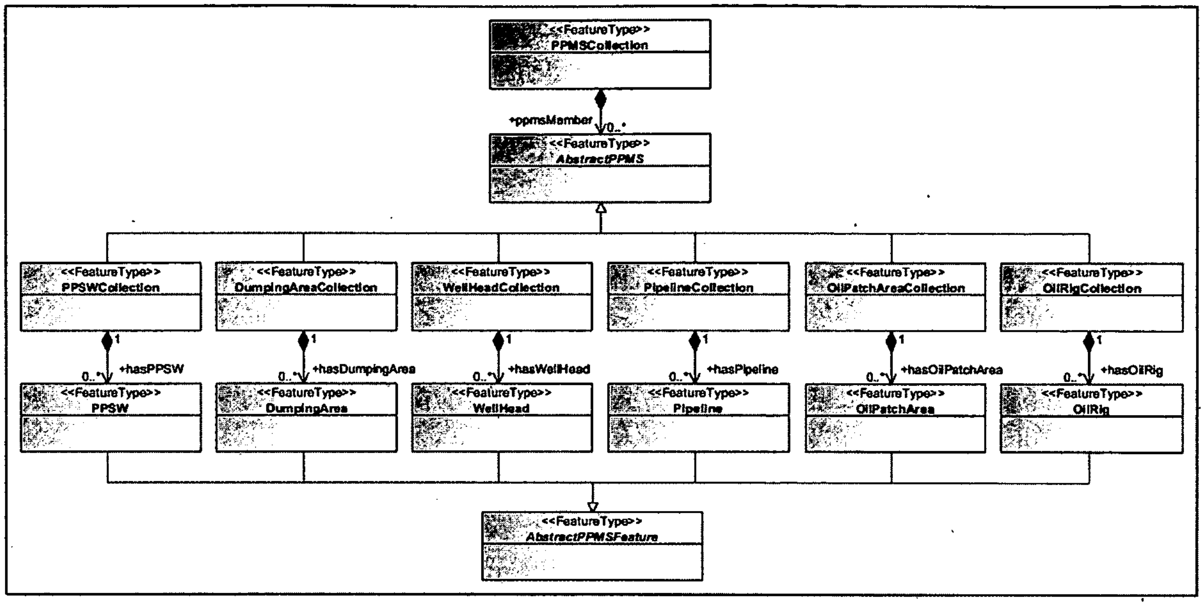


Figure 12 - Sub types and relative relationships of the `AbstractPPMS` class. All of the specializations of a PPMS derive from the Abstract class to allow common methods to be defined, but each specialized PPMS augments the resources maintained to provide information specific to the object being represented.

Furthermore, it was judged helpful for many reasons (e.g., reducing the amount of duplicated information, providing an intuitive data structure, etc.) to subset the initial set of this abstract element in order to create a new feature, complementary but independent, called `EventPPMS` (Figure 13). This class collects the information that should be commonly updated in the case a new "event" happens to the PPMS. The criterion followed for the demarcation of information is whether the nature of the data for a given attribute is judged to be more static or dynamic. An `AbstractPPMSFeature` instance may therefore have an infinite number of `EventPPMS` instances. The value related to the risk index is part of this latter class due to its inherent dynamic nature. Furthermore, the list of `EventPPMS` associated to a given site forms a meaningful history that can be used to track its behavior through time integration.

Each PPMS feature implementation class also presents additional attributes and relationships.

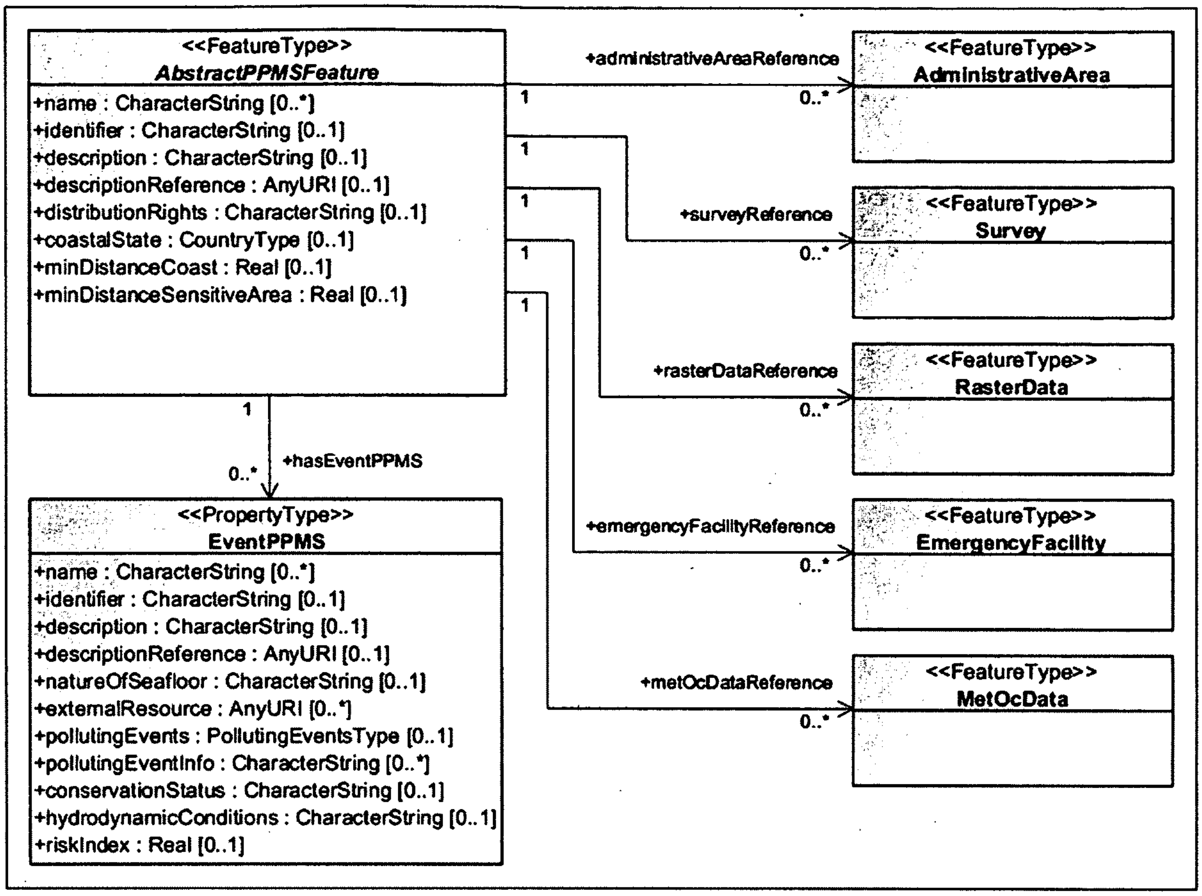


Figure 13 - Attributes and associations of the `AbstractPPMSFeature` class.

As an example, Figure 14 outlines attributes and relationships proposed for Potentially Polluting Shipwrecks (PPSW). This class may have different optional attributes. Most are derived from the `hydro` domain, which is already present in the IHO Registry (IHO, 2012). The limited number of additional attributes will become part of a specific domain of the IHO Supplementary Dictionary.

During the modeling process, many problems have been identified 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 15, top), or a sunken vessel can be associated to many possible wreck sites (Figure 15, middle). In some cases, a site inspection (e.g., by diver or ROV) is required to resolve uncertain associations (Figure 15, bottom). The many-to-many relationship between `SunkVessel` and `PPSW` classes is the solution adopted for this particular problem (Figure 14), since it allows for expression of the uncertain association of ships and sites.

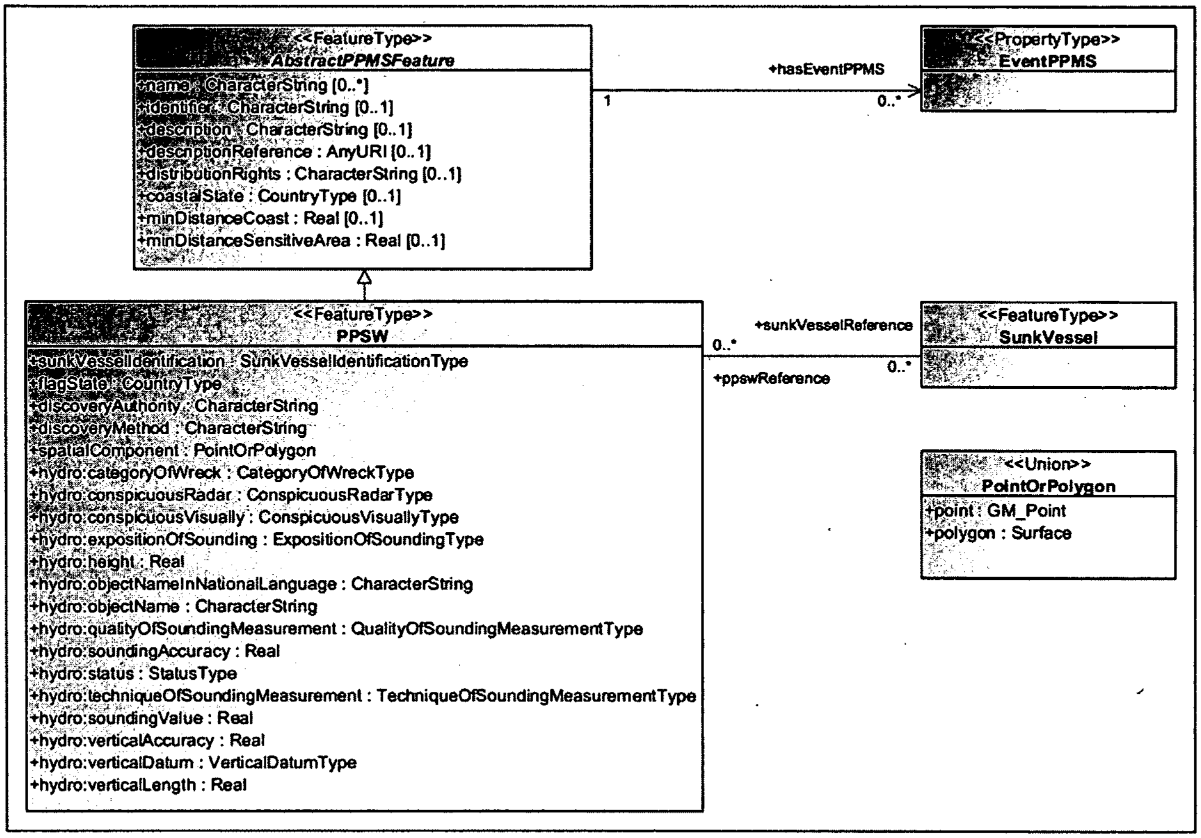


Figure 14 - Attributes of the *PPSW* Class derived from the `AbstractPPMSFeature` Class with the detail on the `PointOrPolygon` union used for the spatial component. Note particularly the many-to-many relation between the `SunkVessel` and `PPSW`, expressing the possibility that any one `SunkVessel` might 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.

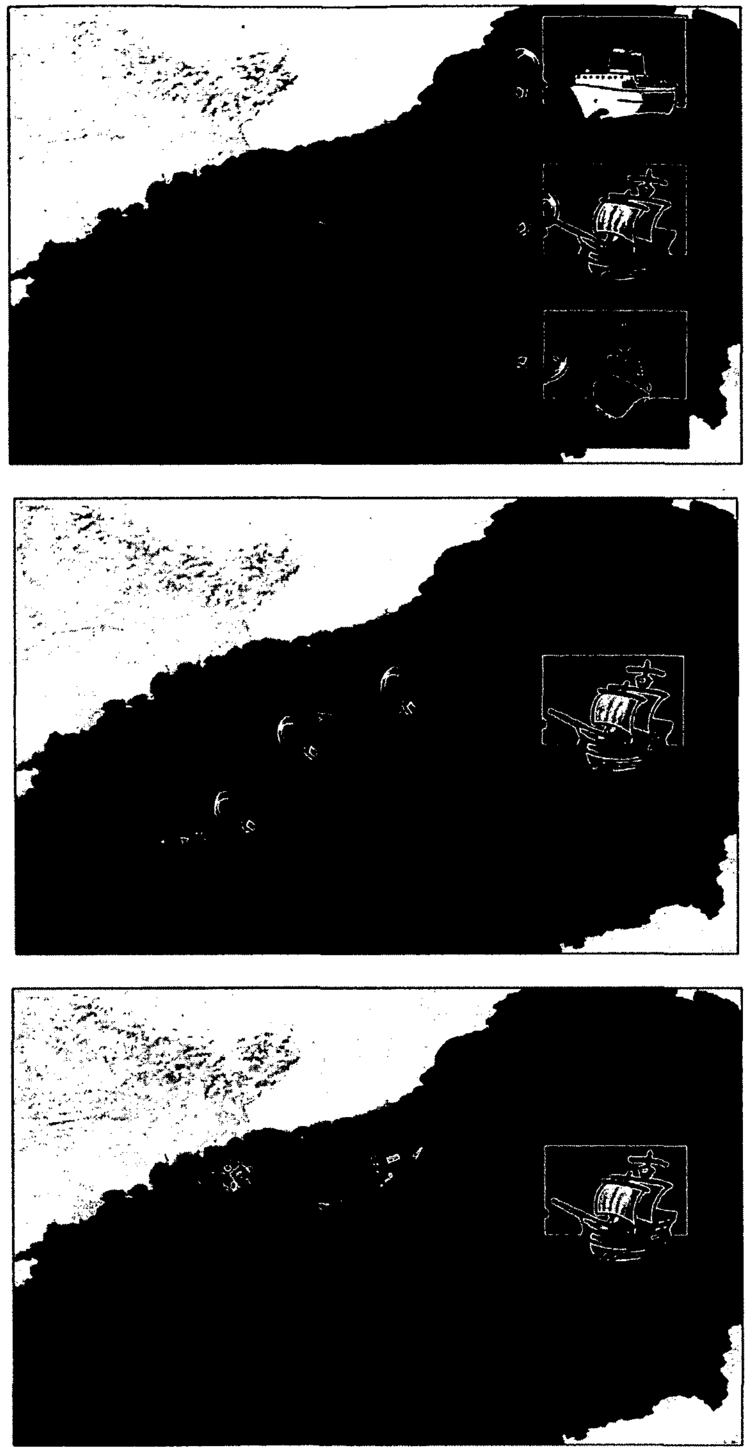


Figure 15 - Examples of different possibilities of the many-to-many relationship between `PPSW` and `SunkVessel` classes. Many shipwrecks may be associated with one PPSW (top) if the provenance of the wreck is not known, while one shipwreck may be associated with many PPSWs (middle) if its location is 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.

For this class, a spatial component represented by a union of a point and a polygon was judged sufficient, while other classes make use of a line as a possible geometric primitive. The adoption of these three simple vector geometries (point, line and polygon) is related to the limitation provided by the IHO S-100 data model as well as based on the consideration that they are the most widely adopted and used in commercial GIS. However, the introduction and the adoption of different and more complex geometries is not a limiting factor for future development of the PPMS GeoDB Product Specifications.

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, and tourist installations, but are not strictly objects in the PPMS sense. As such, they are organized in two related groups: `ResourcesCollection` for marine resources directly or indirectly related to the PPMS, 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 inventory the implementation of these entities may be simply ignored, a specialized application – such as, for instance, risk assessment – will typically require them to be fully populated.

3.2.2 Marine Environmental and Economic Resources Entities

Marine resources directly or indirectly related to the PPMS are described under this group of entities. Their aim is to store information related to possible 'targets' useful to define vulnerability and exposure in the Risk Index evaluation process. Some examples of these instances are: shoreline segments (with the related property of population density, land use classification, etc.), archaeological sites, fishing areas/farms, marine sanctuaries, tourist installations, oil spill fighting authorities, etc.

The cited `AbstractResource` class is defined by the following six concrete Feature Collections (as for the `AbstractPPMS` class, each Feature Collections groups by type an unbounded number of entities):

- `FishingAreaCollection` class;
- `FishingFarmCollection` class;
- `MPACollection` class;
- `HistoricSiteCollection` class;
- `TouristInstallationCollection` class;
- `ShorelineCollection` class.

Any member of the above Feature Collections has as common super type the `AbstractResourceFeature` class (Figure 16).

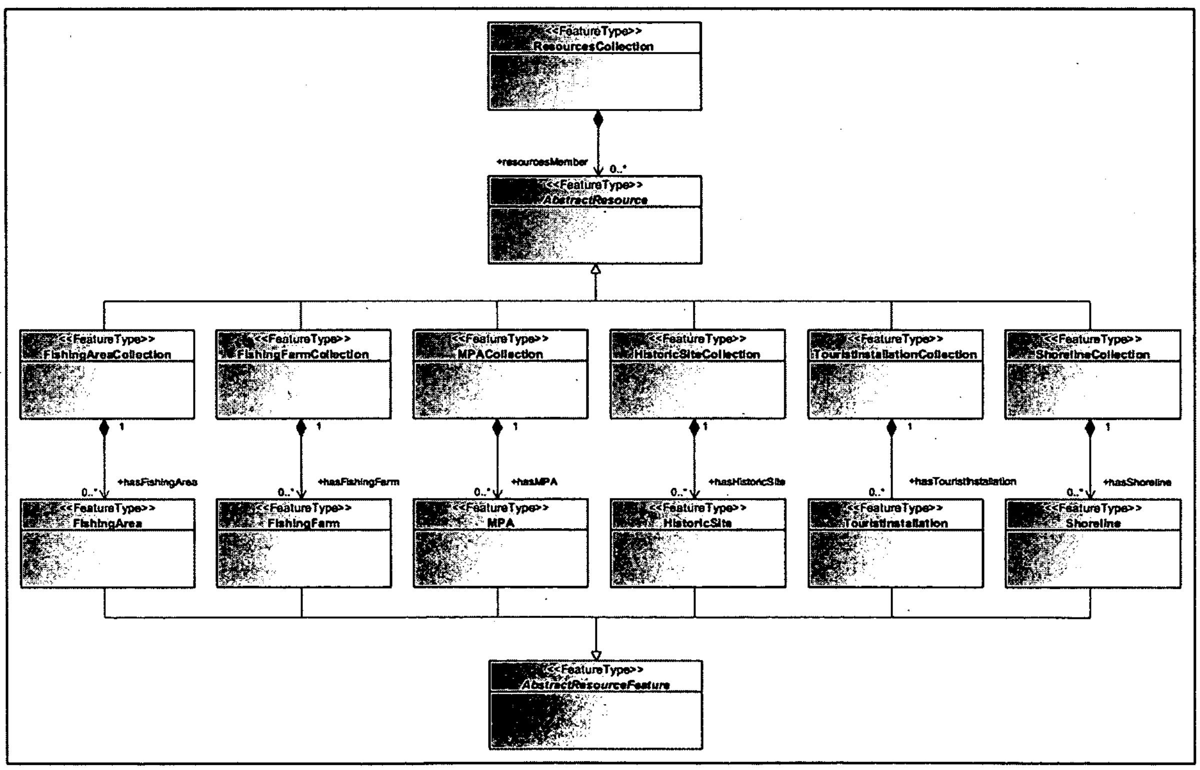


Figure 16 - Sub types and relative relationships of the `AbstractResource` class.

The `AbstractResourceFeature` class has some attributes inherited by all the PPMS Feature implementation classes (Figure 17).

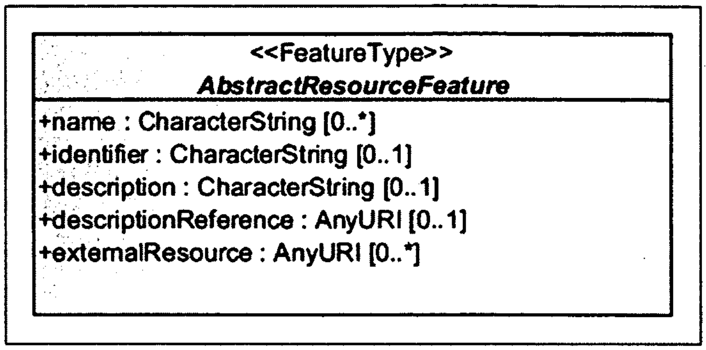


Figure 17 - Attributes of the `AbstractResourceFeature` class.

3.2.3 Complementary Information Required in the PPMS Application

In this third group, information complementary to the previous two entity clusters are stored. Some of them, such as hydrographic (bathymetry, seafloor reflectivity and water column imaging data), oceanographic (average water temperature for the corrosion rate, currents, biodiversity index), and meteorological data (e.g., prevailing winds), are provided through a possible external reference.

This approach was chosen to control the complexity of the PPMS GeoDB PS, and to permit its adoption without the application of deep changes in existing database mechanisms. The approach to import (and thus to convert) some selected information field from practically infinite database types was dismissed. For instance, such an approach creates possible data duplication and may produce potential inconsistencies during update in both directions (from the PPMS GeoDB to the external database, and vice-versa).

Providing attributes used as reference for any type of source was judged to be preferable. This approach works particularly well if the external source has ISO standard metadata since the information mapping and import can be performed smoothly, reducing potential semantic issues (details on this problem are reported in Appendix D).

The AbstractComplementaryInfo class is realized by the following six concrete Feature Collections:

- AdministrativeAreaCollection class;
- SunkVesselCollection class;
- SurveyCollection class;
- RasterDataCollection class;
- EmergencyFacilityCollection class;
- MetOcDataCollection class.

Each member of the above Feature Collections has as common super type the AbstractComplementaryInfoFeature class (Figure 18).

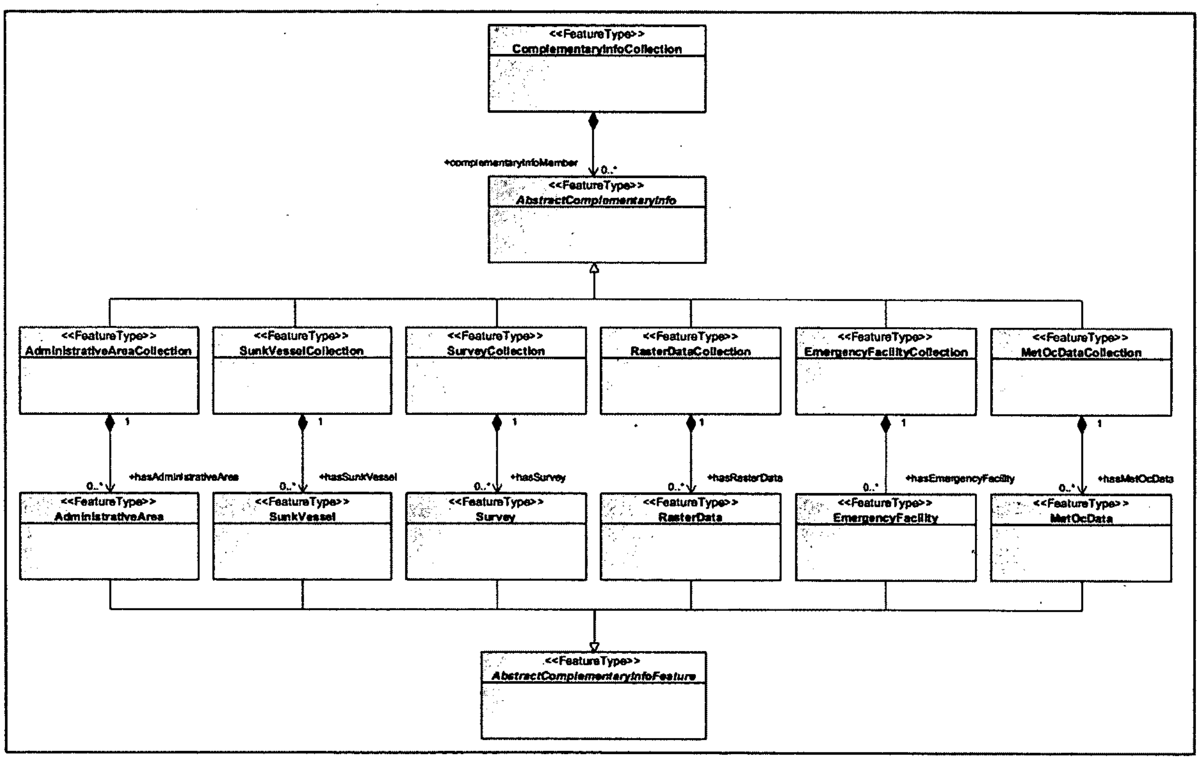


Figure 18 - Sub types and relative relationships of the AbstractComplementaryInfo class.

The AbstractComplementaryInfoFeature class has some attributes inherited by all the Complementary Information Feature implementation classes (Figure 19).

<<FeatureType>>
AbstractComplementaryInfoFeature

+name : CharacterString [0..*]
+identifier : CharacterString [0..1]
+description : CharacterString [0..1]
+descriptionReference : AnyURI [0..1]
+externalReference : AnyURI [0..*]

Figure 19 - Attributes of the `AbstractComplementaryInfoFeature` class.

3.2.4 Metadata and Metadata Collections

A key element of the PPMS GeoDB is the wide use of ISO 19100 Series Metadata (Appendix D), and the related S-100 profile currently in development (Figure 20).

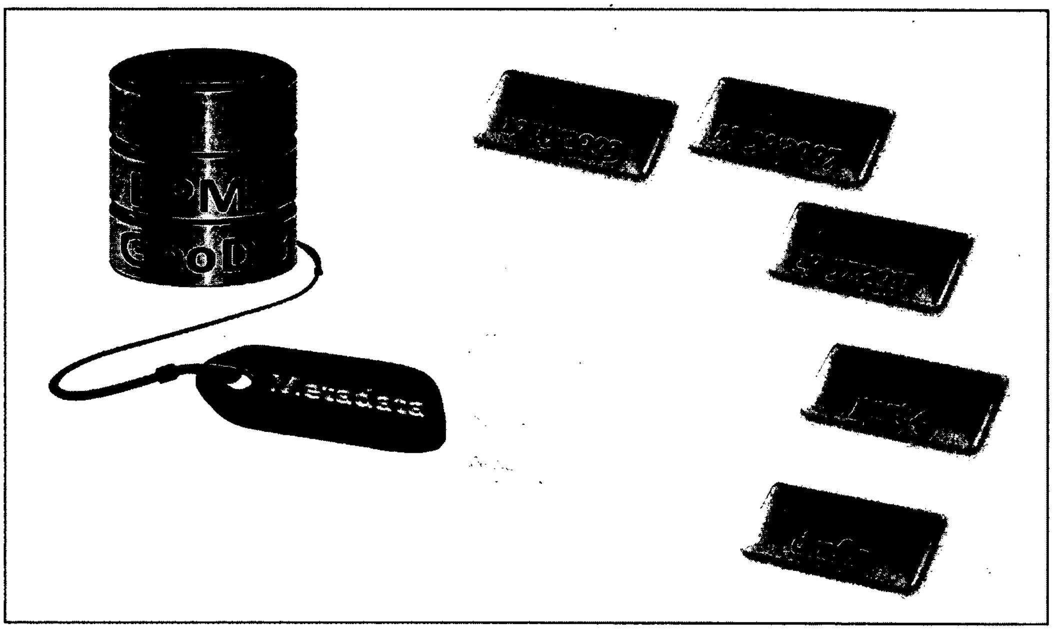


Figure 20 - Sources for the metadata implementation of the PPMS GeoDB. Metadata that supports multiple levels of search and description (e.g., from presence of data to a specific detail of geospatial projection information) must be allowed to make best (and correct) use of the available data.

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).

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 existing web metadata catalogues.
- Quality Metadata, extended to describe 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 21). 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.

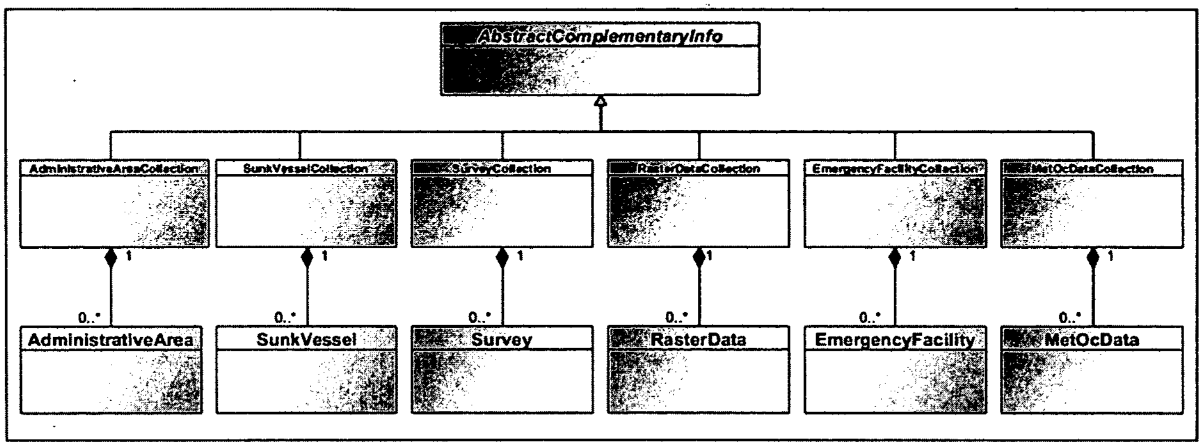


Figure 21 - Features collections of the `AbstractComplementaryInfo` class implemented as collections of metadata (in yellow). Use of metadata as part of the GeoDB (rather than as attached information) allows for simpler integration of data from other databases, and avoids redundancy of data between databases.

The idea is to map the field of these objects to the corresponding ISO metadata fields present in the external source (using a script or whatever application is well suited for this task for the given type of external source). Since a field mapping is created, it can be run at the periodicity decided by the user (e.g., before a query interacts with a given database).

3.3 The Physical Implementation

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.

Independent of the encoding standard adopted, the interchange of PPMS GeoDB data is based on the Product Specifications:

- The content, defined in the Feature Catalogue;
- The structure of a dataset, defined in the Application Schema; and

- The encoding rules that are applied.

Among the many possible encoding standards, the PPMS GeoDB provides an informative implementation using the Geography Mark-Up Language (GML).

3.3.1 Why GML?

The 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 (easy to understand by a computer, compact, and the ability to add metadata), as well as some advantages of text files (universally interchangeable) (Hunter et al., 2007). Further information about XML is provided in Appendix F.

From an IHO S-100 perspective, a number of GML-based Product Specifications are in development. These include:

- The DOALOS (Division for Ocean Affairs and the Law of the Sea) Maritime Limits and Boundaries Product Specification;
- The SNPWG (Standardization of Nautical Publications Working Group) MPA Product Specification;
- The IALA (International Association of Marine Aids to Navigation and Lighthouse Authorities) Aids-to-Navigation (AtoN) Metadata Product Specification.

A common characteristic of these Product Specifications is that they do not require the updating functionality and minimal data volume offered by the ISO 8211 encoding. In this view, the adoption of a GML Profile has been recently invoked (TSMAD, 2012).

Since it is accepted by most industrial companies and research institutions, GML essentially acts as a *de facto* standard in spatial data processing and exchange. In 2007, version 3.2.1 became an international standard as 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). Implementers may decide to store geographic information directly in GML, or to use GML only for schema and data transport.

When the ISO 19136 rules for designing GML application schemas using UML are enforced, it is possible to export the resulting UML model as XML and to automatically transform the resulting XML to the XML encoding of a GML application schema and vice versa (ISO, 2007).

GML encoding mechanisms permit the use of a physical Application Schema to validate data instances that claim conformance with the application schema in an automatic way. GML application schemas are written in XML Schema Definition Language. This is capable of expressing simple constraints (e.g., minimum and maximum values, character patterns). Complex constraints included in the logical model in a formal language (e.g., Object Constraint Language) can be implemented with different mechanisms (e.g., ISO/IEC 19757-3:2006 Schematron) (IHO, 2010; ISO, 2006).

A description of GML both as general overview and highlighting ISO version is provided in Appendix G.

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.

3.3.2 From the Proposed PPMS GeoDB PS to the GML Application Schemas

A number of steps were followed in the creation of GML Application Schemas for a Potential Polluting Marine Sites GeoDB:

- Provide the declaration of a target namespace that is a mechanism for avoiding ambiguity arising from name clashes (ISO, 2007);

- 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.

Due to the complexity and the number of objects involved in this application, several Application Schemas have been created:

- *root.xsd*, containing the root element and the main Feature Collections which can be present in an instance document;
- *ppms.xsd*, used to describe different kinds of Potential Polluting Marine Sites;
- *resources.xsd*, for different types of resources that can be threatened by a PPMS;
- *complementaryinfo.xsd*, for various types of information useful in the PPMS management;
- *ppmsgeodb.xsd*, a wrapper to include all the above Application Schemas.

A detailed description of all the followed steps is provided in Appendix H.

3.4 A Possible Implementation for Data Storage and Query Applications

Even if the PPMS GeoDB PS does not mandate any particular data storage, this section describes a possible implementation for storing and querying GML since it represents a key element in obtaining the full efficiency from this technology, as well as a tool to test and to evaluate the overall feasibility of the proposed Product Specifications.

Because 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 a GML document based on the PPMS GeoDB PS, a basic C++ application was developed.

3.4.1 Issues Related to Working with GML Data

A main issue in working with GML-encoded data is how to integrate XML technology with spatial technology (Manca et al., 2011; Steiniger and Hunter, 2012). The approaches developed for XML storage, transformation, query and index cannot be directly applied for the spatial information existing in GML. In fact, even if the spatial part is expressed with the same tag structure as XML, it needs to be processed by spatial topology technology (Wang et al., 2011). For instance, even if GML can be parsed by the same methods as XML, spatial data nodes have to be parsed as a whole rather than decomposed one by one into nodes or elements.

The main aspects of this issue are: storage, indexing, parsing, and query.

Storage

Since a GML file is a textual document, a simple file system approach is easy to apply. However, a database is a better alternative mechanism for data storage since it has higher performance for indexing and querying mass data (Wang et al., 2011).

Since a GML document has a structure different from a relational database, it has to be converted into a different model using approaches such as model mapping and schema mapping. Different database structures can be used to store GML data: a native XML database, a relational database (RDB), an object-relational database (ORDB), and spatial databases. In term of computation model, the main difference between an RDB and an ORDB is that this latter allows the user of host object language like *C++* or *Java* to access data without the need of translating back and forth through SQL. This characteristic represents an advantage for the adoption from organizations that have embraced object-oriented programming. At the same time, an ORDB management system *"attempts to add OO-ness to tables"* with evident advantages for extensibility and modeling of complex data relationships (McClure, 1997). An RDB could be seen as a subset of an ORDB that lacks the requirement of data and operations encapsulation

(Albarrak and Sibley, 2009). Key elements toward the RDBMS are represented by the simplicity of use and development, as well as the wide variety of end-users tools (Halpin et al., 2008). Finally, since relational database products have been under development and used much longer than object database products, they are more mature products with optimized performances and a large set of functionality (e.g., parallel processing, security) (Yang et al., 2009).

The term 'spatial database' may be applied at a large number of very different type of databases since it is generically used for databases that are in some way optimized to store and query spatial data (Shekhar and Chawla, 2003). This optimization implies additional query types such as, for instance, spatial measurements to calculate the distance between objects (Mamoulis, 2011). As a consequence, the adjective 'spatial' may be added to any database with such a peculiarity.

Indexing

Parsing GML documents stored as files tend to be slow with large datasets. Processing GML data stored in a DBMS has high efficiency utilizing the DMBS features, but the conversion may be expensive. As an alternative to DBMS conversion, GML indexing could be adopted to accelerate queries by directly locating the target data, avoiding traversal of all of the data. In the process of using a GML-oriented query language, some spatial operators can be accelerated with a GML index (Fubao et al., 2010).

In both cases, a structure indexing GML data should extend traditional indices with an additional spatial index. Thus, a GML index structure should be a combination of:

- Text and attribute indices (used to locate text values and assign an index to each attribute name and its value);
- Path indexing (similar to an XML index with the spatial geometry in GML documents regarded as a whole node rather than decomposed into internal coordinate nodes); and
- Spatial indexing.

The text, attribute and path indices are commonly implemented through a classical B+-tree, while the R*-tree or the R+-tree represent possible solutions for the spatial index (Beckmann et al., 1990; Sellis et al., 1987; Wang et al., 2011). Both are variants of an R-tree that is a B+-tree like structure storing multidimensional rectangles (Beckmann et al., 1990). They are completely dynamic: insertions and deletions can be intermixed with queries without any global reorganization (Beckmann et al., 1990).

Parsing

Since a GML file is a textual document, text parsing approaches are applicable. A GML file can be parsed by the same methods as XML, like DOM and SAX, with the only difference being that spatial data nodes in GML must be parsed as a whole rather than decomposed one by one into nodes or elements. An alternative approach is to use XML Data binding that directly delivers the data in an object-oriented representation generated by a compiler. This is the solution used for the proposed implementation, mainly due to the semantic control that is provided. More details about XML Data Binding technique are provided in Appendix I.

Query

In XML, many query languages have been proposed (e.g., Lore, XQL, XML-GL, Quit, etc.). XQuery, defined by the W3C, has become the standard in XML query. MonetDB/XQuery, Saxon and BaseX are well known XQuery implementations.

With spatial operator extensions to XQuery, some GML query languages have been proposed:

- GQuery added external spatial operators into XQuery;
- GQL presented a complete data model and definition of GML query; and
- XML/GML prefilter was proposed in building native GML processors.

GML Query Language (GQL) extends XQuery by adding a spatial data model and operators. Thus, it effectively supports querying non-spatial and spatial information in a GML document, making GQL widely accepted in writing standard GML query expression. The GQL query expression is similar to XQuery expressions and fully supports the XQuery syntax.

If the GML data are stored in a database (e.g., a relational database), the native query language (e.g., SQL) usually represents the best option. It is arguable in this case whether there is a need to remap the applied queries to one of the GML native query languages.

3.4.2 Data Management Framework

A pure XML database does not represent, at the moment, the best choice given the expense involved in its adoption (Ahmad, 2011). It has also been debated whether XML can be effectively used as a database language, since it best supports 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, even if not necessarily the best language.

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 performance;
- They are largely considered a safe choice by corporations; and
- A producer could hesitate to switch suddenly to a new technology.

These 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.

From a general point of view, a Database Management System (DBMS) is a set of programs and libraries used to create, maintain and efficiently use large sets of data. In large organizations, the control of the development of a database is usually in the hands of database administrators (DBAs) and other specialists due to the high complexity that this type of system can reach.

For the realization of the prototype of an application able to manage PPMS GeoDB GML files, five open source DBMS were evaluated as a possible solution:

- *MySQL*, a relational database management system (RDBMS) that runs as a server providing multi-user access to a number of databases (GPL license for non-profit and academic institutions). It is used in many high-profile, large-scale World Wide Web products, including Wikipedia, Google and Facebook;
- *Oracle*, an object-relational database management system (ORDBMS) produced and marketed by Oracle Corporation (free use for non-commercial purpose);
- *Apache Derby*, a Java relational database management system that can be embedded in Java programs and used for online transaction processing. It represents a good option for networking;
- *Postgre SQL*, an object-relational database management system (ORDBMS) released under an MIT-style license. This database works well for some well-known web frameworks (e.g., *django* and *turbogears*);
- *SQLite*, an embedded relational database management system contained in a relatively small C programming library but able to guarantee reliable database transactions that follow a set of properties called atomicity, consistency, isolation and durability (ACID-compliant). *SQLite* uses a dynamically and weakly typed SQL syntax that does not guarantee the domain integrity (e.g., it is possible to insert a string into an integer column), which gives rise to criticism. Unlike client–server database management systems, the *SQLite* engine has no standalone processes with which the application program communicates. Instead, the SQLite library is linked in and thus becomes an integral part of the application program.

Of the five options, the *SQLite* option was found to be the best fit to the requirements of simplicity and scalability. The *SQLite* approach permits a monolithic application that does not require any skills of database administration. However, in the case of implementation of the

PPMS GeoDB PS in a productive environment, one of the other DBMSs should be selected to address increasing data complexity and sharing as well as to best fit the adopted internal organization. Regardless, this transition is not expected to be particularly problematic, as reported in several studies (Bi, 2009; Yang and Hai-Yang, 2010).

More details about *SQLite* and the related SQL Queries are provided in Appendix L.

3.4.3 GML-enabled Geospatial DBMS

The DBMS does not represent the only actor in the storage and manipulation process. In fact, due to various approaches to storage, parse and query, the development of a GML-enabled geospatial data management system should integrate different technologies to answer different user requests.

This framework may be divided into such different logical layers as GML storage, data processing, and interfaces for users, or applications that need geospatial data manipulation service.

The GML storage layer should provide the possibility to store the data by alternative means (e.g., DB-based and file-based approaches). For the DB-based approach, model-mapping and schema-mapping in DB-based storage should provide to the users a flexible support for their applications. At the same time, different query mechanisms should be designed in the processing layer for different storage strategies (database and files system). The framework should provide two 'classic' forms of user interface: GUI and API.

A generic GML data management system is outlined in Figure 22, in which some generic data processing features (e.g., query processor, spatial indexing) and possible storage solution (e.g., object-oriented relational database management system) are represented.

An example of a framework of GML Data Management is described by Wang (2011). In this approach, GML data are converted from their native tree model to an object-relational model, GML query expressions written in GQL are translated to SQL and then executed. An index summarizes GML data and generate an index structure for accelerating query execution.

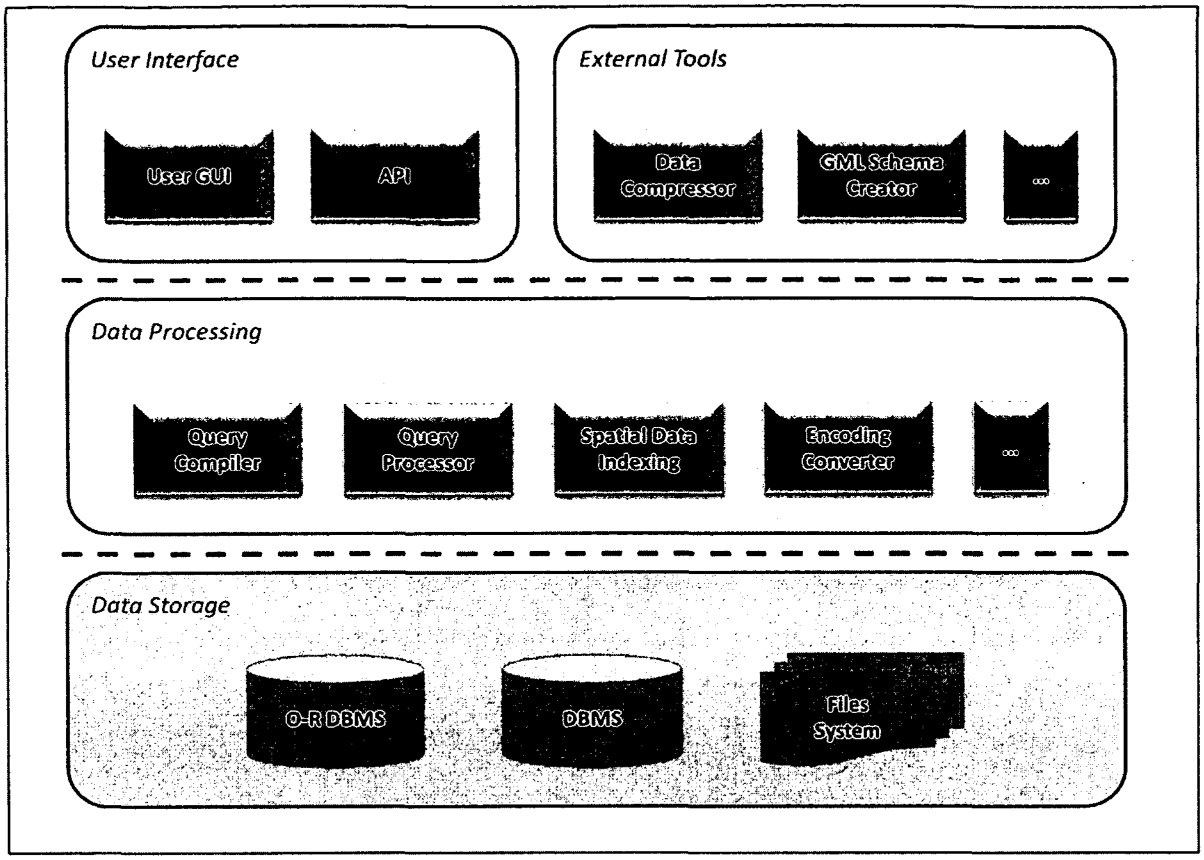


Figure 22 - GML data management framework. Adapted from (Wang et al., 2011).

This described framework was used as a reference to develop a similar structure for PPMS GeoDB data management.

3.4.4 The PPMS GeoDB Manager

The PPMS GeoDB Manager has several of these ingredients:

- A Graphical User Interface;
- A custom library that also represents a potential API for input/output PPMS GeoDB GML data from and to the relational database (encoding converter);
- An SQL query generator and validator that helps the user in query creation;
- Availability of a query processor and spatial data indexing through the SQLite library; and
- Creation of different outputs (geographic and not).

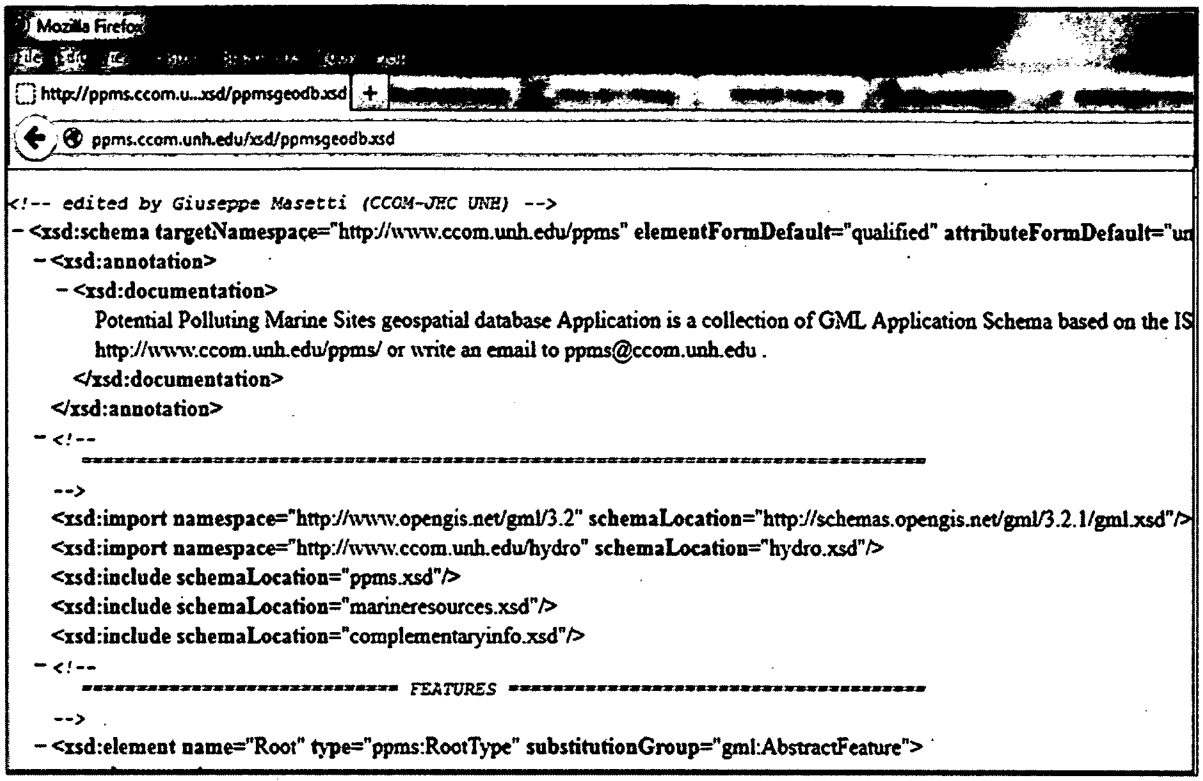


Figure 23 - PPMS GeoDB GML Schemas available on the project website.

The requirements for the target application have been:

- Reading and writing well-formed and valid GML files based on PPMS GeoDB Schemas hosted on the project website: `http://ppms.ccom.unh.edu` (Figure 23);
- Providing a minimalist user-friendly GUI with such basic operations as:
  - Display data as tables but also maintaining the GML tree structure;
  - Basic queries (e.g., selection by values, sort);
  - Basic statistics products (e.g., pie charts), but also performing advanced operation typical of SQL:
    - Spatial queries;
    - Creation of different typology of SQL-based relationships between specific columns of two or more tables (i.e., Join).
- Data importing from other (geographic) formats:
  - ESRI Shapefile;

- XYZ, csv, spreadsheets (.xls); and
- Raster (jpeg, tiff, geotiff, png, etc).

- Data exporting to other (geographic) formats:
  - ESRI Shapefile;
  - Google KML;
  - XYZ, csv, spreadsheets (.xls); and
  - Raster (jpeg, tiff, geotiff, png, etc).
- The internal storage solution should be:
  - Reliable and capable of storing relatively large amounts of data;
  - Compact and cross-platform portable;
  - Readable with generic COTS GIS software.
- Supporting multiple languages, through UNICODE (e.g., UTF-8);
- Supporting a spatial index to speed up geographic queries (e.g., R-Tree);

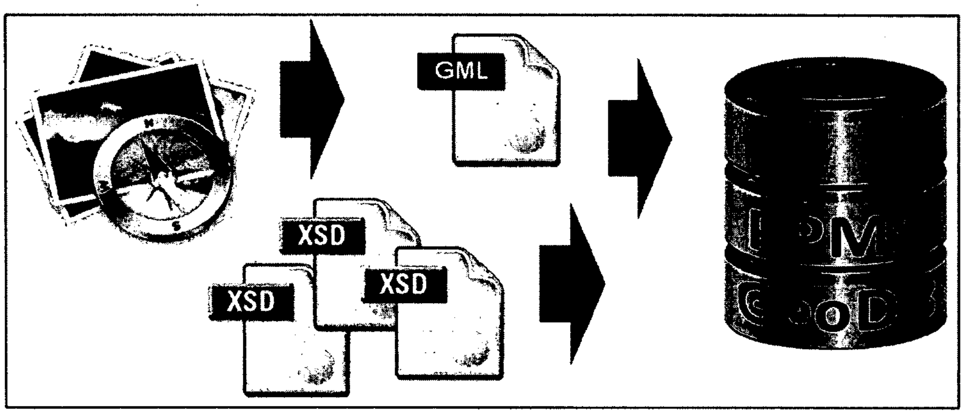


Figure 24 - A PPMS GeoDB may store not only GML data, but also images and schemas.

- Storage of as much data as possible from the PPMS GeoDB Exchange Set into a monolithic file (Figure 24):
  - GML information (.pdb);
  - PPMS GeoDB schemas (.xsd);
  - Images (e.g. vessel photos); and

- Raster data (jpeg, tiff, geotiff, png, etc).

- Enable the user to customize and calculate a Risk Index based on:
  - PPMS GeoDB information; and
  - External sources of (geographic) data (e.g. shp, dbf, csv, etc.).
- Use open-source libraries, including the type of license as an element of choice:
  - GNU Lesser General Public License (LGPL);
  - GNU General Public License (GPL);
  - Mozilla Public License (GPL);
  - Etc.
- Be cross-platform, at least:
  - Windows (x86 and x64);
  - Linux;
  - Mac OS X.

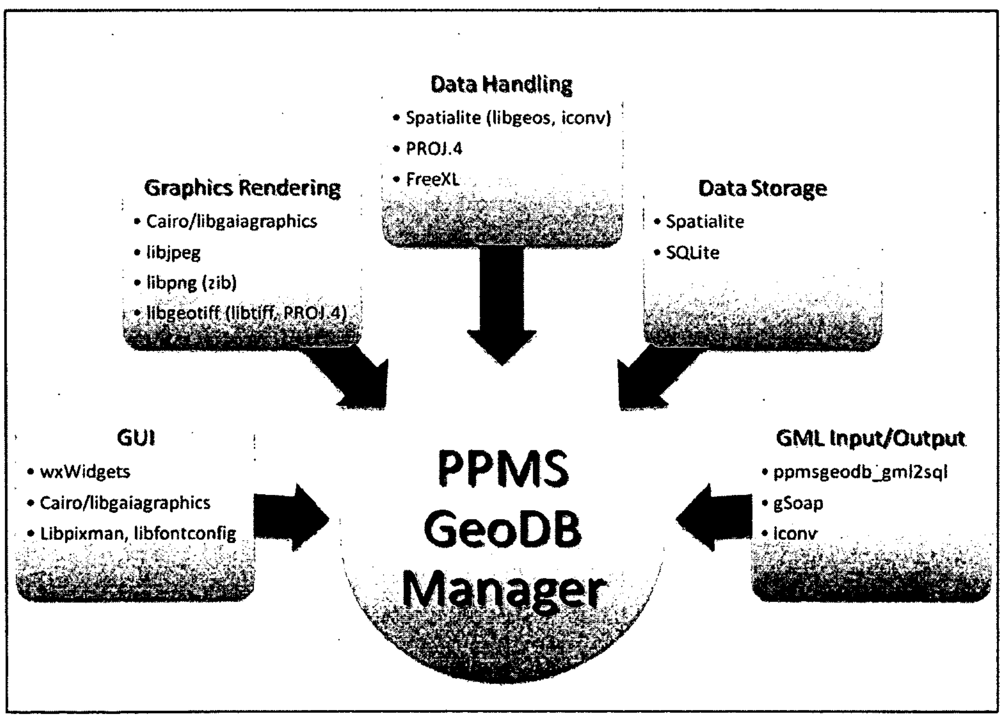


Figure 25 - Main components and related libraries used in PPMS GeoDB Manager.

Even if the PPMS GeoDB Manager is an Alpha version at the time of publication of this work, almost all these requirements are present in the built application. A large part of the facilities are provided using existing well-known open-source libraries as graphically represented in Figure 25.

In the next paragraphs the solutions adopted are briefly discussed.

GML Data / Relational Database Mapping

Mapping GML data into relational data represented a crucial step during the development phase. 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 an 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 available. A possible evaluation criterion for these approaches is the number of relation redundancies produced in the relational schema (since they could create anomaly problems). These redundancies are mainly related to the relationships present as reference in the GML tree data structure that are described using one or more XML Schemas.

- 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.
- In the structural-based approach, the constraint information represented by the XML Schema (or XML Document Type Definitions) 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.
- 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 (Atay et al., 2007; Liu et al., 2006; Lv and Yan, 2006).

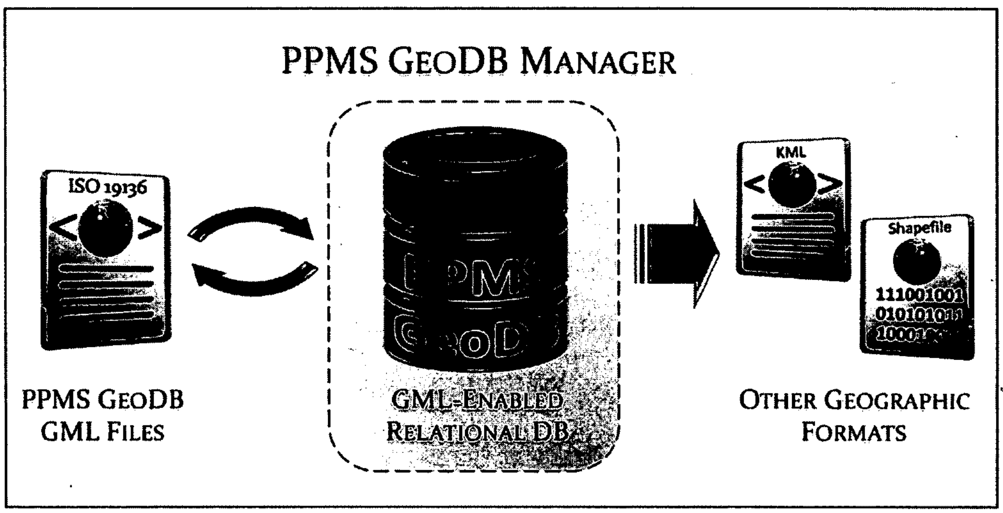


Figure 26 - Inputs and outputs of an application that manages PPMS GeoDBs. Data have to be stored in a relational database for efficiency and support reasons, but the mapping from hierarchically structured GML files to flat-table RDBMS is complex, and requires careful handling to ensure consistent and complete representation.

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 advantages of XML Data Bindings to store the PPMS GeoDB information into a dedicated relational database (Figure 26). The implementation of this approach is basically transparent to the user, since all the operation of validation, import, query and export are internally managed by the application interface. A dedicated C++ library was developed to perform the GML data input/output duty, largely based on a widely adopted open source XML Data Binding library (i.e., *gSOAP*) for GML data reading.

Since the GML data are not stored internally as XML (but within a SQLite-based database), this structure is commonly called an XML-enabled database.

Data Validation and Manipulation

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 at a raw representation of the XML structure (elements, attributes, and text). Thus, the developer has to write a substantial amount of bridging code to transform information encoded in XML to a representation more suitable for the application.

The PPMS GeoDB Manager adopts instead an approach called XML Data Binding. This approach, skipping the raw representation of XML, delivers the data in an object-oriented representation generated by a compiler from an XML schema (Kolpackov, 2007; Surhone et al., 2010). XML Data Binding is a more efficient way to handle the GML documents, given the complexity of the PPMS GeoDB Application Schemas.

Among the available libraries that implement XML Data Binding, *gSOAP* was selected mainly for its large adoption, which is assumed as synonym for reliability, and the license type. The *gSOAP* tools represent an automated Simple Object Access Protocol (SOAP) and XML data binding for C/C++ based on compiler technologies, auto-code generation and advanced mapping methods. Practically, they automatically map native and user-defined C/C++ data types to semantically equivalent XML data types and vice-versa. Furthermore, *gSOAP* supports the entire XML Schema 1.1 standard (except XPath expressions and assertions). Finally, the *gSOAP* adoption could make easier a future web-service development.

More details about *gSOAP* tools and the XML Data Binding technique are provided in Appendix I.

GUI and Graphics Rendering

The Graphical User Interface (GUI) for the PPMS GeoDB Manager used *wxWidgets* to support cross-platform development. The selection of *wxWidgets* is based on the consideration that it is, at the moment, one of the most complete and cross-platform GUI toolkits, as well as the fact that is totally free for personal and commercial use. Furthermore, the use of a platform-native software development kit (SDK) preserves the system's native look-and-feel. The *Cairo* and *Libgaiagraphics* libraries are used to support rendering of 2D graphics primitives on output. These are used here for:

- Image buffers;
- PDF format; and
- SVG format.

Data Handling and Storage

As already discussed, the main reason behind the choice of *SQLite* is its simplicity and its lightweight library that implements a full SQL engine. This combination make this library ideal for a prototype application such as the PPMS GeoDB Manager.

Furthermore, *SQLite* was adopted with its spatial extension called *Spatialite*. This open source library extends the *SQLite* core to support fully-fledged Spatial SQL capabilities.

From many point of views, the combination of *SQLite* with *Spatialite* extension represents a complete Spatial DBMS solution similar to *PostgreSQL/PostGIS* without server-client complications.

GML Data Input and Output

Ppmsgeodb_gml2sql is a library developed with the PPMS GeoDB Manager used to:

- Initialize the PPMS GeoDB data structure;

- Read/write data from/to GML files based on PPMS GeoDB Schemas (.xsd); and
- Store and retrieve PPMS GeoDB data from *Spatialite* DBs.

This library basically performs the conversions reported in Figure 27.

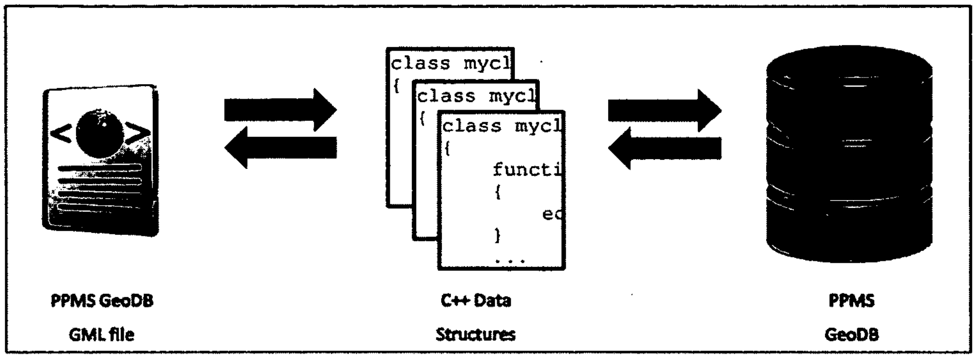


Figure 27 - GML Data to Relation Database conversion through C++ Data Structures.

Summarizing, the PPMS GeoDB Manger was outlined and developed as a prototype C++ application able to manage PPMS GeoDB GML files through a *SQLite/Spatialite* database. The main functions implemented inside the application are:

- (Spatial) SQL query compilation and execution; and
- Import and export in several commonly-used (geographic) formats.

This application relies largely on available open source libraries in order to increase its reliability as well as to reduce the time required for its development. Furthermore, this application was not intended to be used as a GIS since the *Spatialite*-built DB can be directly used from an existing open source GIS (i.e., *QuantumGIS*).

The next chapter will describe a risk index built on top of the PPMS GeoDB PS. Both the data structure and this index will be tested using some real data. At the same time, this operation represents a first benchmark for the PPMS GeoDB Manager. Finally, the processed data will be exported to be used as data layers for a publicly-available Web Map Server (WMS) implemented using the open source *MapServer* project.

CHAPTER 4

AN EXAMPLE APPLICATION: THE MARINE SITE RISK INDEX (MASIRI)

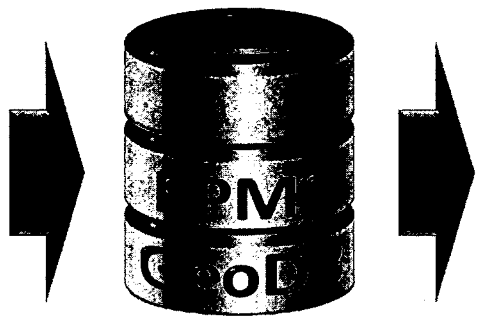
4.1 Introduction

Health, environmental and safety issues are an inherent part of industrial societies. Public opinions are always more sensitive to risks of major accidents happening near populated, environmentally and economically sensitive areas. The reactive approach usually followed in the past - that is, face the problem only when the pollutant starts to leak - is nowadays scarcely acceptable (Michel et al., 2005). In fact, the public demand for 'proactive' interventions is growing, so as to remove any significant threat of future pollution (Basta and Kennedy, 2004). The environmental community refer to PPSWs as *'oil time bombs'* (Girin, 2004).

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.

Although the main target of the PPMS GeoDB Application is a PPMS inventory (Figure 28), its implementation can be a tool for each phase of the disaster management cycle: emergency response, recovery, development, mitigation, and preparedness (Figure 29).

- ✓ PPSW
- ✓ Dumping Areas
- ✓ Archaeological Sites
- ✓ Fishing Areas
- ✓ Fishing Farms
- ✓ Tourist Installations
- ✓ Shoreline
- ✓ Administrative Areas
- ✓ MPA
- ✓ Meteorologic data
- ✓ Oceanographic data
- ✓ ...



- ✓ PPMS Inventory
- ✓ Risk Assessment
- ✓ PPMS Monitoring
- ✓ Risk minimization
- ✓ Disaster management
- ✓ Emergency response
- ✓ ...

Figure 28 – Feature inputs (on the left) and primary aim (PPMS Inventory) together with other possible applications (on the right) for a PPMS GeoDB product.

If the pollutant is easily removed or represents a hazard to navigation, or if there is a cooperative Responsible Party or other sources of funding, the intervention option is easily justified. Conversely, expensive mitigation operations to be conducted in deep waters without an easily available source of funds can become difficult to realize. More generally, some results obtained in the past (e.g., the oil recovery from the *Prestige*) have clearly demonstrated that there are very few limits related to the technology available for oil or wreck removal (Michel et al., 2005). Thus, the true limiting factor is funding.

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.

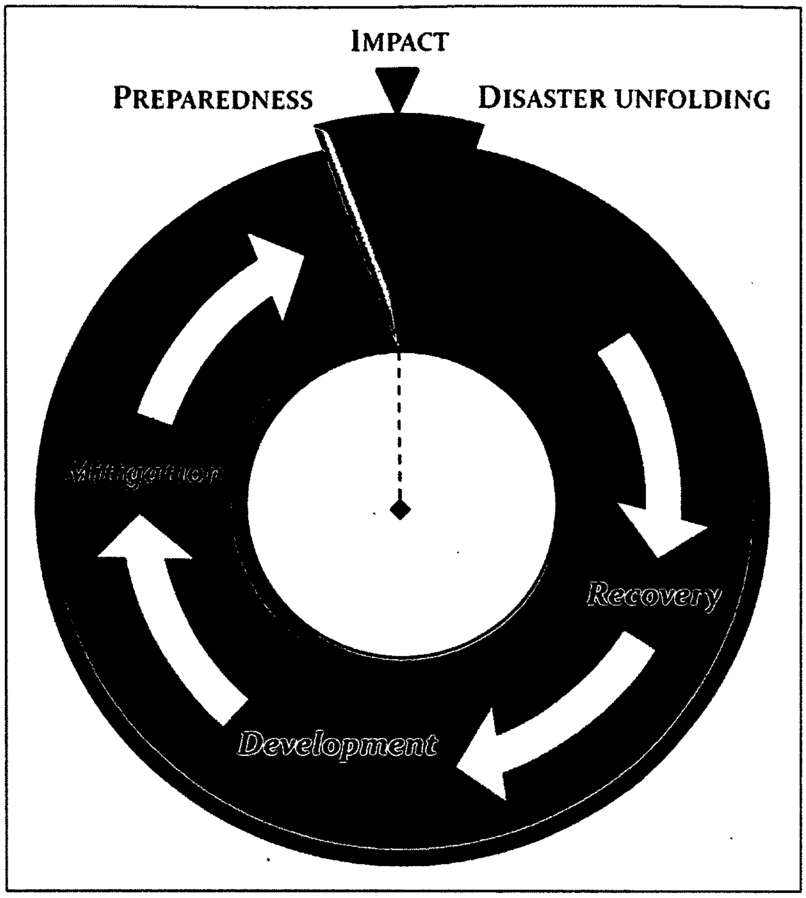


Figure 29 - The disaster management cycle. Used correctly, the PPMS GeoDB could provide key information at all stages of the cycle.

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.

In the remainder of this chapter, a possible risk assessment tool is presented, through the calculation of the Marine Site Risk Index (MaSiRI), preceded by some clarifications about the general frame adopted (e.g., the semantic of some terms) and an overview of previous work on this topic.

4.2 Adopted Definitions

Since some terms adopted in the risk assessment field can have various and sometimes subtly different meanings according to the thematic context in which they are applied, the focus of this section is to provide definitions about the terms used in the MaSiRI.

4.2.1 Disaster

The term 'disaster' usually indicates a serious and relatively sudden disruption of a community's life causing human, material, economic or environmental losses larger than the societies' resources to cope with them (ISDR, 2004). In the last few decades, this term has become related not only to nature-driven disasters, but also to events related to human actions. The monitoring of potential marine disasters through integrated tools like the MaSiRI is part of the aims of the PPMS GeoDB Application (Figure 28).

4.2.2 Risk

In MaSiRi, the definition of 'risk' is derived from (Schneiderbauer and Ehrlich, 2004) as the probability of harmful consequences resulting from a given hazard – e.g., an oil release from a PPMS – to a given element at danger or peril – i.e., any marine resource – over a specified time period. Conventionally, the risk is expressed as a function of hazard and vulnerability. Some disciplines also include the concept of exposure to indicate the physical aspects of vulnerability, building the so called risk triangle (Figure 30) (Crichton, 1999). Examples of exposure applied to the PPMS GeoDB Application are the evaluation of the number or Historic Sites or Tourist Installations surrounding a given PPMS (later proposed as possible refinements).

4.2.3 Hazard

A 'hazard' is a potentially damaging event which may cause loss of life or injury, property damage, social and economic disruption or environmental degradation. It is important to recognize that, if an event takes place in an uninhabited region with low potential threat to the environment, it does not represent a hazard or, better, it is a hazard of negligible magnitude.

Hazard characteristics (like magnitude, duration, spatial extent and frequency) are also related to cultural perceptions (Gravley, 2001).

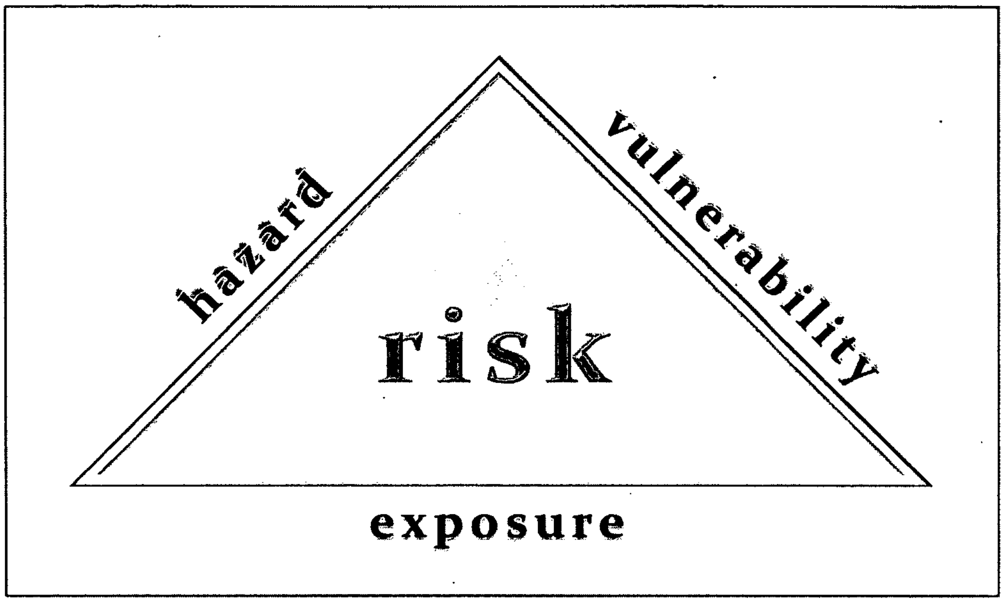


Figure 30 - The risk triangle and its components: hazard, vulnerability, and exposure.

Hazards are often grouped into three classes:

- Natural (e.g. earthquake);
- Social (terrorist attack); and
- Technological.

The hazards related to PPMS are mainly part of this last group. Since hazards may have interrelated causes, an oil spill from a shipwreck might be triggered by other hazards such as, for instance, a tsunami or a tornado. A hazard has only the potential to cause negative consequences on marine resources. These consequences become reality as a function of the vulnerability of each resource at risk.

4.2.4 Vulnerability

'Vulnerability' is determined by physical, social, economic, and environmental factors or processes, which increase the susceptibility to the hazard impact (ISDR, 2004; Schneiderbauer and Ehrlich, 2004).

It is related to economic development of a country (Alexander, 2000). In fact, the poorest societies have usually the lowest resources and capabilities to reduce vulnerability (Schneiderbauer and Ehrlich, 2004). In Figure 31, different types of societies are presented:

- Not developed and with few resources, at point A;
- With a slightly increasing economic development where risk grows faster than the capacity of mitigation, at point B;
- Point C represents the minimum vulnerability with societies that are highly industrialized and developed; and
- Point D represents highly developed societies where danger grows faster than risk mitigation (Alexander, 2000).

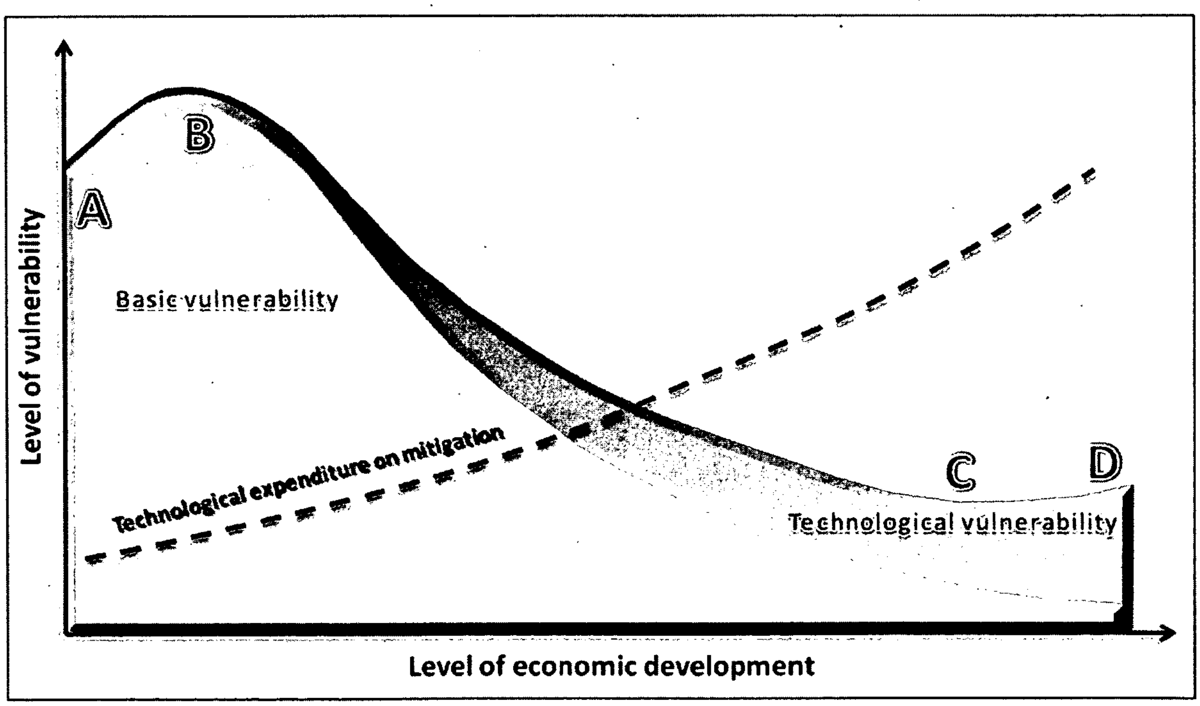


Figure 31 - Vulnerability related to economic development and, in red, technological expenditure on mitigation. Adapted from (Alexander, 2000).

Vulnerability usually changes slowly over the time required for the development process, but sometimes it increases suddenly (i.e., in the case of a sequence of disasters). An environment already affected by a hazardous event is usually more susceptible to future events. This consideration can also be applied with a more general point of view: each disaster (natural, social or technological) has the effect of suddenly increasing the overall vulnerability of a society or of an environment (Figure 32). Making an example related to PPMS, the vulnerability of a coast is different if it is affected by a single oil release or by a series of spilling events from the same shipwreck.

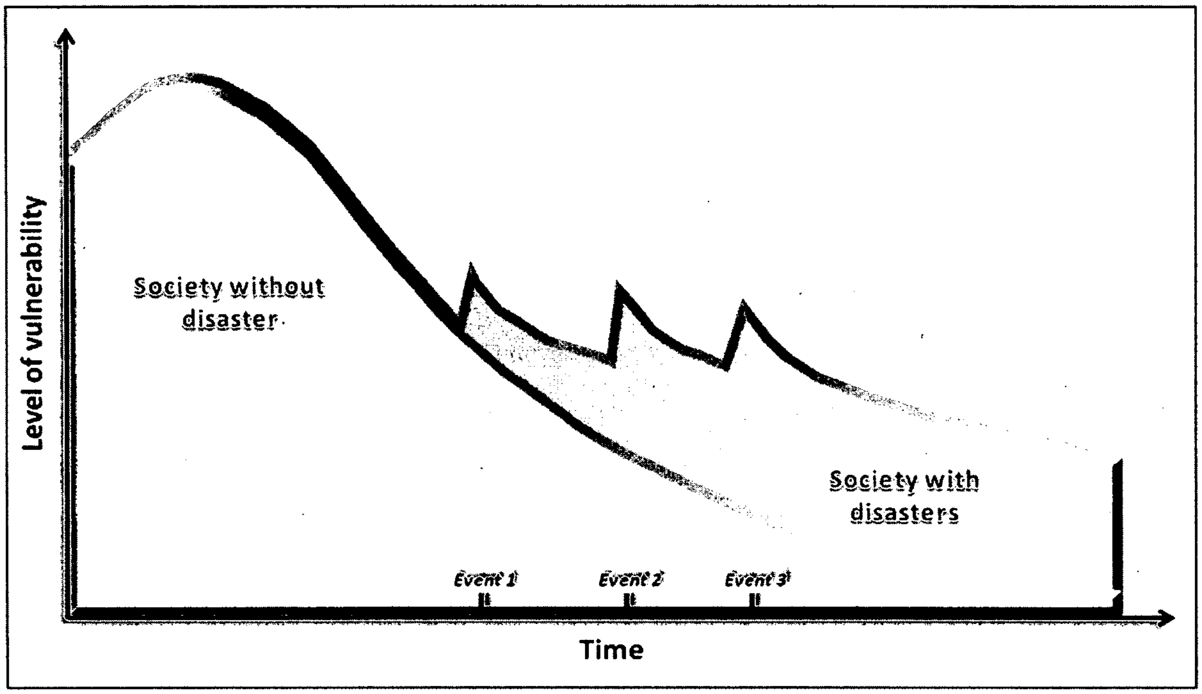


Figure 32 - Effects of a series of disasters on the level of vulnerability of a society.

For each event, it is possible to identify a pre-event and a post-event vulnerability or resilience (Figure 33). The first one, often called 'susceptibility', is mainly related to the physical assets and financial resources available. In the post-event phase, the ability to cope with and recover from the effects of a hazardous event – related to the institutional setting, organizational infrastructures and strength of social networks – plays a fundamental role.

Vulnerability changes with the severity (e.g., the amount of oil) and the type (heavy oil, chemical weapons, etc.) of hazard. It is difficult to determine 'proxy' indicators with the characteristics of being easily available, frequently updated and sufficiently uncorrelated. The weight to give to each indicator is also hard to define, and it may be based on expert knowledge.

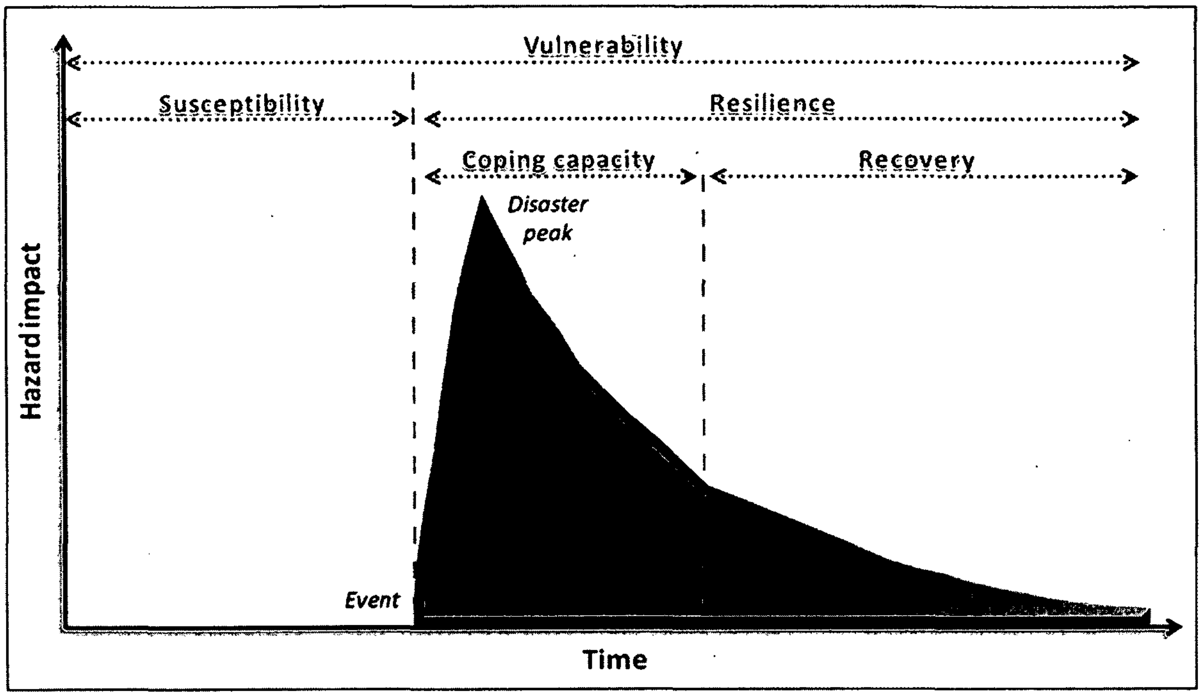


Figure 33 - Terms related to the vulnerability.

4.2.5 Exposure

'Exposure' represents the element at risk. It can be the number of people located in the area, but also infrastructures such as the surface of a MPA, the number of fishing farms, the value of biodiversity, etc. The combination of both hazard and exposure is sometimes called physical exposure.

4.2.6 Risk assessment

The risk R is a function of: hazard H, vulnerability V, and exposure E. This concept is usually expressed by some kind of 'risk equation'. The followed one is commonly used (Peduzzi et al., 2002; Schneiderbauer and Ehrlich, 2004):

$$R_{ah} = H_{ah} \times E_a \times V_{ah} \tag{1}$$

In the above formula, subscript *'h'* is a function of the type of hazard and subscript *'a'* is related to the geographical region affected by the hazard. The risk is null when just one of the three components is zero.

Thus, a risk is a likelihood of harmful consequences on a certain marine resource; a hazard represents the potential for a PPMS to cause this negative consequence; the exposure represents the total amount of resources at risk, and the vulnerability gives for each marine resource indications about the possibility to be significantly affected by the hazard. Paraphrasing an example presented in (Okrent, 1980), crossing the Atlantic presents the same hazard of drowning for people in a liner as in a rowing boat, but the risk is very different due to the vessel vulnerability.

Coming back to the PPSWs, over time the probability of leakage for the remaining oil trapped in a shipwreck equals one. But the main issue is whether the oil will periodically leak for years or if a sudden collapse will cause a large oil spill. Even if these opposite scenarios release the same amount of oil to the environment, the results on the environment and the economic impact will be very different. The natural oil degradation process can partially accommodate low release rates. This process is primarily driven by microorganisms, and some studies demonstrate that the recovery is faster than previously assumed (Elmgren et al., 1983; Hazen et al., 2010; Redmond and Valentine, 2011; Ritchie et al., 1994).

Another big issue is related to the definition of a global risk index. For example, it is difficult to define a risk index that permits correct comparison of a dumping area in Australia with a potentially polluting shipwreck in Africa. Not only the number of people affected can be very different, but also the hazard impact can differ significantly in scale (from local to global), in danger (frequency or magnitude), and in time (from several weeks to decades).

A possible different approach to estimate the risk is to look at impacts from previous hazardous events:

- The frequency is calculated by counting the number of events for the given area divided by the number of years of observation (average frequency per year);

- The vulnerability can be computed by dividing the impacts (that is, the realized risk) from previous hazardous events by the physical exposure (e.g., the shoreline length).

The normalization of the costs by the physical exposure may allow comparison between different coastal areas, as it suppresses the difference of exposed shoreline length. Although this approach is theoretical valid, the costs have a very large range and variability as pointed out in the next section. At the same time, the relatively small number of PPMS events makes this approach statistically weak and with the paradox to focus on a given area only when the disaster has already happened. However, in the future when we have large adoption of PPMS GeoDBs, this approach could provide interesting auxiliary information if applied to a historical and standardized collection of PPMS events.

4.2.7 Cost of intervention

The range of relative costs for oil recovery from a wreck is mainly a function of the site's characteristics. It can vary from a relatively simple case of a wreck site in shallow water containing low-viscosity oil, with an average response cost of $1-3 million, to more complicated scenario with depths of more than 300 meters, high viscosity oil and poor wreck conditions. In this latter situation the mitigation of the pollution threat could easily cost upwards of $100 million (Michel et al., 2005). Some examples of costs of oil removal from shipwreck sites around the world are listed in A further consideration related to this data is the high cost for the removal of relatively small amounts of oil. It is also important to differentiate between cases where large oil quantities are accessible from outside the hull and cases where small quantities of oil are drifting freely within the vessel (O'Brien, 2010). Furthermore, double hull tank characteristics of newer tankers, even if very effective in reducing the wreckage risk, represent an additional challenge (and cost) in case of recovery operations.

Table 5).

These values have to be very carefully evaluated, since they come from different and absolutely un-standardized cost calculations. However, they are probably sufficiently meaningful to provide an order of magnitude estimate of the cost per metric ton (MT) of oil removed for each event.

A further consideration related to this data is the high cost for the removal of relatively small amounts of oil. It is also important to differentiate between cases where large oil quantities are accessible from outside the hull and cases where small quantities of oil are drifting freely within the vessel (O'Brien, 2010). Furthermore, double hull tank characteristics of newer tankers, even if very effective in reducing the wreckage risk, represent an additional challenge (and cost) in case of recovery operations.

Table 5 – Several examples of costs for oil spill removal around the world.

Vessel Name	Year	Location	Depth	Oil Removed	Cost	Source
Cleveco	1995	US	21 m	1,100 MT	$ 3 million	(Davin and Witte, 1997)
Erika	1999	France	100-130 m	11,200 MT	$ 250 million+	(O'Brien, 2010)
Lukenbach	2002	US	56 m	325 MT	$ 20 million	(Hawkins, 2003; McGrath, 2011)
Osung No. 3	1997	S. Korea	70 m	20 MT	$ 5.8 million	(O'Brien, 2010)
Prestige	2002	Spain	3650 m	13,199 MT	$ 120 million+	(O'Brien, 2010)
Solar 1	2006	Philippines	639 m	9 MT	$ 6 million	(O'Brien, 2010)
YUIL no. 1	1995	S. Korea	70 m	634 MT	$ 5.6 million	(O'Brien, 2010)

4.2.8 Factors influencing the intervention decision

The factors evaluated to decide on an intervention to minimize the environmental risks and to permit a correct evaluation of the costs are usually the following (Melchers, 2005; Svensson, 2010; Zelo et al., 2005):

- The state of the wreck itself (type, dimensions and status of conservation);
- Pollutants carried (type, volume and distribution in the vessel structure); and
- Shipwreck site (distance from the coast, depth, sediment and weather, distance from oil spill intervention operators, distance from areas of economic or environmental interest).

The environmental risks of a shipwreck site are strongly related to (Lindström, 2006; Melchers, 1999, 2002; MLAUS, 2009; NRC, 2003; Reynolds, 2005):

- Wreck status of conservation (corrosion of internal and external structures);
- Type of pollutants carried (level of environmental hazard);
- Distance from the coast (impact on the coastal environment and economic activities);
- Distance from possible oil spill intervention operators;
- Distance from areas of economic or environmental interest; and
- Weather in the area (wind blowing towards coast or sea).

The choice of the type of intervention is usually based on (Michel et al., 2005; Monfils, 2005; Nawadra and Gilbert, 2002; SPREP, 2002; Yender et al., 2008):

- Wreck status of conservation and dimensions (i.e., recovery of the entire shipwreck);
- Type of pollutants and their relevant behavior (floating, solving, evaporating, sinking, etc.);
- Distribution and volume of pollutants (complexity of a localized intervention); and
- Weather characteristics of the site (sea state, prevailing winds, etc.).

The costs of the intervention are related to (Mervin, 2000; Monfils et al., 2006):

- Type and condition of the wreck (complexity of work);
- Dimension of the wreck, volume and distribution of pollutants (number of work days);
- Type of pollutant (cost for disposal);
- Depth of the site (complexity of work);
- Distance from the coast and from oil spill intervention operators (size of supply vessels, number of day for mobilization and logistics); and
- Weather in the area (number of stand-by days).

These costs have been classified for wrecks on the basis of technological difficulties in an IOSC publication and they are reported in Table 6 (Michel et al., 2005).

The decision to intervene on PPMS must be based upon a risk assessment and a cost/benefit analysis for different types of intervention.

This evaluation should allow decision on which intervention is less expensive, time-consuming, and risky. The most appropriate type of intervention should be compared with environmental impacts of a release from the wreck, the economic implications on the coastal areas and relative remediation costs. The intervention should be considered only if the costs posed by the oil or other pollutants release outweigh the cost of a proactive action, taking into account the economic damage and social consequences caused by repetitive or massive spills from the shipwreck.

Table 6 – IOSC cost of intervention classified by difficulty level. Adapted from (Michel et al., 2005).

Difficulty level	Definition	Cost (USD)
Simple	Wreck in shallow and protected waters Local mobilization	1 – 3 millions
Moderate	Wreck at moderate depth (20 – 50 meters) in an area with some weather limitations Regional mobilization	2 – 5 millions
Complex	Wreck at deep depth (50 – 250 meters) in an area with weather limitations, in open water Poor wreck condition Long mobilization	5 – 20 millions
Highly complex	Wreck at extreme depth (> 250 meters) in an area with limitations, in open water Poor wreck condition Long mobilization	20 – 100 millions

4.3 Review of Related Projects

Several existing projects have been considered during the process of developing a risk index built on top of the PPMS GeoDB Product Specification. This section provides an overview for some of them judged to be particularly interesting and whose concepts are directly part of (or indirectly influenced) the MaSiRI definition.

4.3.1 A Cantabrian coastal Risk Index

Castanedo (2009) created a GIS application to support spill response planning along the Cantabrian coast. They assumed the dependency of risk *R* for oil spill events related to two factors for each coastal environment:

$$R = H \times V \qquad (2)$$

The hazard *H* that an oil spill reaches a specific target is based on numerically generated data (due to the lack of long period measured data) using an analysis similar to the *NOAA Trajectory Analysis Planner* (Galt and Payton, 1999). Using a 2D Lagrangian transport model, meteorological and oceanic forcings were used to create a modeled track database of 519 potential oil spills and 878 target points, both in water and on the shoreline (Figure 34). The developed response system is based on the statistical analysis of this database.

The vulnerability factor *V* is a function that weights three components: physical I_P, biological I_B and socio-economic I_{SE} characteristics of the coast, using the NOAA Environmental Sensitivity Index (Petersen et al., 2002). The physical index I_P was calculated from the coast exposure to waves and the mean shoreline slope. The biological index I_B evaluates the ecological impact. It is based on:

- The conservation state I_C, which takes into account the current structural and functional status of the water body in which each segment is included;
- The singularity value I_S, which considers the conservation value of the segment according to its legal protection status; and
- The resilience factor I_r, from the power of the community to recover following the perturbation caused by an oil spill and the speed with which it is able to do so.

The socio-economic index I_{SE} estimates the economic damage due to losses related to interrupting activities and the cleaning cost.

The parameters of the vulnerability factor were estimated by means of a survey involving different stakeholders as:

$$V = I_P(0.71 I_B + 0.31 I_{SE}^{4.4})$$ (3)

To apply the model, the Cantabrian coast was divided in segments with a length of 200 meters and classified as estuaries, rocky shores, and sandy beaches (Figure 35). The evaluation of the risk so calculated provides a rank for the coastal protection.

This index is strictly oriented to the risk assessment for the coastline that is subdivided in segments. Thus, this approach does not evaluates the presence of other marine resources that can be affected by a PPMS event. The adoption of parameters coming from the stakeholders' survey underline the inherent difficulties of merging different proxies in a unique risk index value. Finally, this study does not take into account information (except the geographic position) about the PPMS and the amount of pollutant.

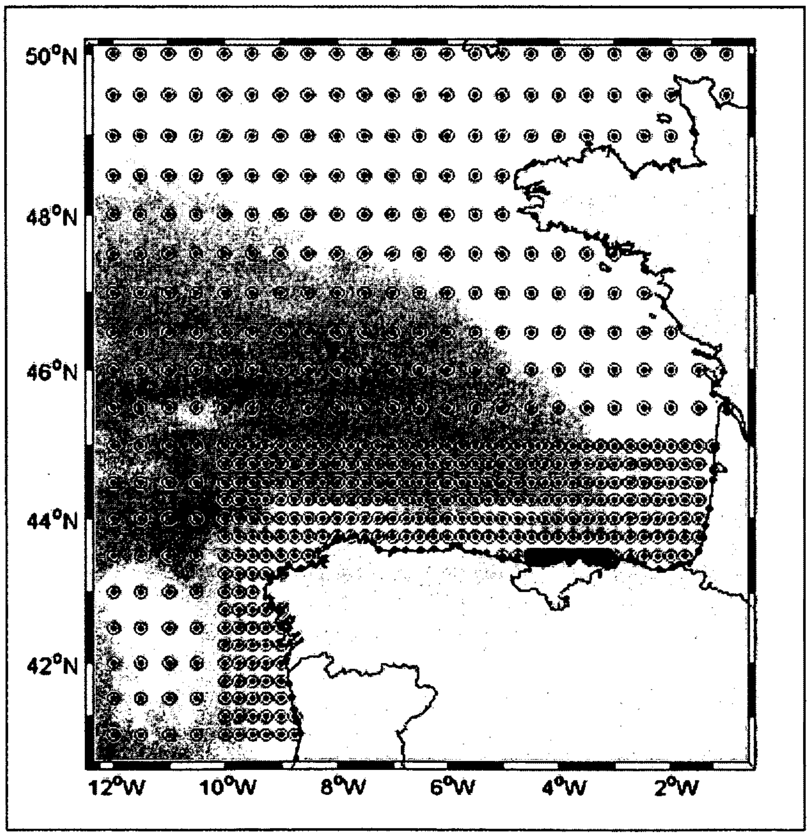


Figure 34 - Simulation domain and potential launch, in blue, and contact points, in red (Castanedo et al., 2009).

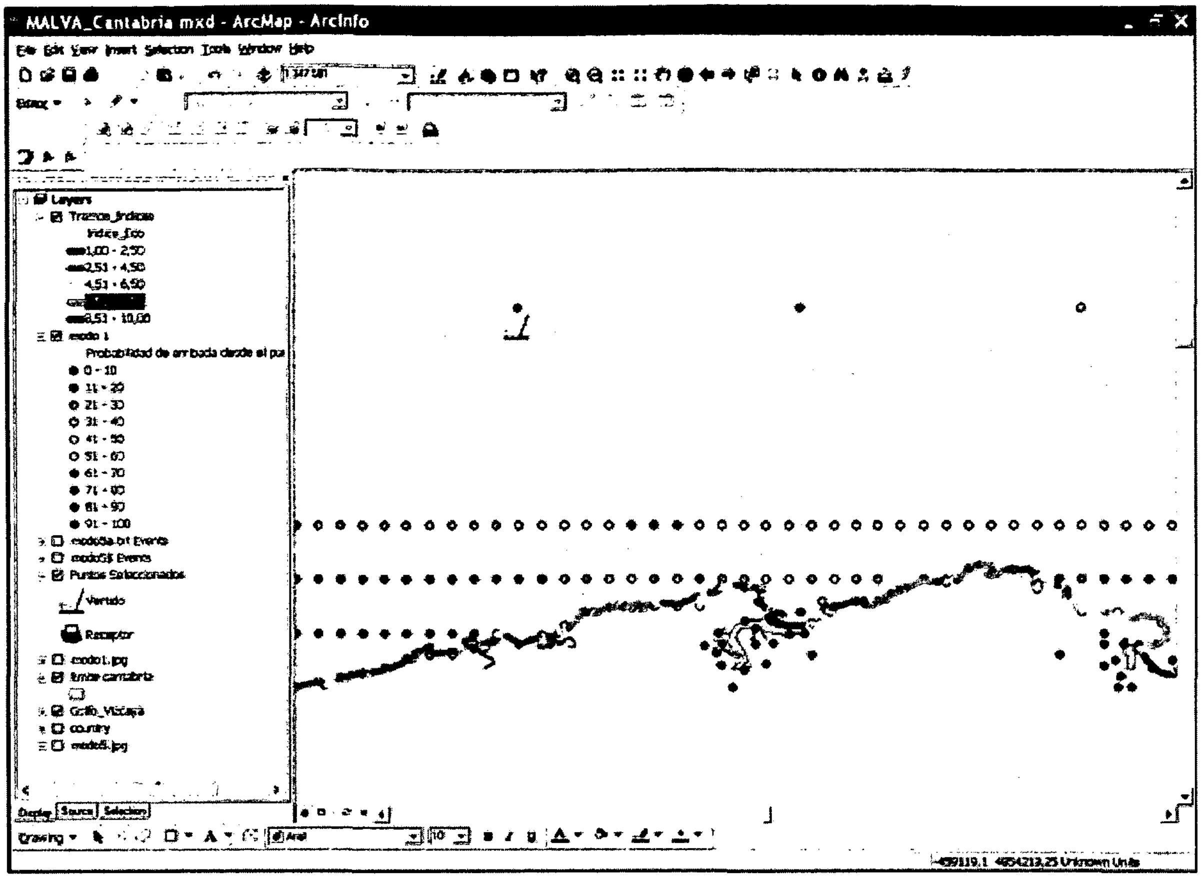


Figure 35 - The ship icon shows the initial oil spill location, colored circles indicate percentage of spills reaching points, and colored shoreline segments stand for vulnerability index (Castanedo et al., 2009)

4.3.2 A Brazilian Risk Index

This work uses GIS to assign oil sensitivity indices to the northern littoral region of São Paulo, Brazil (Pincinato et al., 2009). The natural clean-up capacity was defined by 3 factors:

- The quadrants from which waves with greater impact over the shoreline departed;
- The shoreline orientation; and
- Presence of structures such as shelters.

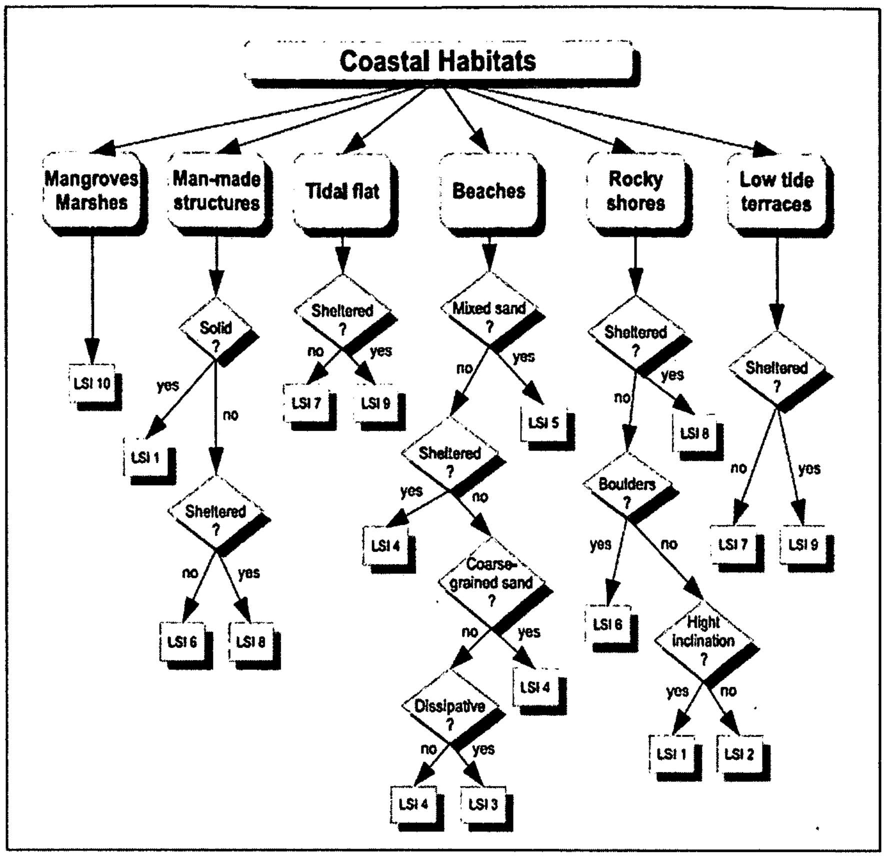


Figure 36 - Decision tree used to assign the Littoral Sensitivity Index (Pincinato et al., 2009)

The decision tree to assign the Littoral Sensitivity Index is based on (Figure 36):

- Environments (reflective beach versus rocky shore);
- Substrate type (fine-grained sand, boulders, etc.);
- Littoral inclination (> 30°, 5°-30°, <5°); and
- Exposure (sheltered or exposed).

Although this study does not provide a real risk index, it describes an interesting approach that permits classification through a decision tree of different types of coastline from the point of view of its sensitivity to the presence of pollutants. However, the collection of some of this information related to the coastline may require significant field work.

This requirement could represent a limitation on adoption at a large scale, although the use of this information when available is highly suggested.

4.3.3 The NOAA RUST

The NOAA RUST criteria to prioritize potentially dangerous wrecks are based on three ship characteristics:

- Construction year (after 1910);
- Steel hull; and
- Dimensions (greater than 200 feet or 1,000 GT).

This simple algorithm permits reduction of the number of dangerous records from more than 100,000 to tens of wrecks (Overfield and Symons, 2009; Symons, 2010). These criteria have also been integrated with additional components in a project called Remediation of Underwater Legacy Environmental Threats (RULET)(Helton and Symons, 2012).

4.3.4 The Barrett Risk Index

Barrett underlines a failure in NOAA's RUST criteria in that they ignore the surrounding environment, i.e., a shipwreck site in a low productivity environment has the same level of attention as one present in a marine protected or fishing area (Barrett, 2011).

The Barrett index risk – proposed to provide an environmental impact measure of American and Japanese WWII shipwrecks – is based on five factors (Barrett, 2011):

- Two related to the construction characteristics of the vessel (tonnage and type); and
- Three intrinsic of the shipwreck site (depth, distance from shore, and surrounding marine diversity).

Each factor can assume a value between 1 and 10 (increasing with danger) and their sum produces the final risk rating (maximum score is 50). The tonnage factor, common to the RUST criterion, is based on the assumption that large ships may carry larger fuel reserves and ordnance (Table 7). The type factor provides 10 additional points for oil tankers due to their significant oil capacity.

Table 7 – Barrett's tonnage factor classification.

Class (in MT)	Points
24 – 1,000	2
1,000 – 5,000	4
5,000 – 10,000	6
10,000 – 25,000	8
25,000 – 65,000	10

The depth factor – assigned on the basis of the depth corresponding to the vessel position reported on the ETOPO 1km bathymetric map – was determined through discussion with marine salvage professionals, in accordance with the industry's best practices and equipment (Table 8). Although the wrecks in shallow waters have a higher corrosion rate, underwater remediation operations are easier to conduct.

Table 8 – Barrett's depth factor.

Class (feet)	Points
0 – 200	1
200 – 1,000	5
> 1,000	10

The distance-from-shore factor is based on the traditional maritime boundaries (Table 9).

Table 9 – Barrett's distance-from-shore factor.

Class (NM)	Points
0 – 3	2
3 – 12	4
12 – 24	6
24 – 200	8
> 200	10

The marine diversity factor is created on the basis of a normalized biodiversity grid (Tittensor et al., 2010) as reported in Table 10.

Table 10 – Barrett's biodiversity factor definition.

Tittensor Index	Points
0.00 – 0.15	2
0.16 – 0.32	4
0.33 – 0.48	6
0.48 – 0.64	8
0.65 – 0.80	10

The Risk Index does not take in account oceanographic and climatic conditions (currents, air and water average temperature, pH, etc.). By the application of the Risk Index to a dataset containing 3,000 WWII Pacific shipwrecks, a near normal distribution was found, with an average score of 26 and 3.67% of the records (9 sites) scoring between 40 and 50 points. A total of 99% of the records in this dataset lay within Exclusive Economic Zones.

The major weakness in this work is the lack of a mechanism to link the vessel tonnage to the effective amount of pollutant.

4.3.5 The Kassomenos Risk Index

Kassomenos proposed a formula in order to combine various risks from an oil spill incident (Kassomenos, 2004). The values used come from some risk analysis tables produced on the basis of the spatial and temporal risk of an incident correlated to prevailing wind fields (in both northern and southern Crete) and the possibility of an early intervention by the local oil spill fighting authorities. The formula also takes in account the presence of various sensitive social and financial activities and environmentally protected areas. The probability of the threat occurring is defined as high, medium or low.

The impact in the financial activity of the population or in the ecosystems is defined from 0 to 3, as reported in Table 11.

Table 11 – Kassomenos' impact value definition.

Impact	Value
No impact	0
Noticeable impact	1
Damage	2
Major damage	3

The potential risk is measured by a weighted point rating system with three levels (Table 12).

Table 12 – Kassomenos' potential risk classification.

Probability	Points
Low risk	1
Medium risk	5
High risk	10

The probability points are multiplied by the impact rating for each activity to obtain a weighted risk rating.

Infrastructures classified as vulnerable to oil-spills are protected areas, locations of special natural beauty, the zone of fishing and fish farming, and tourist zones (with tourist installations and beaches).

The points assigned to each risk level are introduced in such a way as to separate the different risk levels for the various activities of the citizens and threats to the environment, quantify the risk and discriminate the results. The values are combined in Risk Units (*RU*) for each month and area:

$$RU_{ij} = a_i \times b_j \times (c_k + d_m + e_n) \quad (4)$$

where a_i is the month factor (1, 5 or 10) calculated from the number of oil spill incidents in the month i (1, ..., 12); b_j is an area factor (1, 2, 3) obtained looking at the spatial density of the polluting events; c_k is related to the wind flow regime and the percentage of day in which winds blow towards the coast (1, ..., 3); d_m is the score assigned as a function of the type of area affected (i.e., 3 for a fishing area, 2 for a tourist installation, 1 for protected zones); e_n is an indicator of early fighting (from 1 to 5, as a function of the distance from oil spill fighting authority).

The risk is considered high when the combined *RU* is more than 150 (value that represents the beginning of the upper half of the scale).

This work adopts a mechanism to estimate the hazard based on the distribution in time and space of oil spill incidents. As previously discussed, the number of PPMS events is not judged sufficient for correctly evaluating a risk index related to PPMSs. Furthermore, the known presence of a big wreck with tons of pollutant will be mostly ignored by this approach until a disastrous release happens.

4.3.6 The DEEPP Risk Assessment

The DEEPP Risk Assessment is based on four parameters {Alcaro, 2007 #51:

- Pollutant volume;
- Distance from the coast or sensitive areas;
- Nature of the product involved (fuel or cargo); and

- Exacerbating factor (age of the wreck).

The different pollutant volumes are divided in four classes and based on the vessel type (Table 13).

Table 13 – DEEPP classification of volume based on vessel type.

Class of Volume	Pollutant Volume (m^3)	Vessel type
1	V < 100	Vessels up to 500 GWT, lost containers, small coastal tankers
2	100 < V < 500	Vessels from 500 to 3,500 GWT, small chemical tankers, barges, coastal oil tankers, container ships
3	500 < V < 2500	Vessels from 3,500 to 25,000 GWT, coastal oil tankers, barges, chemical tankers, container ships
4	V > 2500	Vessels over 25,000 GWT, chemical tankers, oil tankers

The distance from the coast or sensitive areas is classified using the distance covered by a floating pollutant in presence of a wind speed blowing to the coast (Table 14).

Table 14 – DEEPP classification of distance based on distance from the coast.

Class of Distance	Distance from coast (NM)	Period to reach the coast
1	d < 1	A few hours
2	1 < d < 10	1 day
3	10 < d < 50	A few days
4	d > 50	One or more weeks

The nature of floating pollutant was divided in hydrocarbons and liquid chemicals. For the hydrocarbons, four classes were defined by persistency in the environment:

- Light fractions (up to C10) evaporate within a few hours;

- Diesel oils and equivalent products (C9 to C20) evaporate up to 30%-40 % within a few days while the heavier fractions will disperse naturally in the environment due to the low viscosity of these products categories;
- Intermediate fuel oils (IFO 180 and 380 now called ISO RME180 and 380) and heavy fuel oils (HFO 700 also called now ISO RMK 700) evaporate only at a few per cent, remain on the sea surface for a long time, and are persistent on the shore lines; and
- Crude oils were divided in light crude and heavy/medium crude: the first ones represent an intermediate class between Diesel oils and IFO/HFO, the second are inserted in the IFO/HFO class.

Liquid pollutants are categorized in four classes:

- X, Y, Z on the basis of MARPOL Convention (IMO, 1978); and
- Other substances (OS).

Among the many projects surveyed, the DEEPP approach is the one evaluated as more suited and applicable since it is relatively simple to adopt but, at the same time, takes into account:

- The relationship among pollutant volume and vessel tonnage as a proxy in case of lack of information;
- The spatial offset of the PPMS from the coastline, but also other sensitive areas; and
- The time factor related to corrosion effects.

However, the DEEPP approach has some significant limitations: for instance, there is no information (as well as in the previous methods) about the confidence level for the risk index value calculated (if it is just estimated on the vessel tonnage, it is clearly less reliable), and there is a weakness in the evaluation of the environment at risk surrounding a PPMS.

4.3.7 Integrated overall analysis

The analysis of the risk indices described in the previous sections highlighted very different approaches and, thus, has indirectly confirmed the subjective nature of any risk assessment method.

Almost all the proxies used in the described methods present weaknesses and/or high implementation costs (since they are linked to particular information only available after a certain amount of field-work). Furthermore, some of them are based on simulation (Castanedo et al., 2009) and statistical analysis on particular datasets (Kassomenos, 2004) not easily available or linked to a specific local implementation.

However, it is possible to cluster some of the different elements used on the basis of three classes:

1 Site-specific proxies (first of all, the amount and the type of pollutant);

2 Area-specific elements (e.g., meteo-oceanographic elements, presence of resources at risk, etc.);

3 Economic indicators that are directly linked to the capability of a region or a state to address the effects of an environmental disaster or, for instance, the dispersal at sea of a limited amount of hydrocarbons.

Any of the presented studies provides a good coverage of all of these three classes, although different approaches are presented. As already argued in the section about the DEEPP risk assessment (Alcaro et al., 2007), it appears to be the most reliable and widely applicable. However, it presents the weaknesses of missing any reference to the surrounding environment and the absence of a risk quality mechanism.

Based on these considerations, the initial idea to adopt and apply one of the existing risk index method was abandoned in favor of a definition of a general risk assessment methodology that can take advantage of what is judged positively in each of the previous projects, and with the intent to overcome (or, at least, to reduce) the weaknesses.

For instance, the effects of a pollutant release on the marine environment should not be ignored in the risk assessment process and, thus, the adoption of a proxy as done by some of the presented methods is required (Barrett, 2011; Castanedo et al., 2009). At the same time, it is also necessary to have an approach to 'weight' the quality of the selected proxies. For instance, in the case of two indices at different spatial resolutions (global and local scales) and both providing indications of marine biodiversity, the quality of the information coming from the one at local scale is probably more reliable than a global value representing the average over a wide area. Thus, a mechanism that follows the risk assessment with the intent to provide its quality must be constituted.

The definition of a methodology that addresses the underlined issues is the object of the next section. It will provide a particular focus on:

- Selection of a limited, but meaningful, number of core proxies applicable to almost all known PPMS records;
- Definition of an approach in order to classify possible additional elements (useful for a risk assessment) with the requirement to reduce cross-correlation between proxies; and
- Evàluation of quality in the proxies used in the risk assessment through a level of confidence that is totally missing in the studied projects.

4.4 MaSiRI's Structure

The Marine Site Risk Index (MaSiRI) is a tool for risk assessment based on the information collected using the PPMS GeoDB Product Specification. It may be considered independent from the GML implementation previously described. However, in the project described in this thesis, it is integrated in a 'pilot' PPMS GeoDB implemented using GML encoding.

Its main aim is to provide an index that evaluates possible environmental impacts of each PPMS on the surrounding area and shoreline, analyzing and 'weighting' some of the information present in the proposed Product Specifications. For instance, the proximity of historical and archeologically significant sites increases the risk related to PPMSs being close to the area. At the same time, the proposed structure takes advantage of the preceding review of previous studies with the intent to rectify the elements judged too limited, incomplete or potentially misleading.

The basic outline of MaSiRI can be described similarly to the DEEPP risk index, in that it provides a qualitative (i.e., unit-less) numerical value built from a series of look-up tables linked by a particular equation for combining the results. However, it differs from DEEPP in many aspects that come from other work (e.g., the evaluation of the environment surrounding a PPMS) and for the innovative idea to provide information about the uncertainty and range of calculated risk index. Higher values of the risk index indicate higher likelihood of a significant event.

In order to reduce the intrinsic subjectivity of any risk assessment process, the proposed risk index follows a fixed series of steps and defines look-up tables to identify resources at risk, additional threats, distance of available assets, coastal observations, etc. (Figure 37). The results of this automated process may be partially modified by a registered experienced user (with an Expert Correction Factor), e.g., on the basis of external information such as records of similar incidents or by analyzing previous events on the same PPMS stored in the GeoDB.

The process consists of three steps. The first step provides a filter to the incoming data, in order to remove any PPMS for which the risk index would be unsuitable or undefined. The second step estimates a rough but generally applicable risk evaluation based on data elements that all PPMS will have available as 'core' data. The last step allows the system to take advantage of existing additional information (not always available but meaningful) on PPMSs in order to refine, improve or extend the risk index when more data is available. Each of these steps will be further explained later.

At the end of these three steps, an issue - common to any risk index - is the need to have some indication of the reliability of the resulting value, in other words a meta-risk value to concisely describe the quality of the risk assessment process. The lack of such information can erode users' trust in the validity of the risk index, besides having potentially dangerous effects in the decision makers' choices. For instance, identification of the *'most potentially polluting shipwreck'* in the DEEPP database mobilized public opinion to push for an intervention, but when investigated, the Italian Coast Guard found it was empty of any carburant (Alcaro et al., 2007; Baccicalupi, 2008). This represents a clear example of the need to provide information about the risk assessment process (e.g., about the statistical approach adopted to estimate the amount of remaining carburant). The solution adopted in MaSiRI is to provide a Risk Level of Confidence (*RLC*) that has a range from 1 to 100 and tends to increase with the more 'confidence' in the resulting index along the risk assessment process (outlined in Figure 37).

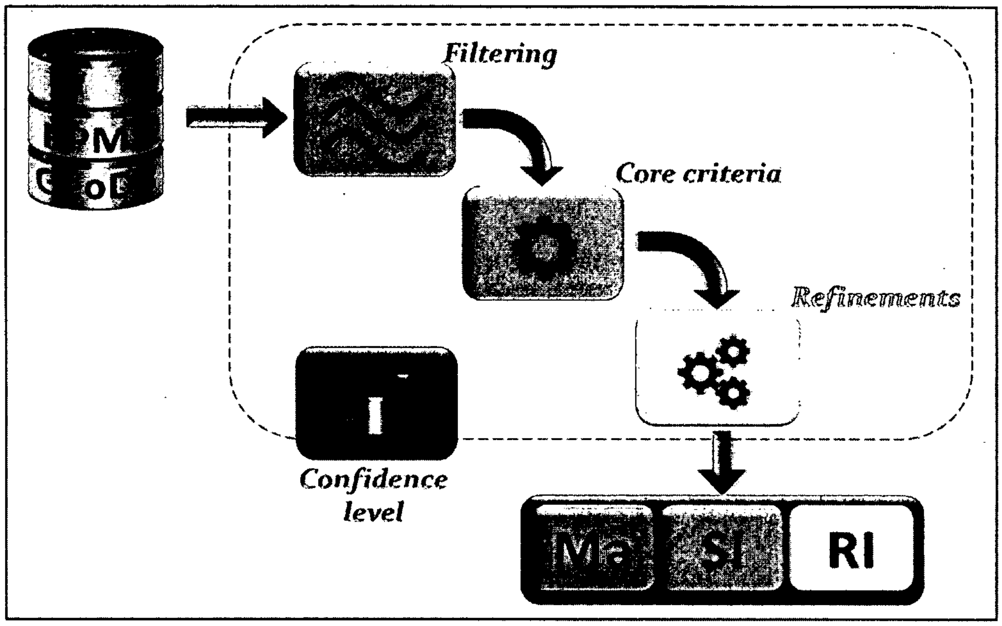


Figure 37 - Main steps in the MaSiRI calculation based on the information collected in a PPMS GeoDB.

4.4.1 Filtering

The filtering step avoids the application of MaSiRI to a given PPMS in cases where insufficient information is present in the GeoDB (in which case the index is set to `unknown`) or they do not match threat criteria related to size, time and material, e.g., nuclear waste (the MaSiRI is set to `not applicable`). The decision to leave out radiation hazards at the present stage of MaSiRI is mainly due to the complexity in implementation required to correctly evaluate the risk related to the different types of radioactivity (Hamblin, 2008; Morton et al., 2009; Poinssot and Gin, 2012). The filtering approach used here is based on (Lindström, 2006; Overfield and Symons, 2009).

4.4.2 Core Criteria

The basic idea of the core criteria step is to provide a rough indication of the risk index based on three elements common to each typology of PPMS:

- Volume of pollutant;
- Distance from the shoreline; and
- Type of pollutant.

The different pollutant volumes are divided into ten volume classes (*VC*) (Table 15) and the weight distribution is the same as the one followed in the DEEPP Project (Table 13). Since the potentially polluting volume is often unknown, the estimation approach suggest by the DEEPP Project (Alcaro et al., 2007) is adopted (when required), although a low level of confidence will be assigned in these particular cases.

Table 15 – MaSiRI classification of pollutant volume.

Class of Volume	Pollutant Volume (m^3)
1	V < 50
2	50 < V < 100
3	100 < V < 200
4	200 < V < 400
5	400 < V < 600
6	600 < V < 900
7	900 < V < 1200
8	1200 < V < 1800
9	1800 < V < 2500
10	V > 2500

The criterion to classify the distance classes (*DC*) is the distance covered by a floating pollutant in the presence of an average wind speed of 5-10 knots blowing towards the coast

(Table 16), based on the global average for wind speed at 10 m of height over the ocean (Archer and Jacobson, 2005; Lehr et al., 2002).

Table 16 – MaSiRI classification of distance based on distance from the coast.

Class of Distance	Distance from coast (NM)	Period to reach the coast
1	d < 2	Less than 1 hour
2	2 < d < 5	1 hour
3	5 < d < 8	A few hours
4	8 < d < 12	Half a day
5	12 < d < 16	1 day
6	16 < d < 20	A few days
7	20 < d < 28	Half a week
8	28 < d < 54	1 week
9	54 < d < 96	A few weeks
10	d > 96	1 month

Since the selection of sensitive areas may be differently driven on a country or regional basis, the distance from these areas (or the minimum between this value and the distance from the coast) is not part of the core criteria, but a candidate for a risk refinement at local/regional scale.

The hazard represented by the polluting volume is scaled as a function of the distance from the coast. The ratio between the polluting volume and the distance class represents the Site Scale Factor (*SSF*) and has a range between 0.1 and 10:

$$SSF = \frac{VC}{DC} \quad (5)$$

The nature of floating pollutant was divided in hydrocarbons and liquid chemicals. The pollutant classes (*PC*) are defined by persistency in the environment and the assigned values are based on the non-linear effects produced by the latter two classes compared to the first two (Table 17).

Table 17 – MaSiRI classification of pollutants defined by persistency.

Class of Pollutant	Type of Hydrocarbon Pollutant (Liquid Chemicals)	Persistency
2	Gasoline (MarPol OS)	Evaporate within a few hours (up to C9)
5	Diesel Oils and equivalents (MarPol Z)	Evaporates up to 30%-40% within a few days (C10 -C20)
8	Light Crude Oils (MarPol Y)	Represents an intermediate class
10	Heavy/Medium Crude Oils (MarPol X)	Evaporates only at a few per cent (e.g. IFO/HFO)

The resulting *SSF* (from 0.1 to 10) is combined with four pollutant categories (related to hydrocarbons and chemical materials) providing a first evaluation of the risk index:

$$MaSiRI_{core} = \frac{VC}{DC} PC \tag{6}$$

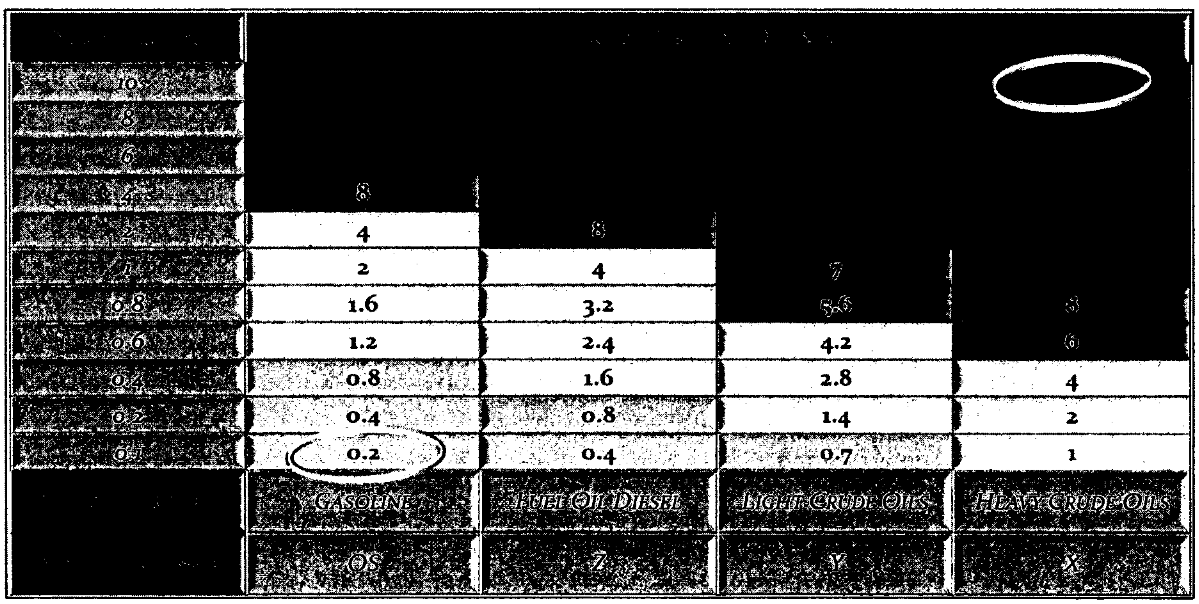


Figure 38 - Risk Index derived from the Scale Factor and the type of pollutants (the same classification followed in the DEEPP Project is adopted).

As represented in Figure 38, the resulting value spans the range 0.2 to 100. Making some extreme examples, a large tanker carrying crude oil and sunk very close to the coast and in shallow waters will have a MaSiRI value close to 100 (upper right corner), while a relatively small vessel sunk in the open ocean while it was carrying gasoline will have a very low risk index (lower left corner).

While PPMSs that lack information on position (and thus distance from the coast) and polluting content are directly filtered out at the filtering stage (MaSiRI is set to `unknown`), the polluting volume may still be roughly evaluated using the size and the type of site. However, the process followed to evaluate this value represents important information for the use of MaSiRI (since it can easily provide 'false positive' PPMSs) and, thus, it is represented through a Risk Level of Confidence (*RLC*) represented by a number not bigger than 100 that increases with the reliability on the calculated risk index. When there is no better information and the polluting volume is based on a method such as the one described in "The 2004 IOSC study" section, based on (Michel et al., 2005) and (Etkin, 2002), the *RLC* is set to a low value (e.g., 20 compared to a maximum of 80 for a direct and recent measurement). However, the adoption of refinements may not only change the MaSiRI but also increase the *RLC* of the obtained value.

4.4.3 Refinements

The refinements are mainly PPMS-type specific. They have the double role to:

- Evaluate a series of meaningful elements related to the PPMS (changing MaSiRI by a variable amount: e.g., in the range ±10); and
- Increase the level of confidence (*RLC*) (e.g., +5 or +10 points) in the resulting MaSiRI.

A list of some possible refinements are:

- Corrosion effects: the corrosion rate is largely region- (and sometimes also site-) specific (as documented in Appendix A);
- Marine biodiversity: some areas contains a smaller number of marine species than others, on these latter the effects of a PPMS event is relatively more dangerous (this factor was used in the Barrett Risk Index and in the Cantabrian Coastal Risk Index);

- Distance from emergency facilities: for instance, the presence of oil recovery facilities (e.g., a skimmer) may reduce the impact of a PPMS event (more details are reported in Appendix C), This factor was adopted by the Kassomenos Risk Index;
- Presence of marine resources in the area surrounding a given PPMS (e.g., the number or Historic Sites or Tourist Installations);
- Site depth: this factor has a double possible interpretation. On one hand, a shipwreck lying on a deep seafloor usually represents a challenge and an adjunctive cost in the case of an oil leak; while, on the other hand, the high pressure at depth and the lower exposure to atmospheric events can preserve the pollutant content from starting to leak or reaching the surface;
- Shoreline characteristics: as discussed in some previously presented works (e.g.,Pincinato et al., 2009), some characteristics of the shoreline are an important discriminant in the evaluation of its vulnerability (e.g., substrate type) as well as its exposure (e.g., shoreline orientation);
- Socio-economic indexes: the need for proxies in order to compare the technological development of a country and, thus, its capacities to handle a PPMS event was discussed in the previous 'Adopted Definitions' section;
- Meteo-climatologic elements: some examples could be chemical and physical characteristics of the water column, and climatic conditions (prevailing winds);
- Oceanographic elements: e.g., the direction and the intensity of currents impacting a PPMS, the bathymetry of the site and the geology of the area.

In any case the sum of the refinements may exceed the maximum value of '100'. To avoid this, but mainly to avoid deformations in the MaSiRI calculation, the refinements may be of any number but their final contribution (to MaSiRI itself and to the related *RLC*) cannot be more than 20 points for each of the following general classes:

- Site-specific class of refinements (R_S);

- Socio-economic and logistic (e.g., emergencies facilities) class of refinements (R_E);
- Coastal (e.g., shoreline characteristics) class of refinements (R_C).

Although the above value (20) is essentially arbitrary, it was based on empirical evaluations and its practical application appears plausible. Its main aim is to avoid data which is potentially cross-correlated from reaching too high a percentage level in the resulting MaSiRI output. The Expert Correction Factor (ECF) represents an additional optional way that, in a controlled environment (that is, a recorded and commented intervention and limited to a range of a maximum of 20 points) an experienced operator may decide to tune the MaSiRI output.

The final formula for the MaSiRI is:

$$MaSiRI = L_{100}\left(\frac{VC}{DC}PC + L_{20}\left(\sum_{i=0}^{n_S} R_{S,i}\right) + L_{20}\left(\sum_{j=0}^{n_E} R_{E,j}\right) + L_{20}\left(\sum_{k=0}^{n_C} R_{C,k}\right) + L_{20}(ECF)\right) \quad (7)$$

where the operator $L_\alpha(x) = \min(\alpha, x)$.

While a requirement for the PPMS GeoDB was global adoptability and validity around the world, two very different approaches could lead the adoption of MaSiRI:

- National/regional scale: each community adopting this risk index may customize the refinements in such a way that they best fit their vulnerability profile (e.g., some areas with high incomes coming from tourism may focus on the marine biodiversity impact of an oil spill);
- Global scale: an international committee develops a common set of refinements based on commonly accepted proxies to coherently represent the relationship between vulnerability and technological/economic development (highlighted in the previous section 'Vulnerability' and in Figure 31).

Since both approaches maintain the same set of core criteria (distance from the shoreline, volume and type of pollutant), the results of the MaSiRI application are comparable in any case at this level.

The selection and the full development of one of the two approaches is out of the scope of this work. However, some refinements adoptable at regional and at global scales are presented and adopted for the MaSiRI application to the test data set.

4.4.4 Relations with PPMS events

A new `PPMSEvent` object represents the availability of new data related to the PPMS. This data can be of different types as described in the annexed Product Specifications. They may represent, for instance, a new oil spill or the execution of a hydrographic survey related to the Potentially Polluting Marine Site. Each new `PPMSEvent` may eventually require a review of the associated MaSiRI, and the result of this iterative process is the redefinition for each PPMS of minimum and maximum values used to catalogue the risks of PPMSs present in the whole database. The range between these two values generally will tend to reduce with the amount of information. This approach represents an innovation with respect to the classic static or slowly dynamic calculation of a risk index. At the same time, the RLC related to a given source of information (e.g., a direct measurement of the pollutant volume) will degrade after some years. Furthermore, the comparison of successive `PPMSEvent` objects provides an interesting trend which is useful to understanding the evolution of a given PPMS together with the amount of related information available.

This approach requires also a visualization method that properly provides to the users this information. A proposed solution for this issue is part of the last section of this chapter.

4.5 The Application of PPMS GeoDB PS and MaSiRI to a Real Database

This section presents the result of the application of many concepts and features developed in the abstract (and previously described in this thesis) to a database containing real data.

The primary goal is the evaluation of the PPMS GeoDB data structure as described in Chapter 2 and in the annexed Product Specifications, testing the proposed conceptual model and validating its physical implementation. This final step also represents a first benchmark for the developed risk index and for the related idea that the concepts developed within the PPMS GeoDB project are capable of supporting more than just a PPMS inventory. However, the developed Risk Index presents several potential weaknesses that are mainly due to the decision to keep it as straightforward, intuitive and easy to adopt as possible. Although simple, the Marine Site Risk Index highlights many of the problems inherent in developing any sort of subjective risk index, and proposes solutions (particularly the uncertainty estimation) to common problems of interest to the community. It therefore has intrinsic value in excess of its use as a test case and illustrative example of a possible application for the PPMS GeoDB Product Specifications.

For the selection of which dataset to use as testing data, the availability of the DEEPP database (kindly provided by Dr. Alcaro) represented an optimum candidate for many reasons:

- It is relatively recent (2007);
- It was structured for use in risk assessment; and
- It covers a sufficiently large region to be significant for the project scope.

The DEEPP database was therefore used as a base, and much publicly available data have been collected in order to update it, as well as to integrate missing or partially erroneous information. In order to provide information useful for the MaSiRI refinement steps, two datasets were requested from the respective authors (and were kindly provided):

- Wind and Wave Atlas for the Mediterranean Sea, an atlas based on wind and wave model data from the European Centre for Medium-Centre Weather Forecasts (ECMWF) models, calibrated by means of satellite altimeter measurements (Cavaleri, 2007);
- Global Patterns and Predictors of Marine Biodiversity Across Taxa, a study based on global patterns and predictor of species richness across 13 species groups (Tittensor et al., 2010).

This latter data source represents a good candidate for the adoption as a global-scale refinement. In fact, it provides a useful and reliable overview for the global marine diversity that can be easily used as a classifier and a proxy for describing the vulnerability of the marine habitat surrounding a PPMS in any part of the world. However, the same advantage of being global entails the disadvantage of a low spatial resolution. The applied classification is similar to the one adopted in the Barrett Risk Index (Table 10), but with two main differences:

- The classes span the range ±10, in order to provide a decrease in the resulting MaSiRI in case of an area with low marine biodiversity;

- The primary coastal taxa are used instead of the average across all taxa (i.e., both coastal and oceanic taxa), since the main concerns about PPMSs are related to the effects on the coastal environment.

The adopted classification for marine biodiversity, based on the Tittensor Index for primary coastal taxa (Tittensor et al., 2010), is reported in Table 18. This index (which ranges from 0 to 0.75) examines global patterns and predictors of species richness across 13 major species groups (ranging from zooplankton to marine mammals) and highlights two major patterns (coastal species showed maximum diversity in the Western Pacific, whereas oceanic groups consistently peaked across broad mid-latitudinal bands in all oceans). Although these source data are judged highly reliable, the *RLC* for the related MaSiRI value is incremented by only 5, due to the spatial resolution issue. The adoption of a different approach, using some regional or local study, could be evaluated with a bigger contribution to the *RLC*, but with the probable disadvantage of losing the characteristic of being globally applicable.

Table 18 – MaSiRI refinement classification of marine biodiversity based on the Tittensor Index for primary coastal taxa.

Tittensor Index	Points
0.00 – 0.15	-10
0.16 – 0.30	-5
0.31 – 0.45	0
0.46 – 0.60	+5
0.61 – 0.75	+10

The Wind and Wave Atlas for the Mediterranean Sea represents a good example for a regional refinement applied to MaSiRI's output from the core criteria. In fact, the intrinsic characteristic of the Mediterranean (a relatively small and closed basin surrounded by fragmented lands and a complicated orography) requires a spatial resolution higher than the one usually provided at global scale for meteorological analysis.

The two basic sources of data for the Atlas have been the altimeters on board of the ERS-1/2 and TOPEX/Poseidon satellites and the operational wind and wave results of the European Centre for Medium-Range Weather Forecasts (ECMWF). The data have been combined providing a ten year data-set from which both overall and point by point statistics have been derived based on 239 points. Among the many types of statistics provided for each of these points (e.g., wave height, peak period) only the direction of major distribution for the mean wind value (prevailing winds) was evaluated for the refinement classification. This approach should provide results at least comparable with other Atlases existing in different regions or at global scale. The definition of the five classes for the deduced prevailing winds is simple and spatially oriented (Table 19) using the direction of connection between the PPSW and the point of minimum distance on the coast (Figure 39).

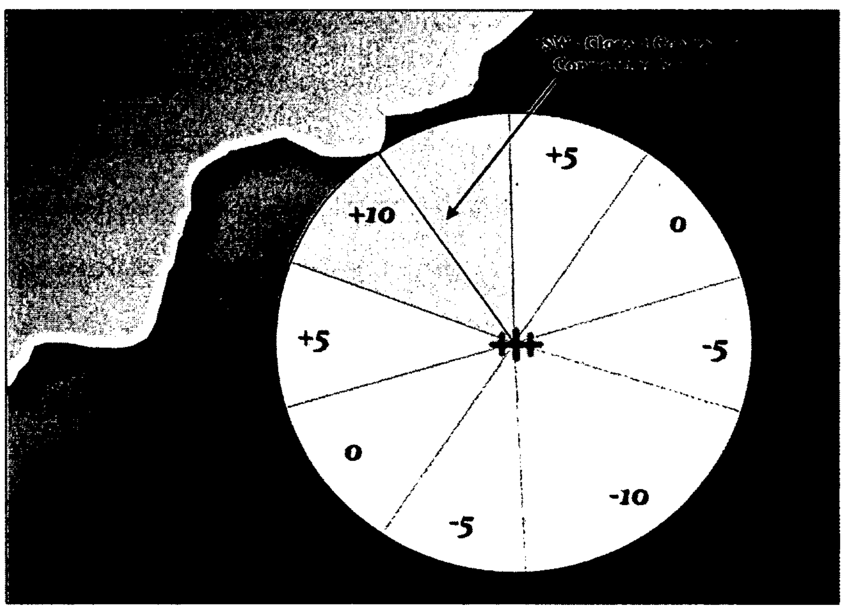


Figure 39 - Spatially oriented classification for the prevailing winds using the direction of connection between the PPSW and the point of minimum distance on the coast.

There are two main weaknesses in the described approach:

- The PPMS could be surrounded by coast, that is, not just on one side (e.g., within a bay, presence of islands): even if this situation could be approached by combining (and normalizing in the proposed range) multiple spatial classification applied to each coastal 'side', it was judged much simpler and still effective to leave to the *ECF* the duty to deal with particular cases like these;

- The direction on an eventual floating polluting dispersal does not always follow the wind direction: the segment direction used as a reference for the classification may require to be rotated by several degrees to match the actual direction of the elongated pollutant slick (Papadimitrakis et al., 2006; Stolzenbach et al., 1977). The evaluation of such an angle could come from the evaluation of a series of possible effects such as, for instance, the presence of an estuary, wind-induced coastal upwelling, surficial currents, etc. (Lentz, 2004; Tkalich and Xiaobo, 2001). These additional evaluations should be part of a local refinement or, again, compensated by the adoption of an *ECF*.

From a more general point of view, the development of a more detailed and parametric (e.g., seasonal) relationship between the direction of an oil pollutant release and atmospheric forcings over a given area is outside the scope of this work. As described in Appendix C, there are several models available for predicting the trajectory and dispersion of oil spills at sea (e.g. NOAA OR&R GNOME (NOAA, 2012)). Based also on these considerations, this refinement contributes +5 points to the *RLC*.

Table 19 – MaSiRI refinement classification of prevailing winds based on the Wind and Wave Atlas for the Mediterranean Sea; the angular values are referred to the line connecting the PPMS and the closest point on the coast.

Prevailing Winds Orientation	Points
> +144° < -144°	-10
+108° – +144° -108° – -144	-5
+72° – +108° -72° – -108°	0
+36° – +72° -36° – -72°	+5
±36°	+10

Two additional refinements (both evaluated as a 5-points contribution to the RLC) are based on:

- Evaluation of corrosion effects - shown combined with depth and submergence period as described in Table 20 - was based on the general indications provided in Appendix A. This refinement, potentially global, had to be regionalized for the Mediterranean.

Table 20 – MaSiRI refinement classification of corrosion effects with time and depth as parameters.

Corrosion Index	Points (depth ≤ 20m)	Points (depth > 20m)
0 – 1 years	-5	-10
1 – 5 years	0	-5
5 – 20 years	+5	0
20 - 50 years	+10	+5
> 50 years	+10	+10

In fact, this basin is grouped within calm seas with low tides where the most severe corrosive attacks are usually found within the first few meters above sea level (conversely, in agitated seas with high tides, serious corrosion attacks can occur at over 10 m above sea level) (Bertolini et al., 2002);

- Site depth: evaluated inversely to the Barrett Risk Index, as a decreasing contribution to the MaSiRI with the increase of depth as a result of a lower likelihood that the pollutant may reach the coast and other marine resources (e.g., the high pressure at depth and the lower exposure to atmospheric events can preserve the pollutant content from starting to leak) (Table 21). This evaluation is also based on the consideration that future technological development will make mitigation operations at depth less expensive.

Table 21 – MaSiRI refinement classification of site depth as inverse likelihood to reach the coast and other marine resources.

Site Depth Index	Points
> 1000 m	-10
400 - 1000 m	-5
100 - 400 m	0
30 - 100 m	+5
< 30 m	+10

4.5.1 Examples

As examples of the process, details related to MaSiRI application are provided for the following shipwrecks:

- The Very Large Crude Carrier (VLCC) *Haven*;
- The cargo ship *Equa*;
- The battleship *Roma*.

VLCC Haven

The VLCC *Haven* is one of the biggest wrecks in the Mediterranean at 344 m length overall (Figure 40). The main body lies on a muddy seafloor off the coast of Arenzano (Genoa, Italy) at a depth of about 75 m (Figure 41), whilst a part of the bow with the bulb is at more than 500 m depth (Masetti et al., 2010).

Figure 40 - The VLCC *Haven*: underway (on the left) and at sinking time (on the right) (sources: *Wikipedia* and *www.haven.it*).

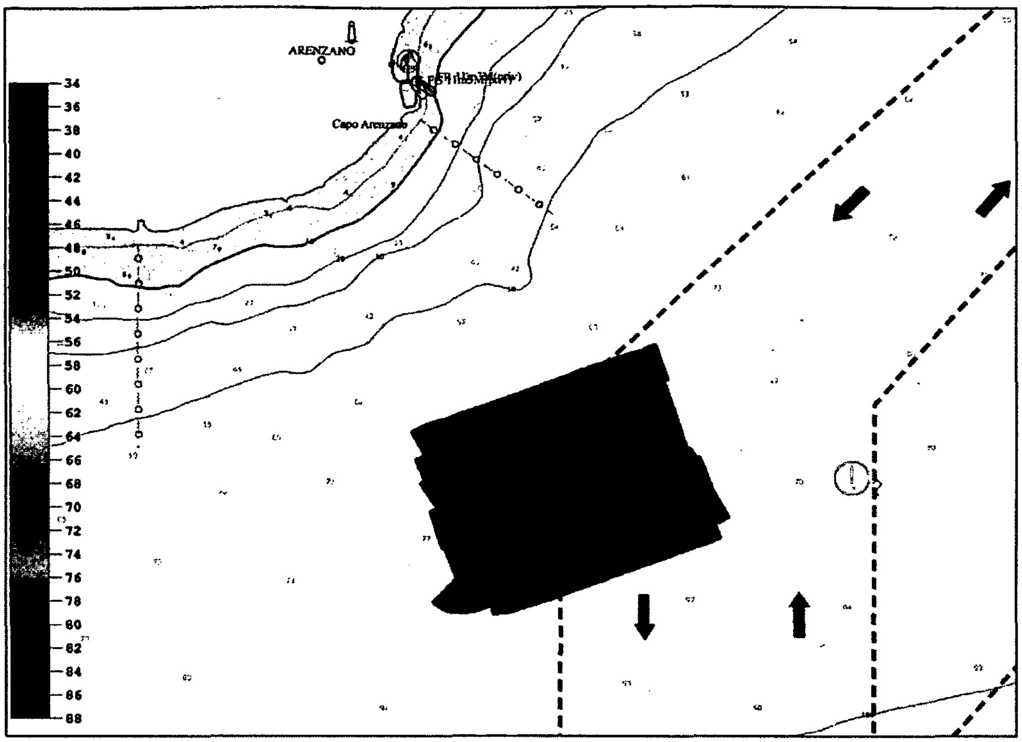


Figure 41 - Bathymetric map (color scale in meters) of the main body of the VLCC *Haven* shipwreck's site.

The sinking – dating back to April 14th, 1991 – was due to a fire caused by a violent explosion abreast of the forward tank. After the sinking, the combusted and semi-combusted residues were removed from the seashore and the seabed as far seaward as the 10-meter isobath, and were pumped from the shipwreck tanks and premises. The residues present on the seabed at a depth beyond 10 m have been left to environmental auto-decomposition processes. Part of the damage compensation was employed to perform both decontamination operations on the shipwreck and interventions of environmental restoration of the sea and coast area, which were most damaged by the accident's harmful effects (Amato, 2003; Masetti et al., 2011).

The application to MaSiRI both for Core Criteria and the described refinements is straightforward due the amount of information available for this PPSW (Amato, 2003). The shipwreck decontamination operations had the target to eliminate an estimated amount of 102 MT of hydrocarbon trapped into the wreck (with a tolerance of 2 MT) and they formally ended on June 12th, 2008 (Liguria, 2008). However, for such a big and complex PPSW it is very difficult to evaluate the actual amount of remaining pollutant. In addition, it was judged necessary to apply to the final result an ECF of 5 due to the known but not 'refinement-evaluated' information about the heavy and well documented damage suffered by the vessel at the time of sinking (Table 22).

Table 22 – MaSiRI application to the VLCC *Haven* shipwreck.

Factor	Value	Class Index	RLC
Distance from coast	1.2 NM	1	//
Pollutant Volume	< 2 MT	1	70
Type of Pollutant (hydrocarbon/chemical)	Heavy/Medium crude oils HFO/IFO	10	//
Core Criteria Result		**10**	**70**
Marine Biodiversity *(refinement)*	0.21	-5	+5
Prevailing Winds *(refinement)*	-40°	+5	+5
Corrosion Effects *(refinement)*	21 years	+5	+5
Site Depth *(refinement)*	75 m	+5	+5
Expert Correction *(refinement)*	Highly damaged at the time of sinking	+5	//
Final Refined Result		**25**	**90**

The MaSiRI output of 25 (with a high RLC of 90) largely differs from the risk evaluation present in the original DEEPP database (i.e., 'serious') since the decontamination operations were subsequent to the information collection phase of that database.

Cargo Ship Equa

The *Equa* was an Italian cargo ship armed and converted to an auxiliary cruiser (and renamed AS105) at the beginning of WWII. Its main aim was to carry mines and depth charges to be used against Allied submarines in Ligurian and Tyrrhenian Seas. The wreckage was due to a collision with the German Unit *UJ220*, due to low visibility, during the night between the 18th and 19th of April, 1944 (Toja, 2004).

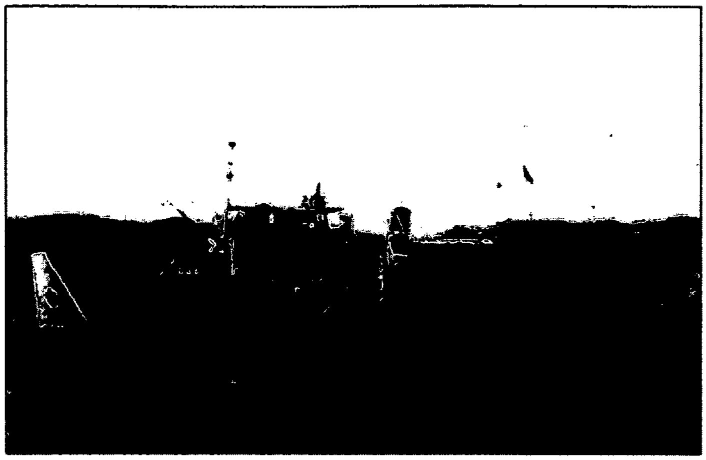


Figure 42 - Photo of the Cargo Ship *Equa* underway (source: *Historic Office of the Italian Navy*).

The wreck is well known by recreational divers since it is quite spectacular, being in good condition, still showing part of its weapons, and partially covered by fishing nets (Figure 43). The *Equa* wreck site lies off Riomaggiore (Italy) in navigation trim on a sandy bottom, at a depth of about 36 m. The wreck was recently cleared of ordnance by the Italian Navy (Toja, 2004).

Figure 43 - Recent underwater photo of the *Equa* wreck (source: *www.scubaportal.it*).

Table 23 – MaSiRI application to the Cargo Ship *Equa* shipwreck.

Factor	Value	Class Index	RLC
Distance from coast	2 NM	1	//
Pollutant Volume	< 100 MT (estimated)	1	40
Type of Pollutant (hydrocarbon/chemical)	Diesel-Fuel-Kerosene	5	//
Core Criteria Result		**5**	**40**
Marine Biodiversity *(refinement)*	0.21	-5	+5
Prevailing Winds *(refinement)*	-130°	-5	+5
Corrosion Effects *(refinement)*	68 years	+10	+5
Site Depth *(refinement)*	34 m	+5	+5
Expert Correction *(refinement)*	The good condition of the wreck are reported by many recreational divers	//	+10
Final Refined Result		**10**	**70**

The MaSiRI estimate is given in Table 23. Although the pollutant volume was estimated from the vessel tonnage (there are no official data or study applied to this wreck), an adjunctive correction of 10 points was applied to the refined result justified by the good condition of the wreck. This information comes from a large number of opportunistic 'inspections' provided by the logs and the underwater images taken by many recreational divers.

The refined result of 10 (with a RLC of 50) is slightly different from the 'moderate' evaluation of the risk provided in the original DEEPP database. However, the MaSiRI result seems to be sufficiently coherent with the type and amount of pollutant that can be present in a relatively small vessel (i.e., 243 GRT and a length overall of 39 meters) as the *Equa*.

Battleship Roma

This PPSW is the wreck of an Italian battleship of 240 m of overall length and a displacement of more than 40,000 MT. Her wreckage represents a particularly sad occasion since she was hit the day after the WWII armistice (September 9th, 1943) by the first free-fall radio-controlled bombs (Fritz X) in military history and, on that occasion, two admirals, the ship's captain and more than 1200 crew members died. Thus, this PPSW represents a war grave (details about the legal aspects of this status are provided in Appendix B).

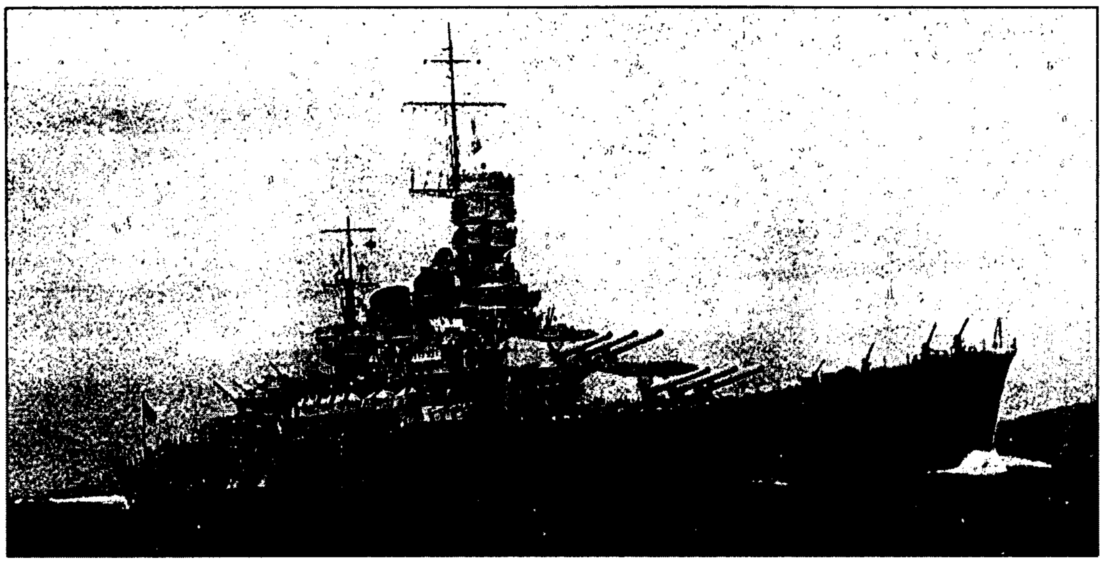


Figure 44 - The Italian Battleship *Roma* underway (source: *Wikipedia*).

The *Roma*'s shipwreck laid undiscovered (notwithstanding a number of expeditions by researchers such as J.J. Cousteau) for many years in the Gulf of Asinara (Sardinia, Italy). She was partially identified on June 17th, 2012 at 1000-meter depth and about twelve nautical miles off the coast (Figure 45)(Donati, 2012). No data about its pollutant content are available at this moment, thus the estimation approach was adopted. In this particular case, due to her GRT, it was necessary to apply low and high estimates to define a range for the MaSiRI output.

Table 24 – MaSiRI application to the RN *Roma* shipwreck.

Factor	Value	Low Estimate	High Estimate	RLC
Distance from coast	12 NM	4	4	//
Pollutant Volume	<500 MT (Low Est.) >2500 MT (High Est.)	5	10	20
Type of Pollutant (hydrocarbon/chemical)	Diesel-Fuel-Kerosene	5	5	//
Core Criteria Result		**6**	**13**	**20**
Marine Biodiversity *(refinement)*	0.33	0	0	+5
Prevailing Winds *(refinement)*	+130°	-5	-5	+5
Corrosion Effects *(refinement)*	68 years	+10	+10	+5
Site Depth *(refinement)*	1000 m	-5	-5	+5
Expert Correction *(refinement)*	Only the position of some parts of the wrecks were discovered	0	+10	//
	The wreck is not cleared of ordnance, even if they mostly exploded at the time of sinking	0	+10	//
Final Refined Result		**6**	**33**	**40**

The MaSiRI output for this PPSW represents a clear case where more investigations are required. In fact, the analysis presents a risk comparable with the previously examined (and much smaller) *Equa* (for the low estimate) but bigger than the *VLCC Haven* as high estimate. In this case, it is also useful to have a Risk Level of Confidence that underlines the lack of information that is behind the values used for the risk index calculation.

Figure 45 - One of the Roma's 90mmm guns discovered last summer.

4.5.2 Whole dataset

The original DEEPP database was integrated and updated with information available from open sources (e.g., newspaper articles, recreational divers' websites). The data import and integration operation was provided using the PPMS GeoDB Manager, and the Risk Index steps were calculated through its SQL query composer. Finally, the output was made publicly available using *MapServer*, using the web architecture described in the next section.

The total number of records reaches almost five hundred, whereas the DEEPP database was of 432 wrecks (Alcaro et al., 2007). Due to the high number of unknown PPSWs (more than 60%), it is possible that there are a limited number of duplicate records. They are almost equally located among French and Italian waters. The largest number (57%) are located both in shallow waters (less than 50 m) and close to the coast (less than 2 nautical miles). This spatial distribution may have different interpretations such as:

- The presence of grounding obstacles along the shoreline;
- Ambush by submarines outside harbors during WWII;
- The ease of locating a shipwreck in relatively shallow waters; and
- The higher presence of recreational diving centers along the coast.

The large amount of high quality information (e.g., underwater images and videos) from recreational divers available on the web underlines the value that this type of information is able to provide.

The distribution of high (Figure 46) and low value (Figure 48) estimates coming from the MaSiRI application is provided. Within the 'known' part, the majority of the assessment is represented by MaSiRI values between 15 and 50.

This output only slightly differs from the results described in the DEEPP project (Alcaro et al., 2007) that was re-charted on the same type of pie chart for easy reading (high estimates in Figure 47 and low estimates in Figure 49). Furthermore, just for this sort of 'comparison', the MaSiRI outputs were classified using a codification similar to one followed in the DEEPP Project (and based on an evaluation of possible combination obtained using only the Core Criteria):

- Minor to moderate: MaSiRI ≤ 5;
- Moderate: 5 < MaSiRI ≤ 15;
- Moderate to serious: 15 < MaSiRI ≤ 50;
- Serious: MaSiRI > 50.

The adjunct information provided by the Risk Level of Confidence to the MaSiRI values accomplishes the scope of providing the risk quality dimension (level of confidence) totally missing in the DEEPP project.

However, it is judged worthwhile to underline how this work describe a possible methodology to develop a risk index application built on top of the PPMS GeoDB, while the scope of the DEEPP Project was *"to provide European coastal States and National Administrations with guidelines and criteria to face the environmental threats which might arise from potentially polluting shipwrecks"* (ICRAMM-CEDRE, 2007). Thus, the database updated during the PPMS GeoDB development was mainly used to validate:

- The conceptual and physical models for the adopted data structure;
- The methodology and the workflow for a marine risk assessment tool based on a PPMS GeoDB; and

- Data storage and management through the PPMS GeoDB Manager.

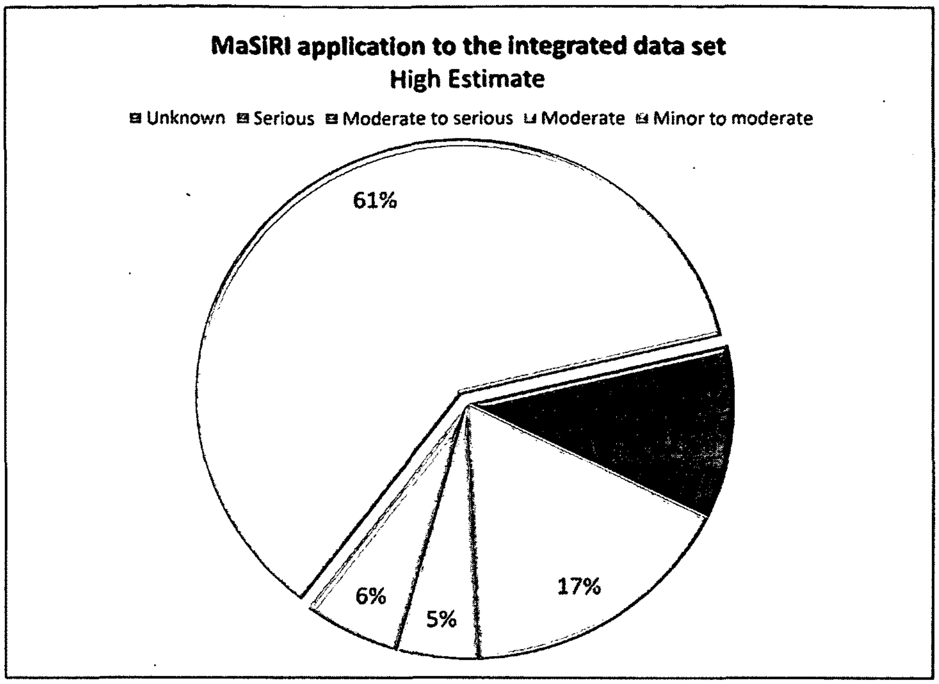


Figure 46 - High-estimate values resulting from the application of MaSIRI to the test data set.

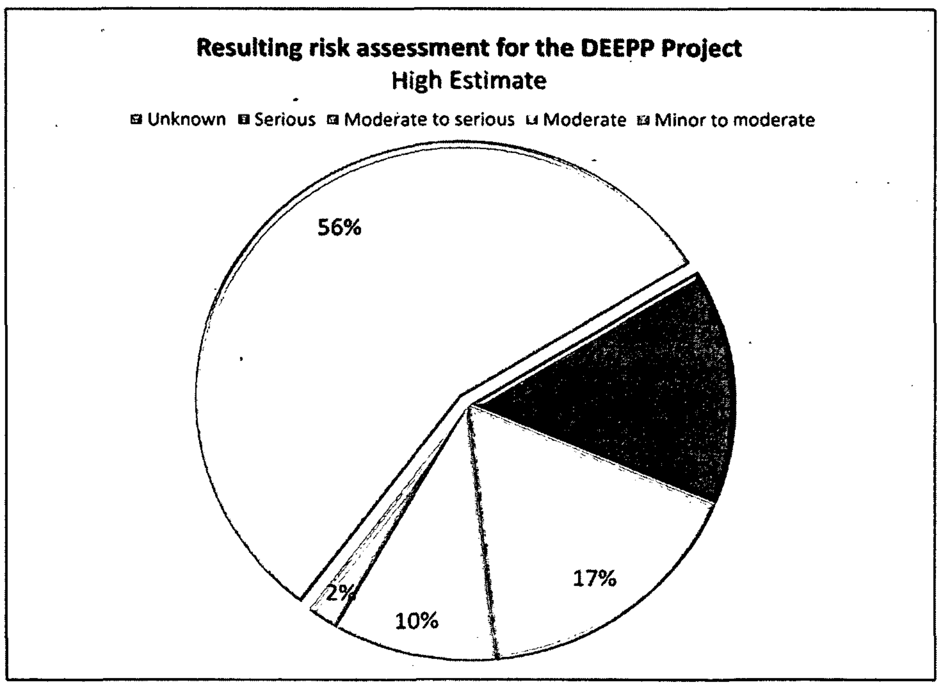


Figure 47 - Risk assessment (high estimates) from the DEEPP risk assessment over the original data set.

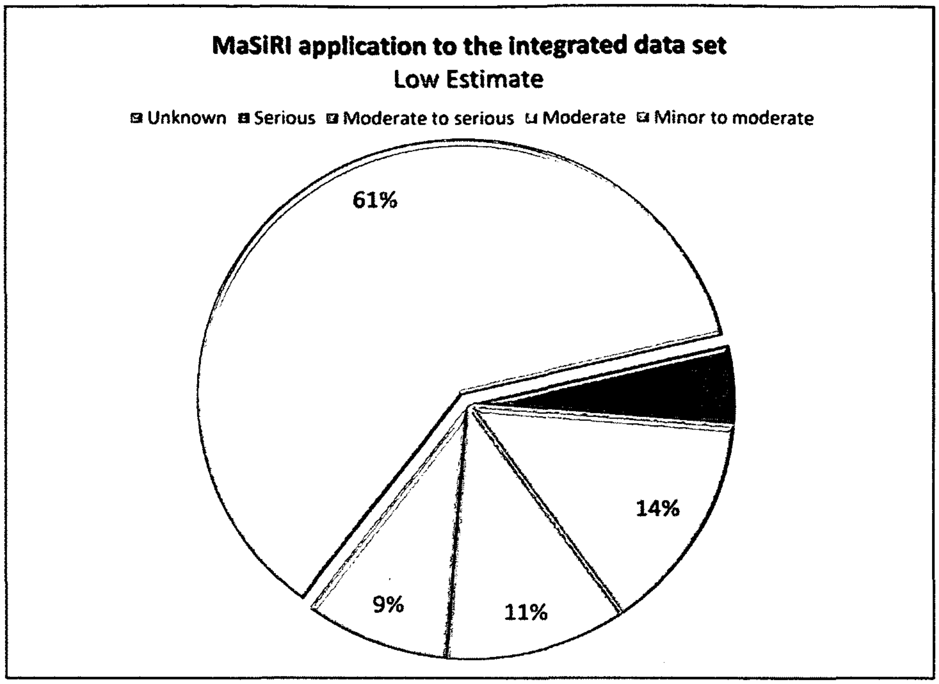


Figure 48 - Low-estimate values resulting from the application of MaSiRI to the test data set.

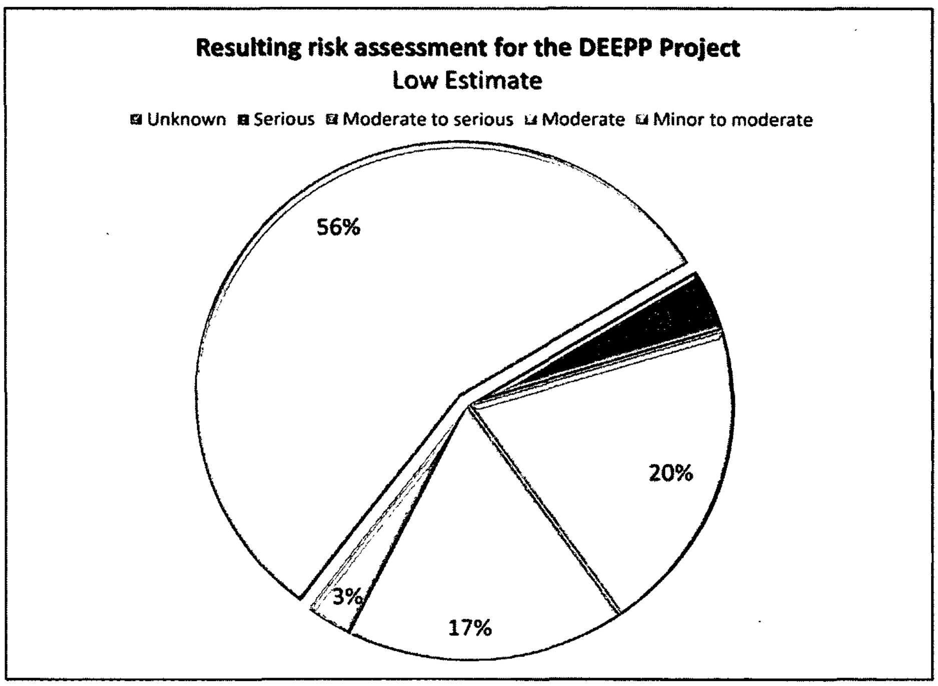


Figure 49 - Risk assessment (low estimates) from the DEEPP risk assessment over the original data set.

As a consequence and although the results are similar (and may be more complete with the RLC feature), the obtained values of risk assessment for the study area represent only a secondary product of this work with limited aspiration of general completeness and validity.

Since the proposed data structure does not only contains PPSWs, publicly available data sets were downloaded and used to populate several `MarineResources` and `ComplementaryInfo` objects:

- Global, Self-Consistent, Hierarchical, High Resolution Shoreline Database: amalgamates two databases in the public domain: World Data Bank II and World Vector Shoreline (Wessel and Smith, 1996);
- Flanders Marine Institute's (VLIZ) Maritime Boundaries Geodatabase: contains two shapefiles (VLIZ, 2012):
  - A polyline file representing the world maritime boundaries;
  - A polygon file with the Exclusive Economic Zones.

Although the relevance to have the above spatial boundaries is quite evident, several legal aspects related to the Coastal and Flag states, as well as the different international rules related to the various international marine areas, are reported in Appendix B.

4.6 The web GIS deployment

This work may be seen as a journey that started with the modeling of a geographic database for Potentially Polluting Marine Sites and continued through its physical implementation, a proposal of data management and a risk index development built on top of the developed Product Specifications. The end of this journey is represented by a solution to deploy both the source and the processed data (partially or completely) on the web for general access by the public.

Among the many possible solutions for building a spatial web application, the *MapServer* project was selected. This project has been open source since 1999 and was initially developed by the University of Minnesota with support from NASA. Written in C, it runs on all the major platforms (MS Windows, Linux and Mac OS X) with the main aim being to render geographic data as image maps.

This selection was based on the following reasons:

- It supports industrial standard data formats and spatial databases;
- It can run with almost any web server and supports the more popular programming languages;
- It is relatively easy to extend and modify the source code to do exactly what the user needs;
- It has a wide user community; and
- It is free and open source.

The *MapServer* solution was fully integrated into the PPMS GeoDB project's website previously developed using only HTML, CSS, Java and JavaScript. The existing functions of the website were:

- Outline the content of all of the parts of this project;
- Provide a repository for PPMS GeoDB XML Schemas;
- Store the proposed drafts for a PPMS GeoDB S-1XX Product Specification;
- Display the Feature Catalogue of the proposed `ppms` domain; and
- Fulfill the URI requirements for being a valid namespace for all the code related to the PPMS GeoDB project: `http://ccom.unh.edu/ppms`.

The *MapServer* added public access to a demo applied on the used test data set. In addition, *MapServer* provides the possibility to deploy the data layer as a WMS or as a non-transactional WFS (that is, transaction requests are not supported and thus the features are read only).

4.6.1 Adopted architecture

The adopted *MapServer* architecture (Figure 50) can be described using the following levels:

- Input geographic data: various OGC standard as well as raster and vector formats, both natively and through the *Geospatial Data Abstraction Library* (GDAL) and *OpenGIS Simple Feature Reference* (OGR) library. Since ESRI *Shapefiles* are effectively the native and highest performance *MapServer* format, they were selected when possible. The *tileindex* (a file containing footprints of raster data coverage) was used to allow *MapServer* to only read the data that intersects the requested map extent, rather than reading all of the data. Other web services (e.g., WMS, WFS) have also been provided as possible data inputs selectable by the web user. Geographic coordinates in WGS 84 were preferred if available; however, *MapServer* provides on-the-fly map projection using the *Proj4* library;
- Configuration file: *MapServer* uses *Mapfile*s that are in an XML-like hierarchical declarative language used to define map settings, styling and server directives. The adopted *Mapfile* contains a series of objects, each one has several parameters and may have other objects as members;
- Application: *MapServer* is a Common Gateway Interface (CGI) program that lies inactive on almost any web server until it receives a request that is a URL containing information about the requested map (it is also possible to use *MapScript*, an alternative program that wraps the C API to that it can be used with many other languages such as Perl);
- Web server: Apache installed on the CCOM-JHC *CentOS* server that is used to serve the HTML page;
- Output products: HTML pages with images representing maps (with scale bar) and a legend for symbols.

An HTML page also represents the main interface between the web user and the implemented *MapServer* architecture. User inputs through that page generate a request that is processed by the CGI program, called *mapserv*. This program, being state-less, must receive in each request all the context information as variables. Using those variables, *MapServer* populates a template HTML file that controls the output for maps and legends. The same template provides commands (e.g., zoom, query) to the user.

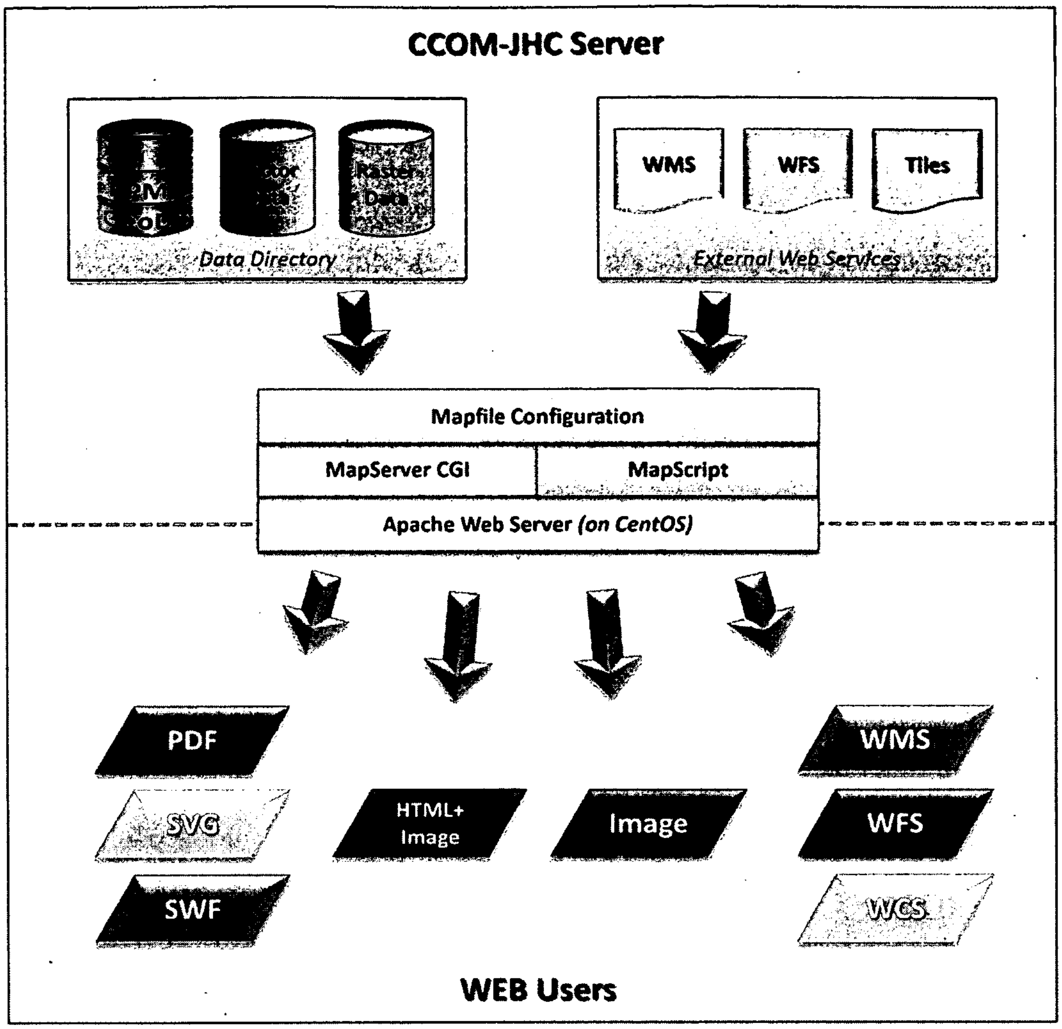


Figure 50 - *MapServer* architecture adopted on the CCOM-JHC server.

4.6.2 Quality Symbol

During this last phase of the PPMS GeoDB development the problem of how to display the information provided by the MaSiRI application was evaluated. The issue is how to show the data in such a way that users could visually evaluate the outputs in the case of a large number of PPMSs present in a given area.

The peculiarity of this problem is that the graphic interpretation of the MaSiRI values without the related level of confidence was considered possibly misleading. Some commonly used solutions were evaluated:

- A dedicated layer for the level of confidence to be rendered at the user's request or displayed with some level of transparency; or
- A pop-up window, a new browser page or a common area on the control panel where the values representing the level of confidence could be visualized.

These solutions were judged sub-optimal from the user's point of view since they can be a source of confusion (e.g., in case of crowded areas) and key information such as "which of the high risk index values is reliable?" are not easy to obtain.

The proposed solution is a combined symbol, called the Quality Symbol, which is a composition of two parts (Figure 51):

- The 'value symbol', a circular area providing the information of the risk level being filled by a color in a color map between green and red. This color map was based on:
  - The sensitivity of blue cones in the human eye is much lower than the others and the selected scale must have mutually exclusive boundaries (Hill and Kelley, 2007);
  - The widely accepted social convention that 'red' represents a dangerous situation, 'green' the absence of hazard, and 'yellow' a potentially risky situation (e.g., the color of traffic lights);
- The 'meta-symbol' (for similarity with the 'metadata' term) that provides information about the quality (in this case, the level of confidence) for the value reported in the 'value' symbol part (Figure 52). This part of the symbol is an annulus surrounding the 'value symbol', has a light grey as background and is filled in black (clockwise from north) in proportion to the confidence (since human color perception is particularly sensitive to changes in the perceived intensity of reflecting objects).

The external radius of the 'meta-symbol' is 10 % larger than the radius of the 'value symbol' (Figure 53). This value has been judged during development as a good tradeoff between the two components of a quality symbol.

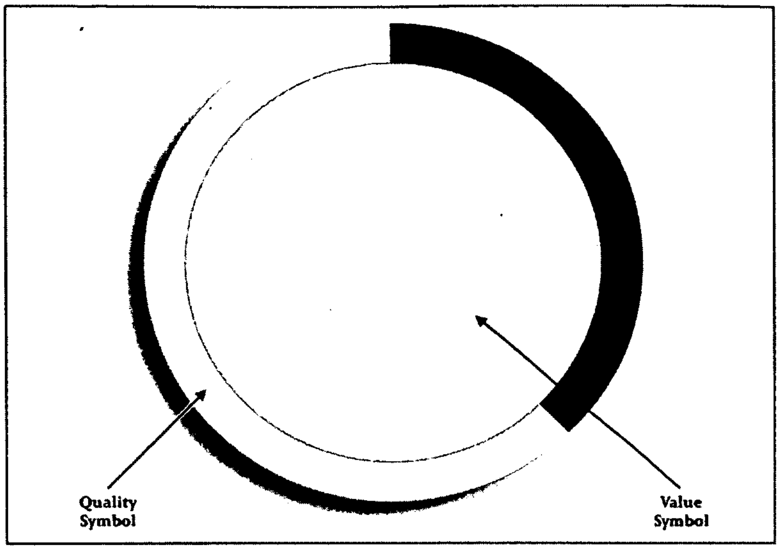


Figure 51 - Example of medium MaSiRI value with indications of the two parts of the combined symbol: the Quality Symbol and the Value Symbol.

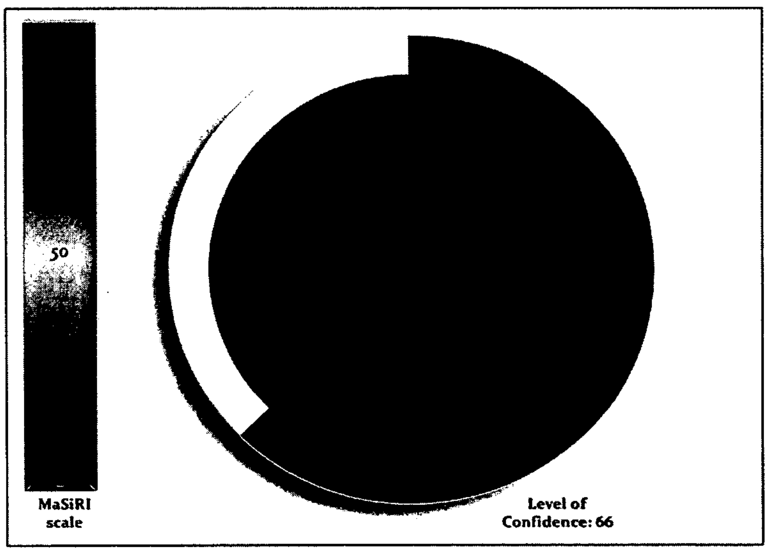


Figure 52 - Example of low MaSiRI value with an associated Level of Confidence of 66; on the left, the color map adopted for the MaSiRI values.

The meta-symbol part for the MaSiRI values is only displayed at large scale, in order to preserve clarity of visualization and to provide the quality information to the user only when it is useful. This combined symbol displays in a local area both the risk index evaluation and the linked level of confidence, without the need for any additional operation from the user.

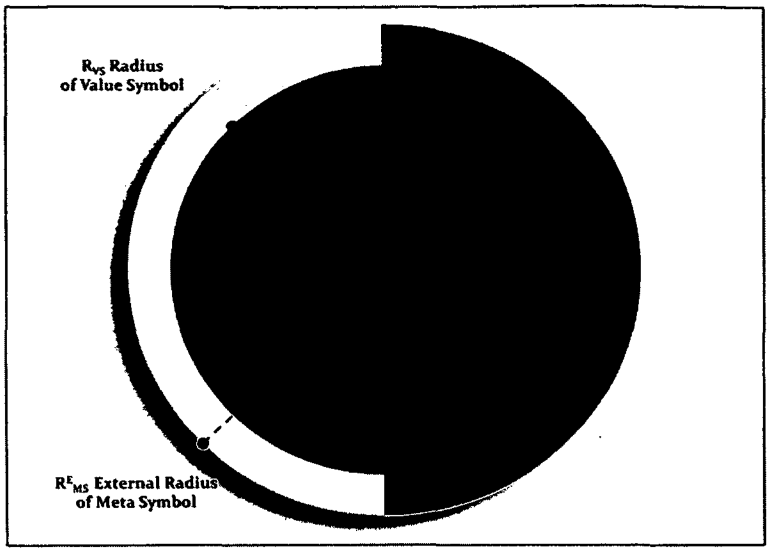


Figure 53 - Example of high MaSiRI value expressed using the combined Quality Symbol with indications of external radius (for the Quality Symbol) and radius (for the Value Symbol).

An attempt to provide information on the range of values present in the MaSiRI database for a particular PPSW was made with the target to display both the high and low estimates of MaSiRI on the same symbol at one time. The results were judged not satisfactory and dismissed due to confusion generated by the presence of too many colors in an already crowded area. Thus, so far, a possible solution for the visualization of these information is clustering them in low and high MaSiRI estimation layers, each of them using a Quality Symbol as previously described.

In addition to providing an additional dimension for data quality representation, the Quality Symbols are relatively easy to render, and work well within the example *MapServer*'s visualization features. This makes them efficient to implement, which is another spur to adoption.

4.6.3 Web user interaction

Figure 54 provides a screenshot of the developed web GIS application from the user point of view.

The *MapServer* features have been fully integrated within the project's website (e.g., the top part in Figure 54 represents a common header and shared menu among all the web pages).

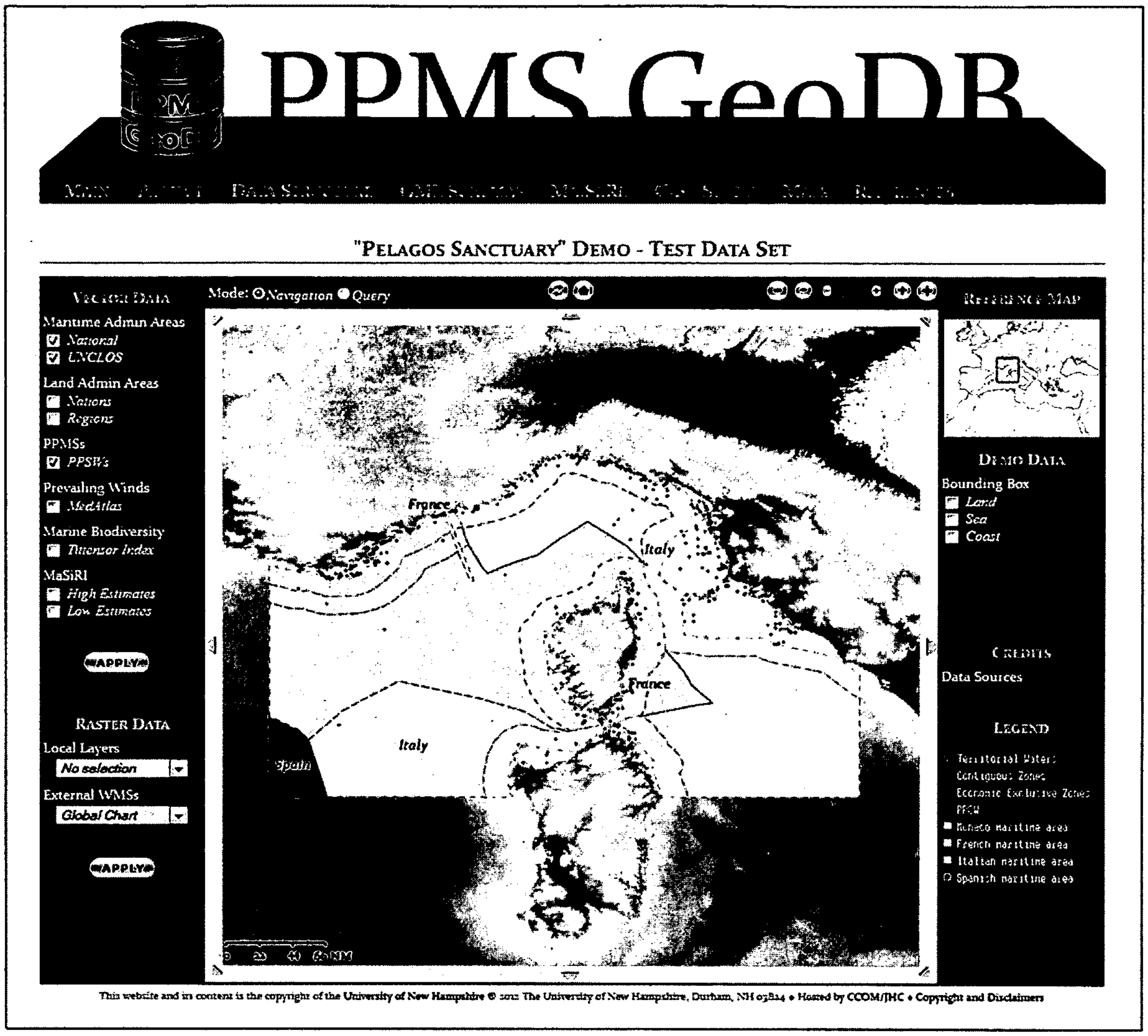


Figure 54 - Screenshot from the PPMS GeoDB Project's website showing the user web interface and the output coming from the activation of several layers.

Only the bottom of this screenshot contains the *MapServer* features (although this application uses a template for the whole page at interaction time):

- The data layers menu (on the left), with a selection of vector, raster and external WMSs;
- The main map (in the middle) with a scale bar and a series of commands (e.g., zoom in/out with different zoom factors, directional pans, etc.);
- The information menu (on the right) with a reference map to locate the extent of the shown main map within Europe, the data set bounding box, the data source credits and the legend (referred to the activated vector data layers).

At the moment, the technology adopted for the website is simply HTML, CSS and a few JavaScripts. The only Java piece of code is primarily used in 'query mode' (the activation command is at the top-left of the main map) to spatially define through mouse interaction the query area. This spatial query (either on a point or in a rectangular area) can provide empty, single or multiple results. An example of single query result in provided in Figure 55.

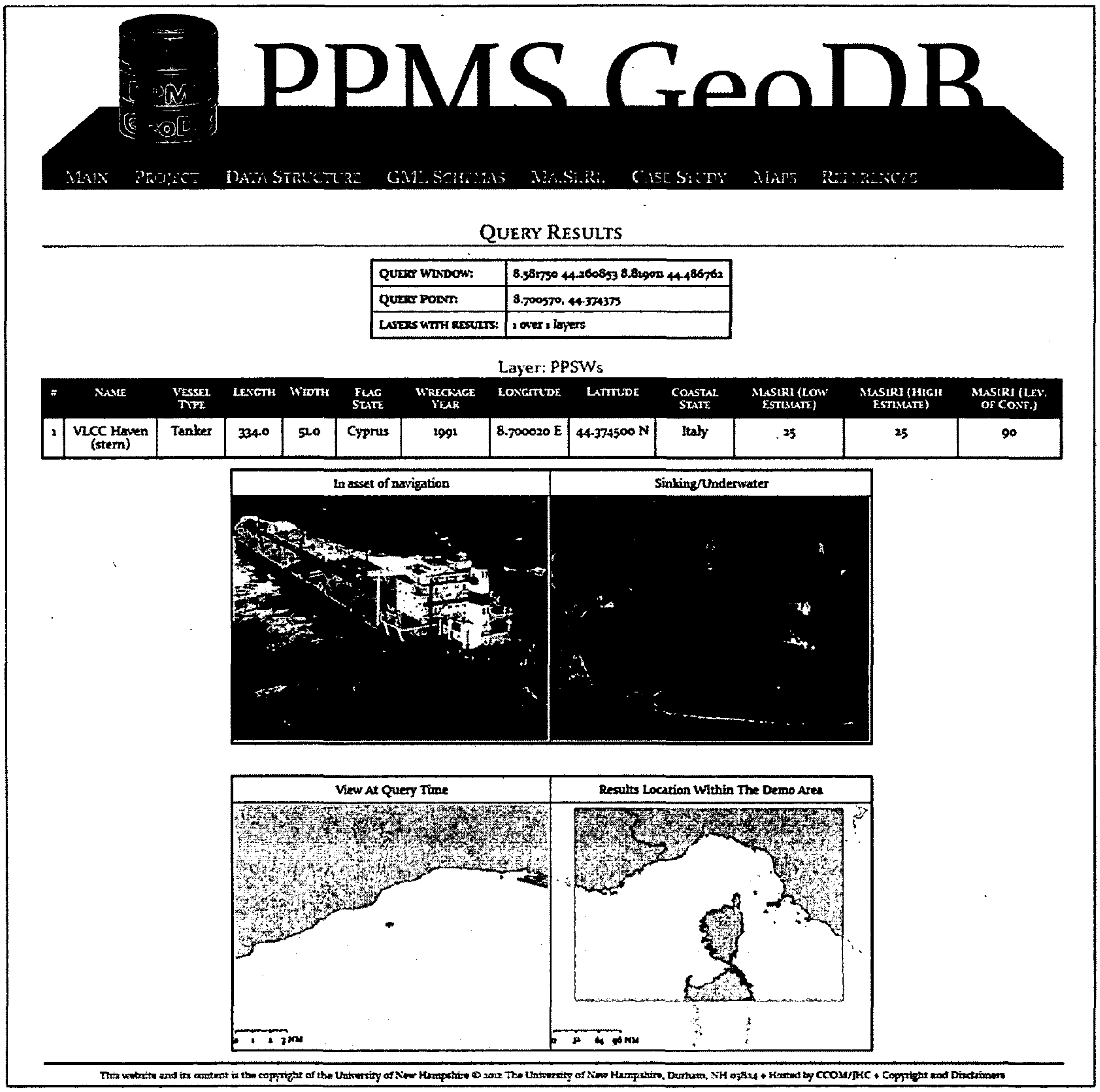


Figure 55 - Example of query with a single result: some selected field are displayed together with PPSW pictures (if available) and maps generated by the *MapServer* rendering engine.

Finally, the Quality Symbol implementation is shown in Figure 56 within an area characterized by the presence of two PPSWs.

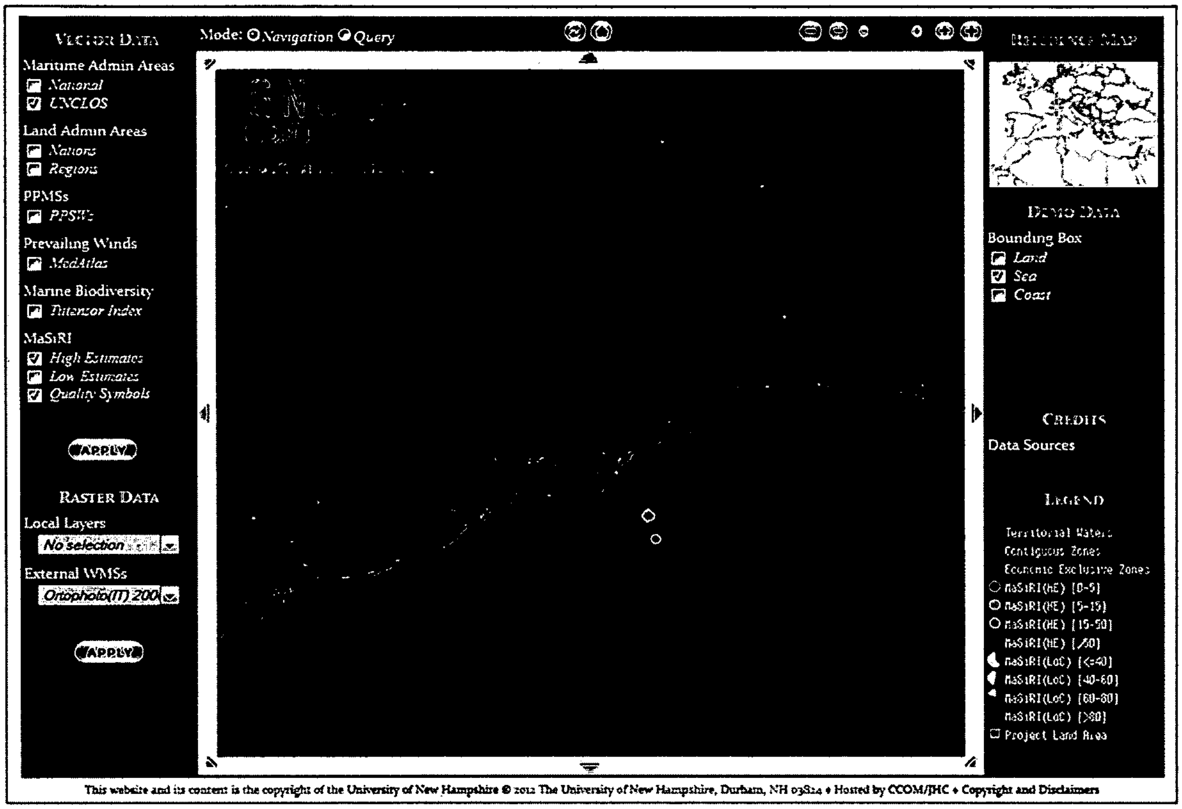


Figure 56 - Screenshot of an area characterized by the presence of two PPSWs. The results from MaSiRI application are visualized using Quality Symbols, which use the inner part of the symbol to represent the risk index value, while the outer part provides information about the estimated quality (level of confidence) of the MaSiRI value.

CONCLUSIONS

Recent pollutant releases from PPMSs have resulted in significant impacts, including loss of marine life, economic impacts to coastal areas, and high cost to mitigate the effects. Standardized collection of information about these sites can contribute to monitoring and reducing these events. The information collected by the proposed PPMS GeoDB represents a contribution to this issue at global and regional 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.

Developed in the S-100 framework, the database 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 PPMS GeoDB was developed to be a practical means for providing a geo-referenced picture of hazardous sites and related marine resources.

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 extensive adoption of existing IHO features and attributes – permits an easy integration with other existing HO's databases.

The release in the near future of the first PSs for the S-100 series will provide interesting elements to improve on and amalgamate with the PPMS GeoDB PS (proposed in the Annex) in such a way that this project could smoothly integrate with the growing S-100 spatial data infrastructure. Similarly, the definition of an IHO profile for the ISO metadata will represent another relevant factor to be taken into account.

Although the main target of the PPMS GeoDB Application is a PPMS inventory, its implementation represents a potential tool for each phase of the disaster management cycle: emergency response, recovery, development, mitigation, and preparedness. In addition, a risk index – representing an assessment of the magnitude of risk associated with any site – was derived to determine the potential impacts of these PPMSs using a GeoDB as outlined in the annexed Product Specifications.

The Marine Site Risk Index (MaSiRI), built on the GeoDB, enables a risk assessment of undersea threats. 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. 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. Based on an extensive analysis of previous studies, MaSiRI compensates for the highlighted weaknesses and provides a risk quality evaluation through a level of confidence that follows in parallel the definition of the risk index.

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. The definition by an international committee of a common set of global refinements could represent a key factor and also an important step towards the adoption of the proposed risk index as part of the PPMS GeoDB PS.

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.

The web GIS deployment of a PPMS GeoDB represents, without any doubt, the simplest way to disclose to the public the collected information, while the related possibility to use the data as a WMS in almost any GIS allows for better development and integration with other regionally and locally available datasets. Although promising, the idea to adopt the Quality Symbol to show the quality dimension for MaSiRI leaves open the issue of value range visualization. A further development of the proposed approach might provide a better solution than just visualizing two distinct layers.

Last, but not least, all the source code, libraries and applications used are free and/or open source. Thus, the adoption of the described suite of technologies does not have explicit cost, except the lack of support that a commercial software company can (or, better, should) provide. In my personal opinion, each new proposal for a geographic standard or application should be provided with some prototype software implementation to push the COTS to offer more than just its simplest industrial implementation. At the same time, a concrete implementation represents an impetus to adoption since, even if the original proposal is extensively modified subsequently, it at least gives any implementer a good place to start.

A preliminary article presenting the PPMS GeoDB to the hydrographic community (Masetti et al., 2012) was positively judged as an application *"[..] for monitoring and maintenance of our fragile marine environment"* providing *"[..] excellent insights into how S-100 can be used to develop the necessary data models and tools to build complex geospatial hydrographic management systems"* (Halls, 2012). This may represent a promising first step in the long journey that could result in the PPMS GeoDB Product Specifications being adopted as part of the IHO S-1XX series. This adoption would represent 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.

LIST OF REFERENCES

Aa T. J., A V. O., Marita L., M O. A., Haldor B. L., Fatima H. (2002) Investigation and risk assessment of ships loaded with chemical ammunition scuttled in Skagerrak. Forsvarets forskningsinstitutt, Kjeller (Norway), pp. 88.

Agumya A., Hunter G. J. (1999) A Risk-Based Approach to Assessing the 'Fitness for Use' of Spatial Data. URISA Journal 11(1):33-44.

Aichele R. O. (2010) RUST - Creating a new data base for maritime salvage responses. Environmental and Historical Research. Information Works Inc., Saratoga Springs, NY (USA), http://www.inforworks.com/rust1.htm.

Albarrak K. M., Sibley E. H. (2009) Translating relational & Object-Relational Database models into OWL models. In: IEEE International Conference on Information Reuse & Integration. pp. 336-341.

Alcaro L., Amato E., Cabioch F. h., Farchi C., Gouriou V. (2007) DEEPP Project DEvelopment of European guidelines for Potentially Polluting shipwrecks. EU D.G. Environment - Civil Protection Unit, Rome (Italy) / Brest (France), pp. 163.

Alexander D. E. (2000) Confronting Catastrophe: New Perspectives on Natural Disasters. Oxford University Press, New York (USA), ISBN 978-1903544006, pp. 282.

Alino P. (2012) Highlights of the Regional State of the Coral Triangle Report (RSCTR). In: Pre-Conference Symposium of the 2012 International Coral Reef Symposium. Cairns (Australia).

Allen G., Owens M. (2010) The Definitive Guide to SQLite. Apress, ISBN 1430232250, 9781430232254, pp. 368.

Amato E. (2003) An Environmental Restoration Programme 12 Year After: the Haven Wreck. In: 9th Information day of CEDRE - The Treatment of Potentially Polluting Wrecks. Paris (France), pp. 17.

Anon. (1984) Per lo Scirè una messa in mare. Recuperati i resti di 22 marinai. La Repubblica. 19 September 1984.

Anon. (1990) Exchange of Notes between South Africa and the United Kingdom Concerning the Regulation of the Terms of Settlement of the Salvaging of the Wreck of HMS Birkenhead. International Journal of Estuarine and Coastal Law:405.

Anon. (2000) Annexed to the decision of the Court of Appeal on the 4th Circuit of the United States, Sea Hunt, Inc. v. Unidentified, Shipwrecked Vessel or Vessels. 221 F.3d 643, 4th Cir. 2000.

Anon. (2002) Danni allo Scirè. Israele chiede scusa. La Repubblica. 4 Novermber 2002.

Archer C. L., Jacobson M. Z. (2005) Evaluation of global wind power. Journal of Geophysical Research 110(D12110):20.

Atay M., Chebotko A., Liu D., Lu S., Fotouhi F. (2007) Efficient schema-based XML-to-Relational data mapping. Information Systems 32(3):458-476.

Australia (1996) Diplomatic Note to the Embassy of the Republic of Turkey. In: Commonwealth of Australia.

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.

Baccicalupi G. (2008) Motonave 'Montone': "Il relitto e' senza combustibile". La Nazione. September 20, 2008, Carrara (Italy).

Baikin V. G., Chak R. O., Divin N. A., Leonov V. I. (1980) Corrosion protection of tanks for the storage of crude oil. Chemical and Petroleum Engineering 16(5):312-314.

Barada M. (2007) International and National legislation to prevent pollution caused by ships and wrecks. In: Workshop Organized by ICRAM on Potentially Polluting Shipwrecks. ICRAMM-CEDRE, Rome (Italy).

Barrett M. J. (2011) Potentially Polluting Shipwrecks. Nicholas School of the Environment. Master Project, Duke University, pp. 35.

Basta D. J. (2010) Opening Remarks. In: International Corrosion Workshop. Newport News (USA).

Basta D. J., Kennedy D. M. (2004) The need for a proactive approach to underwater threats. Marine Technology Society Journal 28(3):9-11.

Beard K. (1997) Representations of Data Quality. In: Craglia M., Couclelis H. (eds) Geographic Information Research: Bridging the Atlantic. Taylor & Francis, ISBN 978-0748405947, pp. 280-294.

Beckmann N., Kriegel H.-P., Schneider R., Seeger B. (1990) The R*-tree: an efficient and robust access method for points and rectangles. In: Proceedings of the 1990 ACM SIGMOD international conference on Management of data. ACM, Atlantic City, New Jersey, United States, pp. 322-331.

Beckmann N., Kriegel H.-P., Schneider R., Seeger B. (1990) The R*-tree: an efficient and robust access method for points and rectangles. SIGMOD Rec. 19(2):322-331.

Bedard Y. (1986) A study of data using a communication based conceptual framework of land information systems. The Canadian Surveyor 40:449-460.

Bedard Y. (1988) Uncertainties in Land Information Systems Databases. In: 8th International Symposium on Computer Cartography (AUTO-CARTO 8). Baltimore (USA), pp. 175-184.

Beddington J., Kinloch A. J. (2005) Munitions Dumped at Sea: A Literature Review. Imperial College London Consultants, London (UK), pp. 90.

Bederman D. J. (1999) The UNESCO Draft Convention on Underwater Cultural Heritage: A Critique and Counter-Proposal. Journal of Maritime Law and Commerce 30(2):353.

Bederman D. J. (2000) Rethinking the Legal Status of Sunken Warships. Ocean Development & International Law (31):98-99.

Bederman D. J. (2000) Rethinking the Legal Status of Sunken Warships. Ocean Development & International Law (31):100-101.

Benka-Coker M. O., Metseagarun W., Ekundayo J. A. (1995) Abundance of sulphate reducing bacteria in Niger Delta Oilfield waters. Bioresource Technology 54(2):151-154.

Bertolini L., Gastaldi M., Pedeferri M., Redaelli E. (2002) Prevention of steel corrosion in concrete exposed to seawater with submerged sacrificial anodes. Corrosion Science (22):1497-1513.

Bi C. (2009) Research and Application of SQLite Embedded Database Technology. WSEAS Transactions on Computers 8(1):83-92.

Bianchi G., Mazza F. (2000) Corrosione e protezione dei metalli. AIM, Milan (Italy), ISBN 978-8821400711, pp. 532.

Birkholz D. (1997) Preservation of Iron Ships in the Marine Environment. Maritime Park Association.

Bobra M. (1991) Water-in-oil emulsification: a physicochemical study. In: Proceedings of the 1991 International Oil Spill Conference. American Petroleum Institute, Washington (USA), pp. 483-488.

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

Bradley N. (2004) The XML schema companion. Addison-Wesley, ISBN 978-0321136176, pp. 312.

Brea-Franco L. O. (2000) Informe sobre el estrado de situaciòn de salvaguardia del patrimònio cultural subacuatico. In: UNESCO (ed) Background Materials on the Protection of the Underwater Cultural Heritage. Paris (France), ISBN TX0014533.

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

Brown J. D., Heuvelink G. B. M. (2007) The Data Uncertainty Engine (DUE): A software tool for assessing and simulating uncertain environmental variables. Computers & Geosciences 33:172-190.

Brown V. J. (2005) Warfare: Iraq's Toxic Shipwrecks. Environ Health Perspect 113(4):A230.

Brustle S. (2012) Archiving image sequences considering associated geographical and nongeographical attributes. In: Carapezza E. (ed) SPIE.

Carbone S. M., Luzzato R., Santa Maria A. (2006) Istituzioni di Diritto Internazionale. G. Giappichelli Turin (Italy), ISBN 978-8834819739.

Carro M. (2010) Data infrastructures and spatial models for biodiversity assessment and analysis: applications to vertebrate communities. Analisi, Protezione e Gestione delle Biodiversità. Università degli Studi dell'Insubria.

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

Cau P., Meloni G. C., Manca S., Soru C., Muroni D. (2011) A Java Based Framework Optimized for Scientific Modeling and Analysis. In: International MultiConference of Engineers and Computer Scientists. Hong Kong, pp. 303-308.

Cavaleri L. (2007) The MEDATLAS wind and wave atlas in the Mediterranean Sea - a combined use of buoy, model and satellite data (abstract only). In: Queffeullou P., Quentel M. L. (eds) Globwave Project Workshop, Brest (France), 20-21 Sep 2007. IFREMER, Plouzane (France), pp. 14-15.

Chawla M., Bhaskar V. (2007) Exposing SOA Enabled C Apps as Web Service. XML Journal.

Cleland J. H. (1995) Corrosion risks in ship's ballast tanks and the IMO pathogen guidelines. Engineering Failure Analysis 2(1):79-84.

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

Cowell A., Gibbs W. (2007) Nazi U-Boat Imperils Norwegians Decades After the War. The New York Times.

Cressman R. J. (1999) The Official Chronology of the U.S. Navy in World War II. Naval Historical Center, ISBN 978-1557501493, pp. 400.

Crichton D. (1999) The Risk Triangle. In: Ingleton J. (ed) Natural disaster management. Tudor Rose Holdings, Limited, London (UK), ISBN 0953614018, pp. 102-103.

Cunningham Dobson N. (2009) The HMS Sussex Shipwreck Project (Site E-82): Preliminary Report. Odissey (ed).

Daniel P., Gilbert T., Hackett B., Hines A., Tiercelin C. (2008) Operational Metocean Products and Services In Support of Marine Pollution Emergency Response Operations. In: International Oil Spill Conference 2008. IOSC, Washington, DC (USA).

Dassonville L., Vauglin F., Jacobsson A., Luzet C. (2002) Quality Management, Data Quality and Users, Metadata for Geographical Information. In: Shi W., Fisher P. F., Goodchild M. F. (eds) Spatial Data Quality. Taylor & Francis, London (UK), ISBN 978-0415258357, pp. 202-215.

Davin J. J. J., Witte J. J. A. (1997) CLEVECO Underwater Oil Recovery: Removing a 50-Year Old Threat. In: International Oil Spill Conference. Fort Lauderdale, FL (USA).

de la Huz R., Lastra M., Junoy J., Castellanos C., Viéitez J. M. (2005) Biological impacts of oil pollution and cleaning in the intertidal zone of exposed sandy beaches: Preliminary study of the "Prestige" oil spill. Estuarine Coastal and Shelf Science 65:19-29.

Dean T. A., Jewett S. C. (2001) Habitat-specific recovery of shallow subtidal communities following the Exxon Valdez oil spill. Ecological Applications 11(5):1456-1471.

Delvigne G. A. L., Sweeney C. E. (1988) Natural dispersion of oil. Oil and Chemical Pollution 4(4):281-310.

Devillers R., Beard K. (2006) Communication and use of spatial data quality information in GIS. In: Fundamentals of Spatial Data Quality. ISTE, London (UK), ISBN 978-1905209569, pp. 237-253.

Devillers R., Bedard Y., Jeansoulin R., Moulin B. (2007) Towards Spatial Data Quality Information Analysis tools for Experts Assessing the Fitness for Use of Spatial Data. International Journal of Geographical Information Sciences 21(3):261-282.

Devillers R., Gervais M., Bedard Y., Jeansoulin R. (2002) Spatial Data Quality: from metadata to quality indicators and contextual end-users manual. In: OEEPE/ISPRS Joint Workshop on Spatial Data Quality Management. Instanbul (Turkey), pp. 11.

Devillers R., Stein A., Bedard Y., Chrisman N., Fisher P., Shi W. (2010) Thirty Years of Research on Spatial Data Quality: Achievements, Failures, and Opportunities. Transactions in GIS 14(4):387-400.

Dexter S. C. (1980) Global Variability of Natural Sea Water. Mater Performance (September 1980):16-28.

Donati F. (2012) Il ritrovamento della Corazzata Roma. Il Corriere di Como. September 23, 2012, Faenza.

Dorsey W. (2000) Historic Salvors, Marine Archeologists and The UNESCO Draft Convention on Underwater Cultural Heritage. In: Houston Marine Insurance Seminar.

Dorst L., Howlett C. (2012) Safe Navigation with Uncertain Hydrographic Data. Hydro International (5):18-21.

Dromgoole S. (2005) Legal protection of the underwater cultural heritage: lessons from the Titanic. Amicus Curiae (61):18.

Duckham M., McCreadie J. (2002) Error-aware GIS Development. In: Shi W., Fisher P. F., Goodchild M. F. (eds) Spatial Data Quality. Taylor & Francis, London (UK), ISBN 978-0415258357, pp. 63-75.

Dueri S., Dahllof I., Hjorth M., Marinov D., Zaldivar J. M. (2009) Modeling the combined effect of nutrients and pyrene on the plankton population: Validation using mesocosm experiment data and scenario analysis. Ecological Modelling 220:2060-2067.

EC (2009) Scientific Principles for Design of Marine Protected Areas in Australia: A Guidance Statement. The Ecology Centre, The University of Queensland, http://www.uq.edu.au/ecology/index.html?page=102441&pid=108450.

Elmgren R., Hansson S., Larsson U., Sundelin B., Boehm P. (1983) The "Tsesis" oil spill: acute and long-term impact on the benthos. Marine Biology 73:51-65.

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(11):2141-2144.

ESRI (1998) ESRI Shapefile Technical Description. An ESRI White Paper. Environmental Systems Research Institute, Inc.

Etkin D. S. (1999) Oil Spill Reference Guide. Cutter Information Corp., Arlington, MA (USA), ISBN B0006RKZ70, pp. 59.

Etkin D. S. (2002) Analysis of past marine oil spill rates and trends for future contingency planning. In: 25th Arctic & Marine Oilspill Programme (AMOP) Technical Seminar. Environment Canada, Ottawa, ON (Canada), pp. 227-252.

Fallah M. H., Stark R. M. (1976) Literature review: movement of spilled oil at sea. Journal Name: Mar. Technol. Soc. J.; (United States); Journal Volume: 10:1:3-18.

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(7):1092-1098.

Fay J. A. (1971) Physical processes in the spread of oil on a water surface. In: Institute A. P. (ed) Proceedings of the Joint Conference on Prevention and Control of Oil Spill. Washington (USA), pp. 463-467.

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

Fingas M., Fieldhouse B., Mullin J. (1996) Studies of water-in-oil emulsions: the role of asphaltenes and resins. In: Proceedings of the Nineteenth Arctic and Marine Oil Spill Program (AMOP) Technical Seminar. Environment Canada, Emergencies Science Division, Ottawa (Canada), pp. 73-88.

Fingas M. F., Charles J. (2001) The basics of oil spill cleanup. Lewis Publishers, Boca Raton, Fla., ISBN 1566705371, pp. 233.

Fisher P. F., Unwin D. (2005) Re-presenting GIS. John Wiley & Sons, Chichester, England ; Hoboken, NJ, ISBN 0470848472 pp. x, 286.

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

France (1999) Diplomatic Notes. In: Journal Officiel de la République Française, p. 3.261.

France (1999) Journal Officiel de la République Française. In, p. 3.260.

Franceschini E. (2002) Manovre militari Israele-USA, profanato il relitto dello Scirè. La Repubblica. 2 October 2002.

Frank A. U. (1998) Building a geospatial data framework - finding the best available data. In: Data Quality in Geographic Information: from error to uncertainty. Hermes, Paris (France), ISBN 978-2866016579, pp. 192.

Fubao Z., Hui C., Jihong G. (2010) Querying GML documents: An XQuery based approach. In: International Conference On Computer and Communication Technologies in Agriculture Engineering (CCTAE). pp. 393-396.

Furieri A. (2012) SpatiaLite Cookbook. http://www.gaia-gis.it/gaia-sins/Spatialite-Cookbook-PDF.zip.

Galt A. J., Payton L. D. (1999) Development of Quantitative Methods For Spill Response Planning: A Trajectory Analysis Planner. Spill Science & Technology Bulletin 5(1):17-28.

Garabello R. (2004) Sunken Warships in the Mediterranean. In: Scovazzi T. (ed) La protezione del patrimonio culturale sottomarino nel Mare Mediterraneo. Giuffre, Milan (Italy), pp. 192-195.

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. Ed. Springer, New York, ISBN 978-3540775843, pp. 244-256.

Gervais M. (2004) Pertinence d'un manuel d'instructions au sein d'une strategie de gestions du risque juridique decoulant de la forniture de donnees geographiques numeriques. PhD, Universite Laval, Quebec et Universite de Marne-La-Vallee, pp. 344.

Gilbert T. (1998) Maritime Response Operations- Requirements for Met/Ocean Data and Services. In: Conference on Meteorological and Oceanographic Services for Marine Pollution Emergency Operations (MARPOLSER 98). World Meteorological Organisation Report #960, pp. 67-89.

Gilbert T. (2001) Report of the Strategic Environmental Assessment, USS Mississinewa Oil Spill, Ulithi Lagoon, Yap State, Micronesia. South Pacific Regional Environment Programme.

Gilbert T., Nawadra S., Tafileichig A., Yinug L. (2003) Response to an Oil Spill from a Sunken WWII Oil Tanker in Yap State, Micronesia. In: International Oil Spill Conference. Vancouver (Canada).

Girin M. (2004) European experience in response to potentially polluting shipwrecks. Marine Technology Society Journal 38(3):21-25.

Glasby G. P. (1997) Disposal of chemical weapons in the Baltic Sea. The Science of the Total Environment 206:267-273.

Gonik A. A. (2003) Prevention of Repaired Steel Tanks from Sulphide Corrosion Cracking. Protection of Metals (40):297-299.

Goodchild M. F. (2008) Epilogue: Putting research into practice. In: Stein A., Shi W., Bijker W. (eds) Quality Aspects in Spatial Data Mining. CRC Press, Boca Raton, ISBN 978-1420069266, pp. 345-356.

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. Springer, Berlin (Germany), ISBN 978-3642019753, pp. 49-59.

Goodchild M. F., Gopal S. (1989) The Accuracy of spatial databases. Taylor & Francis, London ; New York, ISBN 0850668476, pp. xviii, 290 p.

Goodchild M. F., Zhou J. (2003) Finding geographic information: collection-level metadata. GeoInformatica 7(2):95-112.

Gravley D. (2001) Risk, Hazard, and Disaster. University of Canterbury in New Zealand, pp. 7.

Grennan D. (2010) What is the current state of the art for assessment, salvage and response technologies? In: International Corrosion Workshop. Newport News, VA (USA).

Griffin R. B., Cornwell R. L., Seitz W., Estes E. (1989) Localized corrosion under biofouling. Mater Performance 28(3):71-74.

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). IOS Press, Trento (Italy), ISBN 905-1993994 pp. 3-15.

Gubbay S. (2004) Marine Protected Areas in the context of marine spatial planning – discussing the links. WWF UK, http://www.wwf.org.uk/filelibrary/pdf/MPAs-marinespacialplanning.pdf.

Guerin U., Egger B., Penalva V. (2010) Underwater Cultural Heritage in Oceania. UNESCO, Paris (France), http://unesdoc.unesco.org/images/0018/001887/188770e.pdf.

Guttman A. (1984) R-trees: a dynamic index structure for spatial searching. In: Proceedings of the 1984 ACM SIGMOD international conference on Management of data. ACM, Boston, Massachusetts, pp. 47-57.

Halls I. W. (2012) Editorial. International Hydrographic Review (8):5.

Halpin T. A., Morgan T., ScienceDirect (Online service) (2008) Information modeling and relational databases. Morgan Kaufmann series in data management systems. Morgan Kaufmann Publishers, Burlington, MA, http://www.sciencedirect.com/science/book/9780123735683.

Hamblin J. D. (2008) Poison in the well : radioactive waste in the oceans at the dawn of the nuclear age. Rutgers University Press, New Brunswick, N.J., ISBN 978-0813542201, pp. x, 311.

Harper S. (2010) Conveying Data Quality in ENCs to the Mariner – Work Past, Present and Future of the IHO Data Quality Working Group. In: Hydro2010. Rostock-Warnemunde (Germany).

Harris J. H. (2001) The protection of Sunken Warships as Gravesites at Sea. Ocean and Coastal Law Journal (7):75-129.

Hassellöv I.-M. (2007) Pre-Study of Ship Wreck Assessment and Remediation. Alliance for Global Sustainability, Goteborg (Sweden).

Hawkins J. (2003) Hydrographic data key element in SS Jacob Luckenbach oil removal. In: US Hydro 2003. Biloxi, MS (USA).

Hazen T. C., Dubinsky E. A., DeSantis T. Z., Andersen G. L., Piceno Y. M., Singh N., Jansson J. K., Probst A., Borglin S. E., Fortney J. L. (2010) Deep-Sea Oil Plume Enriches Indigenous Oil-Degrading Bacteria. Science 330:204-208.

Heldtberg M., Macleod I. D., Richards V. L. (2004) Corrosion and cathodic protection of iron in seawater: a case study of the James Matthews (1841). In: Metal 2004. National Museum of Australia, Camberra (Australia).

Helton D., Symons L. (2012) NOAA's Remediation of Underwater Legacy Environmental Threats (RULET) Database & Wreck Oil Removal Program (WORP). In: Interspill 2012. London, UK.

Helton D., Symons L., Wagner J. (2012) Potentially Polluting Wrecks in U.S. waters. In: Interspill 2012. London (UK).

Heuvelink G. B. M. (1998) Error Propagation in Environmental Modelling with GIS. Taylor & Francis, London (UK), ISBN 978-0748407446, pp. 144.

Hill F. S., Kelley S. M. (2007) Computer graphics : using OpenGL. Pearson Prentice Hall, Upper Saddle River, NJ, ISBN 978-0131496705, pp. xxii, 778.

Hunter D., Rafter J., Fawcett J., van der Vlist E., Ayers D., Duckett J., Watt A., McKinnon L. (2007) Beginning XML - 4th Edition. Wiley Publishing, Inc., Indianapolis (USA), ISBN 978-0470114872, pp. 1080.

Hunter G. J., Goodchild M. F. (1997) Modeling the uncertainty in slope and aspect estimates derived from spatial databases. Geographical Analysis 29(1):35-49.

Hunter G. J., Reinke K. J. (2000) Adapting Spatial Databases to Reduce Information Misuse Through Illogical Operations. In: Spatial Accuracy Assessment, Land Information Uncertainty in Natural Resources Conference. Amsterdam (The Netherlands), pp. 313-319.

Hunter J. (2002) Heritage or Hazard ? The Oil Tanker Montebello and Its Potentially Dangerous Cargo. In: Ruppe C. V., Barstad J. F. (eds) International Handbook of Underwater Archaeology. Kluwer Academic & Plenum Publishers, New York (USA), ISBN 978-0306463457, pp. 727-736.

ICRAMM-CEDRE (2007) DEvelopment of European guidelines for Potentially Polluting shipwrecks - Proposal. D.G. Environment, Civil Protection Unit, pp. 57.

ICRAMM-CEDRE (2007) Outcomes of the Potentially Polluting Shipwrecks Workshop: technical, legal and financial perspectives. Rome (Italy), pp. 57.

IHO (2010) IHO S-100 - Universal Hydrographic Data Model. International Hydrographic Bureau.

IHO (2012) IHO Geospatial Information Register. International Hydrographic Bureau, Monaco, http://registry.iho.int/s100_gi_registry/home.php.

IMO (1978) MARPOL 73/78 : articles, protocols, annexes, unified interpretations of the International Convention for the Prevention of Pollution from Ships, 1973, as modified by the protocol of 1978 relating thereto. International Maritime Organization.

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

ISDR (2004) Terminology: Basic terms of disaster risk reduction. In: Living with Risk: A global review of disaster reduction initiatives. United Nations, New York (USA) and Geneva (Swiss), ISBN 9211010640, pp. 429.

ISO (1999) ISO/IEC Guide 51 - Safety aspects - Guidelines for their inclusion in standards.

ISO (2002) ISO 19113 - Geographic Information - Quality Principles.

ISO (2003) ISO 19114 - Geographic information — Quality evaluation procedures.

ISO (2003) ISO 19115 - Geographic Information - Metadata.

ISO (2004) ISO 3864-2 - Safety colours and safety signs - Part 2: Design principles for product safety labels

ISO (2005) ISO 19109 - Geographic information - Rules for application schema.

ISO (2006) ISO/IEC 19757-3:2006 Information technology - Document Schema Definition Languages (DSDL) - Part 3: Rule-based validation - Schematron.

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

Jacobson M. (2003) The international compensation regime 25 years on. In: The IOPC Funds' 25 years of compensating victims of oil pollution incidents. IOPC Fund, Kent (UK), ISBN 978-1402083679, pp. 13-24.

James S. J. (2003) The international compensation regime: the insurer's view. In: The IOPC Funds' 25 years of compensating victims of oil pollution incidents. IOPC Fund, Kent (UK), ISBN 978-1402083679, pp. 89-92.

JANAC (1947) Japanese Naval and Merchat Shipping Losses During World War II by All Causes. U.S. Navy Department / Joint Army-Navy Assessment Committee, Washington (USA), pp. 123.

Janowicz K., Schade S., Bröring A., Keßler C., Maué P., Stasch C. (2010) Semantic Enablement for Spatial Data Infrastructures. Transactions in GIS 14(2):111-129.

JNCC (2010) Different Types of Marine Protected Areas. Joint Nature Conservation Committee, http://jncc.defra.gov.uk/pdf/MPAsInfoDoc_v2_2.pdf.

Johnson E. (2008) Japan Asked to Avert Disaster From Pacific World War II Wrecks. Bloomberg.

Johnson R. (2004) SS Richard Montgomery Matter. The Sunday Times Magazine. http://www.ssrichardmontgomery.com/images/suntimes.htm.

Jones R. K. (1997) A simplified psuedo-component oil evaporation model. In: Proceedings of the Twentieth Arctic and Marine Oilspill Program (AMOP). Environment Canada, Emergencies Science Division, Ottawa (Canada), pp. 43-61.

Kaffka A. V. (1996) Sea-Dumped Chemical Weapons: Aspects, Problems and Solutions. In: NATO Advanced Research Workshop on 'Sea-Dumped Chemical Munitions'. Kalingrad (Russia).

Kassomenos P. A. (2004) Risk analysis for environmental hazards: the case of oil spills, in Crete. Global Nest: the International Journal 6(1):39-51.

Kent W. (1983) A Simple Guide to Five Normal Forms in Relational Database Theory. In: Communications of the ACM. pp. 120-125.

Kerzner H. (2009) Project management : a systems approach to planning, scheduling, and controlling. John Wiley & Sons, Hoboken, N.J., ISBN 9780470278703, pp. xxiv, 1094.

Kolpackov B. (2007) An Introduction to XML Data Binding in C++. The C++ Source. http://www.artima.com/cppsource/xml_data_binding.html.

Kreibich J. A. (2010) Using SQLite. O'Reilly, Sebastopol, Calif., ISBN 9781449394592, pp. xx, 503.

Lake R. (2004) Geography mark-up language (GML). Foundation for the GeoWeb. John Wiley & Sons, Chichester (UK), ISBN 978-0470871546, pp. 406.

Lans R. F. v. d., Cools D. (2006) Introduction to SQL mastering the relational database language. Addison Wesley Professional, Upper Saddle River, N.J., http://proquest.safaribooksonline.com/?uiCode=uariz&xmlId=0321305965.

Lanter D. P. (1994) A lineage metadata approach to removing redundancy and propagating updates in a GIS database. Cartography and Geographic Information Systems 21(2):91-98.

LaQue F. L. (1975) Marine corrosion : causes and prevention. Wiley, New York, ISBN 0471517453, pp. xi, 332 p.

Laurin F. (1991) Scandinavia's underwater time bomb. The Bulletin of the Atomic Scientists 47(2):10-15.

Law R. J., Hellou J. (1999) Contamination of Fish and Shellfish Following Oil Spill Incidents. Environmental Geosciences 6(2):90-98.

Lazzari L., Pedeferri P. (1981) Protezione catodica. CLUP, Milan (Italy), ISBN 978-8873980285, pp. 496.

Lee D., Chu W. W. (2000) Comparative Analysis of Six XML Schema Languages. ACM SIGMOD Record 29(3).

Lee D., Chu W. W. (2001) CPI: Constraints-Preserving Inlining algorithm for mapping XML DTD to relational schema. Data &Knowledge Engineering 39(1):3-25.

Leewis R. (1991) Environmental Impact of Shipwrecks in the North Sea II. Negative Aspects: Hazardous Substances in Shipwrecks. Water Science and Technology 24(10):299-300.

Lehr W., Jones R., Evans M., Simecek-Beatty D., Overstreet R. (2002) Revisions of the ADIOS oil spill model. Environmental Modelling & Software 17(2):189-197.

Lehr W. J., Overstreet R. (1998) Smoke plume screening tool for in-situ burning. In: The 1998 1st International Conference on Oil and Hydrocarbon Spills, Modelling, Analysis and Control, Oil Spill. Southampton (UK), pp. 121-129.

Lentz S. (2004) The Response of Buoyant Coastal Plumes to Upwelling-Favorable Winds. J Phys Oceanogr 34(11):2458-2469.

Levesque M. A., Bedard Y., Gervais M., Devillers R. (2007) Towards managing the risks of data misuse for spatial datacubes. In: Stein A., Shi W., Bijker W. (eds) Fifth International Symposium on Spatial Data Quality (ISSDQ). International Institute for Geo-Information Science and Earth Observation, Enschede (The Netherlands).

Liguria (2008) Caratterizzatione e Progettatione Preliminare e Definitiva della Bonifica del Relitto della VLCC Haven. Liguria Region, pp. 92.

Lindström P. (2006) Vrak i Skagerrak. In: Forum Skagerrak II. Länsstyrelsen i Västra Götalands Län, pp. 6.

Ling L., Qing W., Huiging W. (2010) GIS Data Management Based on Spatialite Database. Geomatics World (4):71-75.

Liu C., Vincent M., Liu J. (2006) Constraint Preserving Transformation from Relational Schema to XML Schema. World Wide Web 9(1):93-110.

Lu j. (2003) Marine oil spill detection, statistics and mapping with ERS SAR imagery in south-east Asia. International Journal of Remote Sensing 24(15):3013-3032.

Lv T., Yan P. (2006) Mapping DTDs to relational schemas with semantic constraints. Information and Software Technology 48(4):245-252.

Mackay D., Buist I., Mascarenhas R., Patterson S. (1980) Oil Spill Processes and Models. Environment Canada, Ottawa (Canada).

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. Kluwer Academic & Plenum Publishers, New York (USA), ISBN 978-0306463457, pp. 697-714.

Macleod I. D. (2005) In-situ monitoring of the iron shipwreck City of Launceston (1865). Department of Materials Conservation, Western Australian Museum, Fremantle (Australia).

Macleod I. D. (2006) In-situ Corrosion Studies on Wrecked Aircraft of the Imperial Japanese Navy in Chuuk Lagoon, Federated States of Micronesia. The International Journal of Nautical Archaeology 35(1):128-136.

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

Mainetti V. (2002) Considerazioni in tema di esercizio della sovranità archeologica da parte dello Stato costiero. In: Camarda G., Scovazzi T. (eds) The Protection of the Underwater Cultural Heritage. Legal Aspects. Giuffre, Milan (Italy), ISBN 978-8814095658 pp. 239.

MAMMOET (2008) Norwegian government recommends salvage of the U-864. Mammoet Salvage B.V., Schiedam (The Netherlands).

Mamoulis N. (2011) Spatial Data Management. Morgan & Claypool Publishers, Plymouth (UK), ISBN 978-1608458325, pp. 150.

Manca S., Cau P., Soru C., Vargiu A., Muroni D., Peneva E. (2011) A WEB based framework to face issue of marine water vulnerability. In: Computer Research and Development (ICCRD), 2011 3rd International Conference on. pp. 328-332.

Manso M., Wachowicz M. (2009) GIS design: A review of current issues in interoperability. Geography Compass 3:1105-1124.

Masetti G., Calder B., Alexander L. (2012) Potentially Polluting Marine Sites GeoDB: An S-100 Geospatial Database as an Effective Contribution to the Protection of the Marine Environment. International Hydrographic Review (8):27-40.

Masetti G., Orsini F. (2011) I relitti delle navi da guerra. Status e tutela giuridica. Informazioni della Difesa (5):25-31.

Masetti G., Sacile R., Trucco A. (2010) Caratterizzazione remota del fondale marino tramite analisi e mosaicatura del backscatter. In: Atti della 14a Conferenza Nazionale ASITA (9 – 12 novembre 2010). Brescia (Italy).

Masetti G., Sacile R., Trucco A. (2011) Remote characterization of seafloor adjacent to shipwrecks using mosaicking and analysis of backscatter response. Italian Journal of Remote Sensing 43(2):77-92.

Matera M. (2002) Due accordi relative a navi di Stato britanniche. In: Scovazzi T. (ed) The Protection of the Underwater Cultural Heritage. Legal Aspects. Giuffre, Milan (Italy), ISBN 978-8814095658 pp. 63.

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

McClure S. (1997) Object Database vs. Object-Relational Databases. IDC Bulletin (14821E):21.

McGranaghan M. (1993) A cartographic View of Spatial Data Quality. Cartographica 30:8-19.

McGrath G. G. (2011) Utilizing the OSLTF for Response to Spills from Sunken Shipwrecks. In: 2011 International Oil Spill Conference. Portland, OR (USA).

McLaughlin B. (2002) Java and XML Data Binding. O'Reilly & Associates, Sebastopol, CA (USA), ISBN.

Melchers R. E. (1999) Corrosion uncertainty modelling for steel structures. Journal of Constructional Steel Research (52):3-19.

Melchers R. E. (2002) Effect of Temperature on the Marine Immersion Corrosion of Carbon Steels. Corrosion 58(9):768-782.

Melchers R. E. (2005) The effect of corrosion on the structural reliability of steel offshore structures. Corrosion Science 47:2391-2410.

Melchers R. E., Ahammed M. (1994) Non linear modelling of corrosion of steel in marine environments. Surveying D. o. C. E. a. (ed) The University of Newcastle, Newcastle (UK).

Mervin F. (2000) Basics of oil spill cleanup. CRC Press LLC, Boca Raton (USA), ISBN 978-1566705370, pp. 256.

MHU (2009) Port Phillip Bay Fieldwork. Maritime Heritage Unit - Heritage Victoria.

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

Michel K., Winslow T. (2000) Cargo Ship Bunker Tankers: Designing to Mitigate Oil Spills. In: SNAME Joint California Sections Meeting.

Missiaen T., Henriet J. P. (2002) Chemical munition dump sites in coastal environments: a border-transgressing problem. Missiaen T., Henriet J. P. (eds) Chemical munition dump sites in coastal environments. Federal Office for Scientific, Technical and Cultural Affairs, Brussels (Belgium), pp. 1-11.

MLAUS (2009) Wreck Oil Removal Program (WORP). 2009 Demonstration Project Overview. http://www.mlaus.org/content/ItemContent/10431Events.pdf.

Monfils R. (2005) The global risk of marine pollution from WWII shipwrecks: Examples from the seven seas. In: 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.

Morton A., Airoldi M., Phillips L. D. (2009) Nuclear Risk Management on Stage: A Decision Analysis Perspective on the UK's Committee on Radioactive Waste Management. Risk Analysis 29(5):764-779.

Navy U. (1946) Japanese Fuel and Lubrificants. Article 1 Fuel and lubrificant technology. U.S. Naval Technical Mission to Japan, pp. 99.

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

Nawadra S., Gilbert T. (2002) Risk of Marine Spills in the Pacific Islands Region and its Evolving Response Arrangements. In: SPILLCON. Sidney (Australia).

Nelson R. K., Kile B. M., Plata D. L., Sylva S. P., Xu L., Reddy C. M., Gaines R. B., Frysinger G. S., Reichenbach S. E. (2006) Tracking theWeathering of an Oil Spill with Comprehensive Two-Dimensional Gas Chromatography. Environmental Forensics 7:33-44.

NOAA (2012) General NOAA Operational Modeling Environment (GNOME) Technical Documentation. http://response.restoration.noaa.gov/gnome_manual.

NOAA (2012) Office of Response and Restoration. http://response.restoration.noaa.gov/oil-and-chemical-spills.

NRC (2003) Oil in the Sea III: Inputs, Fates, and Effects. National Academies Press, ISBN 978-0309084383, pp. 280.

O'Brien M. (2010) SEA DIAMOND 3 Years On... Dealing with Continual Leakage from Sunken Wrecks. In: First Adriatic Oil Spill Conference. ITOPF, Opatija (Croatia).

OCIMF (1997) Factors influencing accelerated corrosion of cargo oil tanks. Oil Companies International Marine Forum.

OGC (2003) OGC Reference Model. Perceivel M. (ed) Open Geospatial Consortium Inc., pp. 180.

Okrent D. (1980) Comment on Societal Risk. Science 208(4442):372-375.

Olsvik P. A., Brattas M., Lie K. K., Goksoyr A. (2011) Transcriptional responses in juvenile Atlantic cod (Gadus morhua) after exposure to mercury-contaminated sediments obtained near the wreck of the German WW2 submarine U-864, and from Bergen Harbor, Western Norway. Chemosphere 83:552-563.

Olsvik P. A., Brattås M., Lie K. K., Goksøyr A. (2011) Transcriptional responses in juvenile Atlantic cod (Gadus morhua) after exposure to mercury-contaminated sediments obtained near the wreck of the German WW2 submarine U-864, and from Bergen Harbor, Western Norway. Chemosphere 83(4):552-563.

Orth D. (2011) Administering America's Offshore Oil Fields: How Fewer, Performance-Based Regulations Can Produce Better Results. Journal of Environmental Law & Litigation 26(2):509-534.

Overfield M. L. (2005) Corrosion on Deep Gulf Shipwrecks of World War II. In: International Oil Spill Conference. Miami Beach (USA), p. a377.

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

Overfield M. L., Symons L. (2009) The use of the RUST database to inventory, monitor, and assess risk from Undersea Threats. Marine Technology Society Journal 43(4):33-40.

Paik J. K., Thayamballi A. K., Park Y. I., Hwang J. S. (2004) A time-dependent corrosion wastage model for seawater ballast tank structures of ships. Corrosion Science (46):471-486.

PAM (2010) Dumping of toxic and radioactive waste and Human Trafficking in the Mediterranean. Parliamentary Assembly of the Mediterranean, Rabat (Morocco), pp. 4.

Papadimitrakis J., Psaltaki M., Christolis M., Markatos N. C. (2006) Simulating the fate of an oil spill near coastal zones: the case of a spill (from a power plant) at the Greek Island of Lesvos. Environmental Modelling & Software 21(2):170-177.

Parthiot F., de Nanteuil E., Merlin F., Zerr B., Guedes Y., Lurton X., Augustin J. M., Cervenka P., Marchal J., Sessarego J. P., Hansen R. K. (2004) Sonar detection and monitoring of sunken heavy fuel oil on the seafloor. In: Interspill 2004. Trondheim (Norway).

Peduzzi P., Dao H., Herold C. (2002) Global Risk and Vulnerability Index Trends per Yeat (GRAVITY), Phase II: Development, analysis and results. UNDP/BCPR, Geneva (Switzerland), pp. 62.

Petersen J., Michel J., Zengel S., White M., Lord C., Plank C. (2002) Environmental Sensitivity Index Guidelines. Seattle, Washington (USA), pp. 175.

Pfeifer C., Brzozowski E., Markarian R., Redman R. (2008) Quantifying percent cover of submerged oil using underwater video imagery. In: International Oil Spill Conference 2008. Savannah (USA).

Phull B. S., Pikul S. J., M. K. R. (1997) Seawater corrosivity around the world: result from five years of testing. In: Kain R. M. Y., W. T. (ed) Corrosion Testing in Natural Waters. ASTM, Philadephia (USA), ISBN 978-0803120259, pp. 34-73.

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, pp. 23.

Plunkett G. (2003) Sea Dumping in Australia: Historical and Contemporary Aspects. Australian government Department of Defence, Department of Environmental and Heritage, ISBN 0-642295883, pp. 164.

Poinssot C., Gin S. (2012) Long-term Behavior Science: The cornerstone approach for reliably assessing the long-term performance of nuclear waste. J Nucl Mater 420(1–3):182-192.

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, ISBN 978-3540283126, pp. 477-477.

Qiu J., Hunter G. J. (2002) A GIS with the Capacity for Managing Data Quality Information. In: Shi W., Fisher P. F., Goodchild M. F. (eds) Spatial Data Quality. Taylor&Francis, London (UK), ISBN 978-0415258357, pp. 230-250.

Rawson C., Crake K., Brown A. J. (1998) Assessing the Environmental Performance of Tankers in Accidental Grounding and Collision. SNAME Transactions (106):41-58.

Redmond M. C., Valentine D. L. (2011) Natural gas and temperature structured a microbial community response to the Deepwater Horizon oil spill. In: P Natl Acad Sci USA. PNAS, pp. 6.

Reed M., Johansen Ø., Brandvik P. J., Daling P., Lewis A., Fiocco R., Mackay D., Prentki R. (1999) Oil Spill Modeling towards the Close of the 20th Century: Overview of the State of the Art. Spill Science & Technology Bulletin 5(1):3-16.

REMPEC (2004) Regional Information System. Part C: Databanks and Information Services. Section 2: List of Alerts and Accidents in the Mediterranean. IMO/UNEP (ed), pp. 131.

Reynolds K. (2005) Summary on the results of the Potentially Polluting Shipwrecks Workshop, Southampton, June 2005. In: Potentially Polluting Shipwrecks Workshop: Technical, Legal and Financial perspectives. ICRAM-CEDRE, Rome (Italy).

Ritchie W., Bedborough D., Davies J., Dickson D., Hall M., Hepworth R., Kingston P., Mailes J., Monaghan P., O'Sullivan M., Tulloch B., Uscher M., Kingham L. (1994) Wreck of the tanker Braer: The environmental impact of the oil spill. Spill Science and Technology Bulletin of the American Meteorological Society 1(2):101-107.

Roberge P. (1999) Handbook of Corrosion Engineering. McGraw-Hill Companies,Incorporated, ISBN 978-0070765160, pp. 1072.

Roberts C. M., McClean C. J., Veron J. E. N., Hawkins J. P., Allen G. R., McAllister D. E., Mittermeier C. G., Schueler F. W., Spalding M., Wells F., Vynne C., Werner T. B. (2002) Marine Biodiversity Hotspots and Conservation Priorities for Tropical Reef's. Science 295(5558):1280-1284.

Ruane M. E. (2006) At Pearl Harbor, a cautious wait. The Washington Post. 8 December 2006, The Seattle Times Company.

Rumbaugh J., Jacobson I., Booch G. (1999) The unified modeling language reference manual. Addison-Wesley, Reading, Mass., ISBN 020130998X, pp. xvii, 550.

Sato R. (2010) Sea-Dumped Chemical Weapons: Japan. Global Green USA, Washington (USA), pp. 19.

Schneiderbauer S., Ehrlich D. (2004) Risk, hazard and people's vulnerability to natural hazards. European Commission Jointh Reseach Center, Luxemburg (Luxemburg), ISBN 978-3836481625, pp. 121.

Schumacher M. (1979) Seawater corrosion handbook. Noyes Data Corp., Park Ridge, N.J., ISBN 0815507364, pp. vii, 494.

Schwehr K. D., McGillivary P. A. (2007) Marine Ship Automatic Identification System (AIS) for Enhanced Coastal Security Capabilities: An Oil Spill Tracking Application. In: OCEANS 2007. Vancouver (Canada), p. 9.

Schwemmer R. (1996) California and the Second World War – The Attacks on the SS Montebello and the SS Idaho. In: The California State Military Museum.

Scovazzi T. (2004) La protezione del patrimonio culturale sottomarino nel Mare Mediterraneo. Giuffrè, Milan (Italy), ISBN 978-8814107726.

Sellis T., Roussopoulos N., Faloutsos C. (1987) The R+-Tree: A Dynamic Index for Multi-Dimensional Objects. In: 13th VLDB Conference. Brighton (UK), pp. 507-518.

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

Shapreau C. J. (2001) Extension of Express Abandonment Standard for Sovereign Shipwrecks in Sea Hunt, Inc. et al., Raises Troublesome Issues Regarding Protection of Underwater Cultural Property. International Journal of Cultural Property (10):276-314.

Shekhar S., Chawla S. (2003) Spatial databases : a tour. Prentice Hall, Upper Saddle River, N.J., ISBN 0130174807, pp. xxiii, 262.

Short J. W., Lindenberg M. R., Harris P. M., Maselko J. M., Pella J. J., Rice S. D. (2004) Estimate of Oil Persisting on the Beaches of Prince William Sound 12 Years after the Exxon Valdez Oil Spill. Environmental Science & Technology 38(1):19-25.

Simons M. (2003) Discarded War Munitions Leach Poisons Into the Baltic. New York Times. http://www.nytimes.com/2003/06/20/world/discarded-war-munitions-leach-poisons-into-the-baltic.html.

Smil V. (2008) Energy in nature and society : general energetics of complex systems. The MIT Press, Cambridge, Mass., ISBN 978-0262195652 pp. xi, 480.

Sofotassios D., Tzimas G., Tsakalidis A. (1997) A GIS-based Framework for oil-spill emergency management in the Mediterranean Sea. In: JEC-GI '97. Vienna (Austria), pp. 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. pp. 9-26.

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

Staniland H. (1999) South Africa. In: Dromgoole S. (ed) Legal Protection of the Underwater Cultural Heritage. National and International Perspectives. The Hague, London (UK) / Boston (USA), ISBN 978-9041197627, pp. 134.

Steiniger S., Hunter A. J. S. (2012) Free and Open Source GIS Software for Building a Spatial Data Infrastructure. Geospatial Free and Open Source Software in the 21st Century. In: Bocher E., Neteler M. (eds) Springer Berlin Heidelberg, ISBN 978-3642105951, pp. 247-261.

Stephenson P. E. (2005) Titanic Wreck Observations 2005. In: Special Meeting of Naval Architects and Marine Engineers. Landover (USA).

Stiver W., Mackay D. (1984) Evaporation rate of spills of hydrocarbons and petroleum mixtures. Environmental Science & Technology 18(11):834-840.

Stolzenbach K. D., Madsen O. S., Adams E. E., Pollack A. M., Copper C. K. (1977) Review and evaluation of basic techniques for predicting the behavior of surface oil slicks. ISBN, pp. x, 323.

Sui D. Z. (2004) Tobler'r First Law of Geography: a big idea for a small world ? Annals of the American Geographers 94(2):269-277.

Surhone L. M., Timpledon M. T., Marseken S. F. (2010) XML Data Binding. VDM Verlag Dr. Mueller e.K., ISBN 978-6131048012, pp. 72.

Svensson E. (2010) Potential Shipwreck Pollution in the Baltics Sea. Overview of Work in the Baltic Sea States. Swedish Maritime Administration and Lighthouse, Norrkoping (Sweden).

Symons L. (2010) Identification of Data Gaps and Development of Research Priorities. In: International Corrosion Workshop. Newport News (USA).

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, pp. 44.

Tanikawa H. (2003) A revolution in maritime law: a history of the original framework on oil spill liability and compensation. In: The IOPC Funds' 25 years of compensating victims of oil pollution incidents. IOPC Fund, Kent (UK), ISBN 978-1402083679, pp. 51-56.

Thorpe S. A. (1995) On the meandering and dispersion of a plume of floating particles caused by Langmuir circulation and a mean current. 25(4):685-690.

Tittensor D. P., Mora C., Jetz W., Lotze H. K., Ricard D., Berghe E. V., Worm B. (2010) Global patterns and predictors of marine biodiversity across taxa. Nature 466(August):1098-1103.

Tkalich P., Xiaobo C. (2001) Accurate Simulation of Oil Slicks. In: International Oil Spill Conference. Tampa, FL (USA), pp. 1133-1137.

Tobler W. R. (1970) A computer movie simulating urban growth in the Detroit region. Economic Geography 46(2):234-240.

Toja M. (2004) EQUA - L'immersione D.O.C.G. The Daily News of the Sea. http://www.marescoop.com/article90.html.

TSMAD (2012) GML Encoding in S-100. Paper for Consideration by TSMAD. Submitted by UK.

UNESCO (2001) Working Procedures 34. In: Convention on the Protection of the Underwater Cultural Heritage.

UNESCO (2001) Working Procedures 58. In: Convention on the Protection of the Underwater Cultural Heritage.

Unwin D. (1995) Geographical information systems and the problem of error and uncertainty. Progress in Human Geography 19(4):549-558.

Upton H. F. (2011) The Deepwater Horizon Oil Spill and the Gulf of Mexico Fishing Industry. CRS Report for Congress. Congressional Research Service, pp. 17.

USA (1953) Diplomatic Notes. In: United Nations Treaty Series, p. 432.

Valkovic V. (2009) New/Future Approaches to Explosive/Chemicals Detection. In: Application Of Accelerators In Research And Industry: Twentieth International Conference. American Institute of Physics, Fort Worth, TX (USA), pp. 591-594.

VLIZ (2012) Maritime Boundaries Geodatabase, version 6.1. http://www.vliz.be/vmdcdata/marbound.

W3C (2004) XML Schema Part 0: Primer Second Edition.

W3C (2008) Extensible Markup Language (XML) 1.0 (Fifth Edition).

W3C (2009) Namespaces in XML 1.0 (Third Edition).

W3C (2009) XML Schema Patterns for Databinding Interoperability Report. Downey P. (ed) W3C XML Schema Patterns for Databinding Interoperability working group.

W3C (2012) W3C XML Schema Definition Language (XSD) 1.1 Part 1: Structures.

Walford N. (2002) Geographical data : characteristics and sources. Wiley, West Sussex, England New York, ISBN 0471970859 (alk. paper), pp. xv, 274.

Walmsley P. (2002) Definitive Xml Schema. Prentice Hall Ptr, ISBN 9780130655677.

Wang W., Wang F., Qian Z., Zhang L. (2011) GML Data Management: Framework and Prototype. Lecture Notes in Computer Science 6637:101-111.

Wessel P., Smith W. H. F. (1996) A Global Self-consistent Hierarchical, High-resolution Shoreline database. J. Geophys. Res. 101:8741-8743.

Woodward T. (2008) Pacific World War II Wrecks Pose Risk of Toxic Leaks. National Geographic Magazine. http://news.nationalgeographic.com/news/2008/12/081210-pacific-shipwrecks-missions.html.

Yang X., Hai-Yang L. (2010) Research on Interconnection and Correspondence Between Relational Database and Embedded Database SQLite. In: Communications and Mobile Computing (CMC), 2010 International Conference on. pp. 279-283.

Yang X., Procopiuc C. M., Srivastava D. (2009) Summarizing relational databases. Proc. VLDB Endow. 2(1):634-645.

Yender R., Stanzel K., Lloyd A. (2008) Impacts and Response Challenges of the Tanker SOLAR 1 Oil Spill, Guimaras, Philippines: Observations of International Advisors. In: International Oil Spill Conference. Savannah (USA), pp. 77-82.

Yin J., Carswell J. D. (2012) MobiSpatial: open source for mobile spatial interaction. In: Proceedings of the 27th Annual ACM Symposium on Applied Computing. ACM, Trento, Italy, pp. 572-573.

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(1):110-141.

Zargar A. (2009) An operation-based approach to the communication of Spatial Data Quality in GIS. Memorial University of Newfoundland, St. John's, ISBN 978-0769535272, pp. 145.

Zelo I., Overfield M., Helton D. (2005) NOAA's abandoned vessel program and resources and under sea threats project - Partnerships and progress for abandoned vessel management. In: International Oil Spill Conference. Miami, FL (USA).

APPENDICES

APPENDIX A

CORROSION IN SEAWATER

The generic term of 'corrosion' describes the attacks taking place on the surface of many metals and alloys of technological interest. Furthermore, when these phenomena take place in water, the term 'in water' must be added to the term 'corrosion' to distinguish them from those which happen in a gaseous environment at high temperature. In particular, this Appendix illustrates the characteristics of marine corrosion.

The Chemical Process

All (natural) waters have an aggressive action function of:

- Physical-chemical properties of the water;
- Dissolved solids content; and
- Type of metal or alloy.

As a consequence, this rate can be used to calculate the hypothetical permanence in seawater for a wreck before a massive or continuous release of pollutants.

Corrosion in waters is the result of an electrochemical process that consists of two partial steps: an oxidation reaction followed by a reduction process.

The reaction [A.1] represents an oxidation where: "*Me*" is a surface atom of a generic metal, "*n e*" a generic number of electrons, "Me^{n+}" is the metal transformed into a cation.

$$Me - n\,e \bullet Me^{n+} \qquad (A.1)$$

For example, the oxidation reaction of iron can be represented by the following reaction:

$$Fe - 2\,e \bullet Fe^{++} \qquad (A.2)$$

The build-up of the electrons produced by this reaction slows-down and finally stops the corrosion process if a substance "*A*", using the electrons produced by reaction [A.1], is not transformed in another substance "*B*" with a reduction process [A.3]:

$$A + n\,e \bullet B \qquad (A.3)$$

Depending on the pH of water, substances which usually act as reducing agents are oxygen O_2 or hydrogen ions H^+, both normally present in natural waters. H^+ intervenes in acid environments (pH < 7), whereas the oxygen reduction is prevailing in neutral ones (pH • 7). The two possible types of reduction reactions are:

$$2\,H^+ + 2\,e \bullet 2\,H \bullet H_2 \qquad (A.4)$$

$$O_2 + 2\,H_2O + 4\,e \bullet 4\,OH^- \qquad (A.5)$$

The prevailing reduction reaction in a marine environment is that described by reaction [A.5] because the pH value of seawater ranges between 7.5 and 8.5. Thermodynamically speaking, the corrosive process can spontaneously take place only when the final free energy of the system "G_f" is lower than initial one "G_i":

$$G_f - G_i = \bullet G \quad \text{(A.6)}$$

By remembering the equivalence between chemical energy consumed and electrical work produced by a system:

$$\bullet G = -z F E \quad \text{(A.7)}$$

where "*F*" is Faraday's constant (96,500 Coulombs), "*z*" is the electric charge of the involved ionic species and "$E = E_{rid} - E_{ox}$" ("E_{rid}" is the potential of the reduction step and "E_{ox}" of the oxidation one). The corrosive attack is spontaneous only when $E > 0$.

For these reasons different metals show different tendencies to spontaneous oxidation. For example, gold shows a potential higher than the potentials both of oxygen and of hydrogen at every pH value. In this case, we observe always [A.8] and [A.9]:

$$E_{ox} > E_{rid} \quad \text{(A.8)}$$

$$\bullet G > 0 \quad \text{(A.9)}$$

In the E-pH diagram of gold, the broken lines "*a*" and "*b*" refer to potentials of oxygen and hydrogen, respectively, as a function of pH, whereas the thick line marks the boundary of the two zones of stability and corrodibility (Figure 57). So for gold, the corrosive process cannot take place spontaneously.

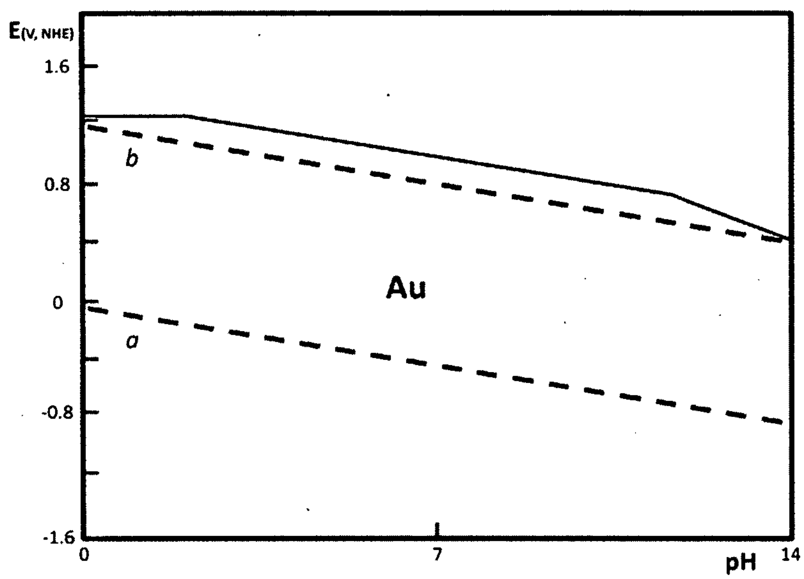


Figure 57 - E-pH diagram for gold adapted from (Alcaro et al., 2007).

The opposite happens in the case of iron and its alloys of technological interest – i.e. carbon- and low alloyed steels – which can spontaneously corrode in a wide range of pH and potential values (Figure 58).

It must be outlined that the "•*G*" values gives information only about the spontaneity of corrosive processes at given conditions of potential and pH, but do not allow prediction of the rate with which they will develop. Corrosion rate is influenced by the formation of corrosion products on the metal surface, which can considerably slow down the corrosion. This is the case of metals like aluminum and titanium which – even if thermodynamically highly corrodible – can form on the surface a layer of extremely protective oxides. This layer practically stops a subsequent advancement of the attack. This is the case of iron and of its alloys when immersed in sea water for a long time.

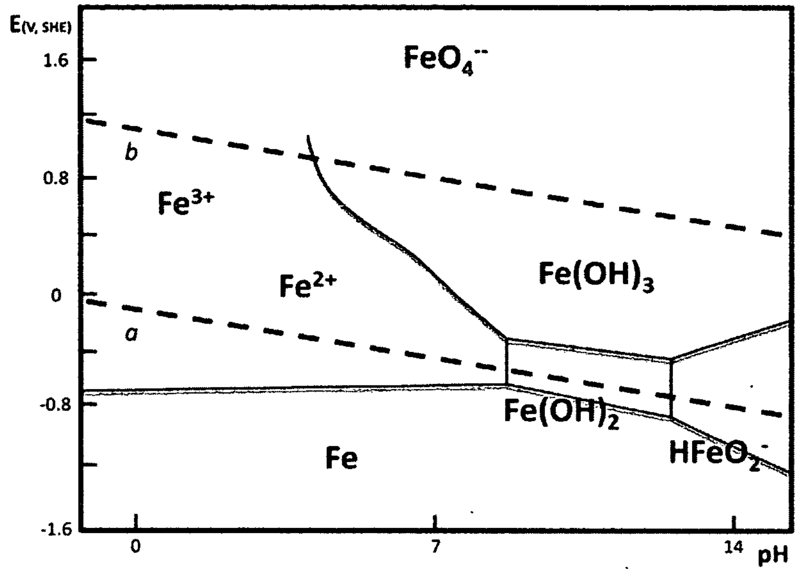


Figure 58 - E-pH diagram for iron adapted from (Alcaro et al., 2007; Roberge, 1999).

Corrosive Properties of Seawater

The type and concentration of dissolved ions in seawater are key factors in corrosion kinetics. These ions influence the protective properties of the corrosion product layers.

For example, Cl^- ions – also at low concentration, i.e., some tens of p.p.m. – radically decrease the resistance of many corrosion products layers present on the surface of metals and alloys. As a consequence, the sea represents one of the most aggressive natural environments because of its salinity.

The quantity of dissolved ions present in seawater may vary in a wide range, depending on geographic and climatic factors. For example, the closed seas – i.e., Red and Persian seas – can have a high value of 44 g/kg. In these seas there is at the same time high evaporation and low supply of fresh or oceanic water. In contrast, the salinity of the Baltic sea is very low – about 8 g/kg – due to the low evaporation and high fresh water supply. In general, the mean salinity of oceans varies in a range between 34 and 36 g/kg (Bianchi and Mazza, 2000; LaQue, 1975; Lazzari and Pedeferri, 1981). In seawater, chloride and sodium ions are the most abundant, followed by sulfate, magnesium and calcium ions. Salinity – in addition to its detrimental effect on the protective properties of corrosion products developed on the metal surface – increases the water conductivity, intensifying localized corrosion phenomena through the enlargement of the areas where the cathodic process takes place.

Temperature also influences the corrosion rate, but its effect on the process is not easy to describe. According to a series of experimental data obtained after five years of immersion tests carried out in different sites, a linear relationship exists between mean annual temperature of the sea and corrosion penetration rate (Melchers, 1999, 2002; Phull et al., 1997). In general, an increase of temperature involves an intensification of corrosion rate (Figure 59).

Another parameter which plays an important role in the corrosion process is the oxygen concentration in sea water. Oxygen is one of the chemical species which allow the corrosion to proceed by absorbing the electrons produced by the anodic process of dissolution of a metal.

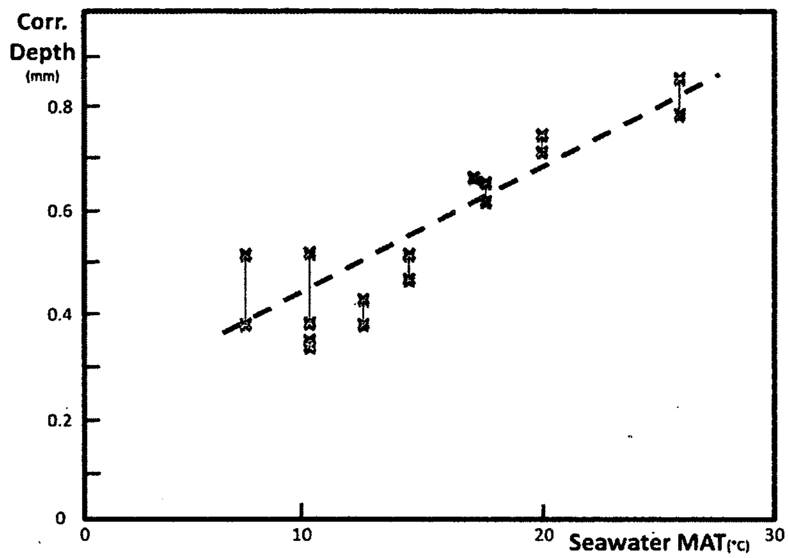


Figure 59 - Corrosion depth on steels as a function of mean annual temperature of seawater. Adapted from (Melchers, 1999).

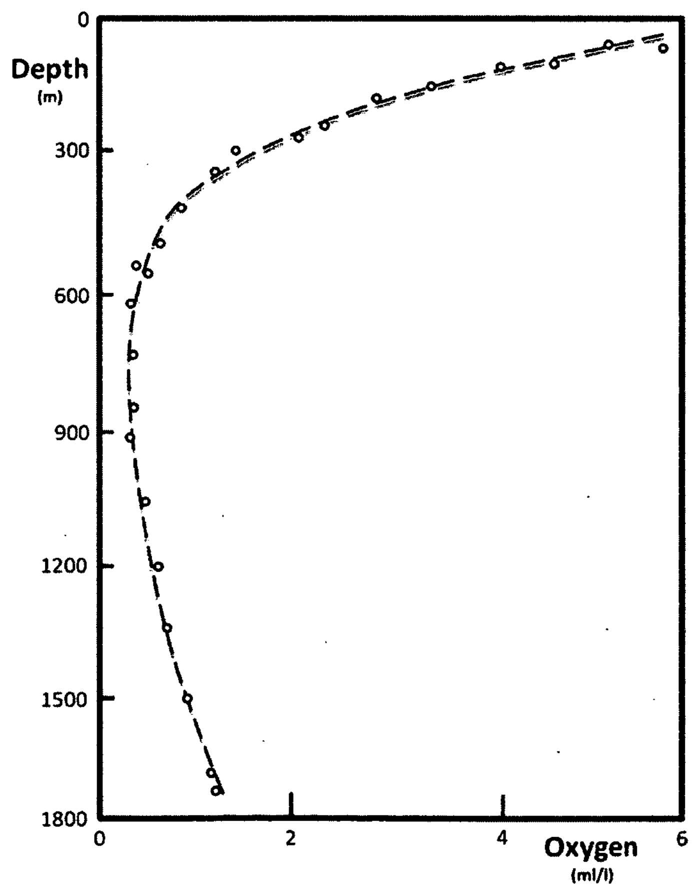


Figure 60 - Oxygen concentration as a function of seawater depth. Adapted from (Dexter, 1980; Roberge, 1999).

At the pH values of sea water (7.5 – 8.5), the cathodic partial process consists of the oxygen reduction as in reaction [A.5]. The oxygen concentration in sea water may vary between 0 and 10 p.p.m. and changes as a function of depth, reaching a maximum value near the sea surface, a minimum at a depth of about 700 meters and then increasing again at greater depths (Dexter, 1980).

Related to the surface value, it is essentially due to the uninterrupted exchange of oxygen with the atmosphere as well as from primary production (mainly phytoplankton activity). Deeper in the water column, the absence of algae and other sources of oxygen due to the sinking of dead organisms of the phyto- and zooplankton which, by decomposing, consume oxygen – reduces greatly the concentration. At levels deeper than 700 meters, in the oceanic context, the increase in oxygen concentration is due to the intervention of masses of Arctic and Antarctic water – rich in oxygen – which, due to their higher density, tend to move in proximity to the seabed. Starting from the above illustrated depth distribution of oxygen concentration, it is very interesting to observe, comparing Figure 60 and Figure 61, that the corrosion rate of carbon steel is strongly related to it (Schumacher, 1979).

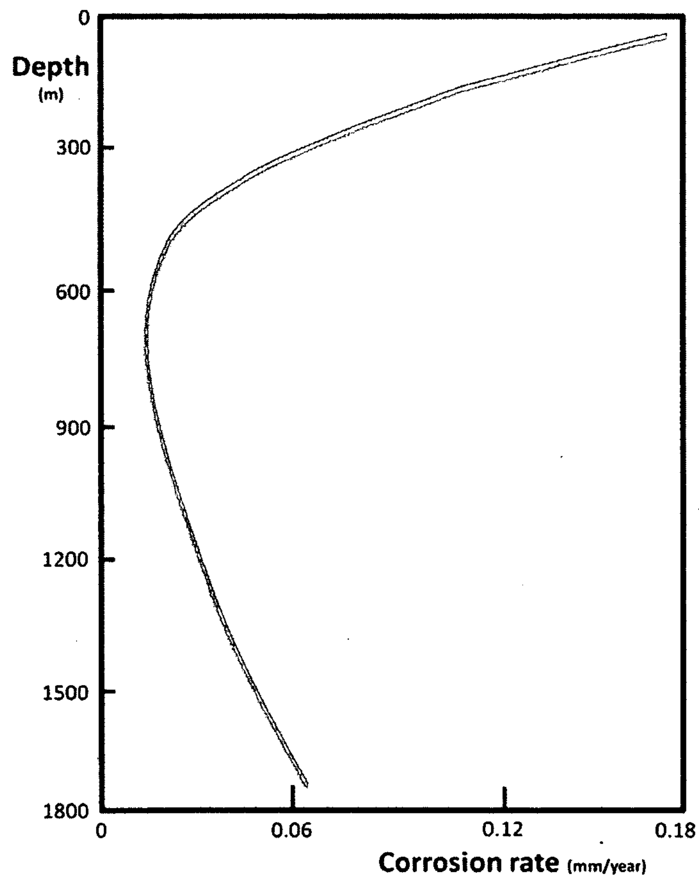


Figure 61 - Steel corrosion rate in function of sea depth. Adapted from (Schumacher, 1979).

The two curves in Figure 60 and Figure 61 have the same shape. As a matter of fact, the highest penetration rate is measured in proximity of the sea surface, where the oxygen concentration reaches its maximum value, whereas the minimum oxygen concentration at about 750 meters of depth corresponds to the minimum penetration rate of corrosion.

Particular attention has to be reserved for carbon steels because they are most widely utilized in ship building. In the following table the most studied carbon steels are listed together with their chemical composition.

Table 25 – Chemical composition of carbon steels utilized in ship building

AISI Nr.	Carbon	Manganese	Phosporus$_{MAX}$	Sulfur$_{MAX}$
1009	0.15 (max)	0.60 (max)	0.030	0.035
1010	0.08 – 0.13	0.30 – 0.60	0.040	0.050
1019	0.15 – 0.20	0.70 – 1.00	0.040	0.050
1020	0.17 – 0.24	0.25 – 0.60	0.040	0.050
1090	0.85 – 0.95	0.60 – 0.90	0.040	0.050

As regards their corrosion behavior in sea water – for short periods of immersion, i.e., up to 1 year – the corrosion rate of steels can be described by the following general equation:

$$p = k\,t^n \qquad \text{(A.10)}$$

where "*p*" is the corrosion penetration in meters, "*t*" is the immersion time in days, while "*n*" and "*K*" are two variables that vary as a function of the steel type between 0.5 (in surface) and 0.33 (in deep waters) and between 3.6 (in surface) and 3.9 (in deep waters), respectively. A graphical representation of [A.10] is provided in Figure 62.

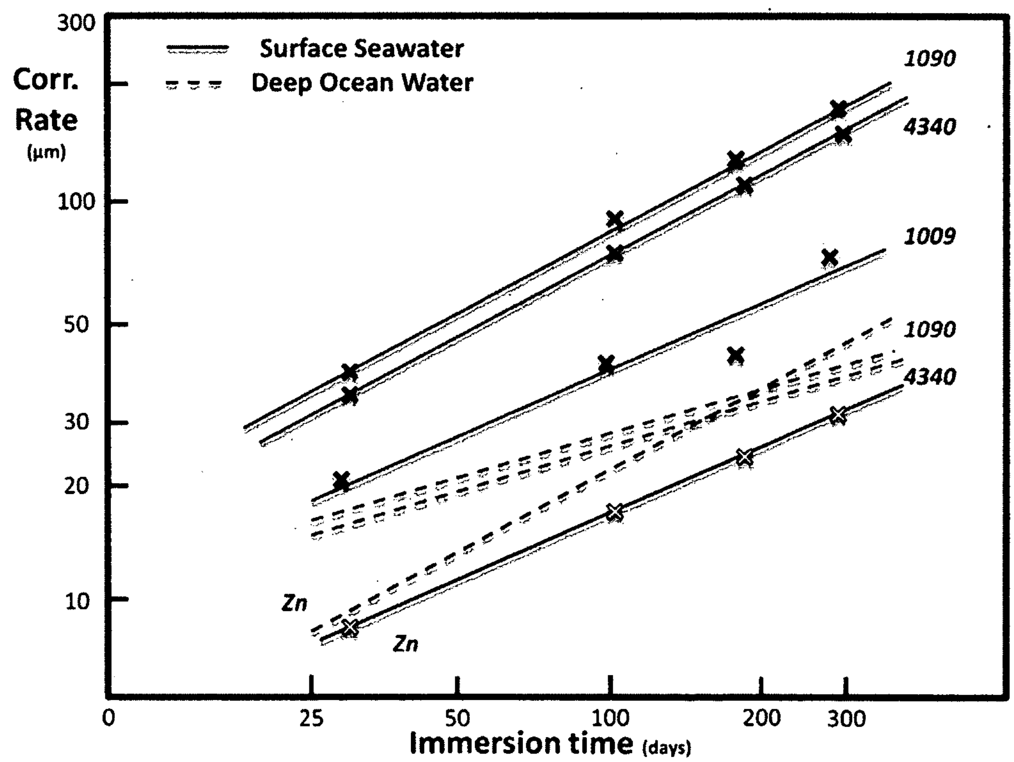


Figure 62 - Log-log diagram for corrosion penetration of steels as a function of immersion days. Adapted from (Alcaro et al., 2007).

The relationship [A.10] can be used in order to extrapolate a hypothesis on the extent of general corrosion penetration for longer immersion times (several tens of years). The curves A (AISI 1009 steel) and B (AISI 1090 steel) so obtained (Figure 63) indicate a tendency of corrosion rate on steels to decrease during the immersion time. This behavior is mainly due to the formation, on the steel surface, of a layer of corrosion products like iron oxides, hydroxides and carbonates which slow down the oxygen diffusion towards the cathodic areas present on the metal surface, and therefore inhibits the overall corrosion process.

In case that some particular conditions exist – e.g., absence of oxygen, continuous supply of SO_4^{2-} ions, availability of organic matter and presence of Fe^{2+} ions – the development and subsequent life of anaerobic organisms like sulfate reducing bacteria (SRB) is possible. With these bacteria the corrosion can continue at a constant rate by utilizing the reduction from SO_4^{2-} to S^{2-} ions as the cathodic reaction (Melchers, 1999; Melchers and Ahammed, 1994).

If these conditions are fulfilled, the corrosion rate of steel can be represented by curve C of Figure 63, based on experimental data obtained from long immersion tests (16 years).

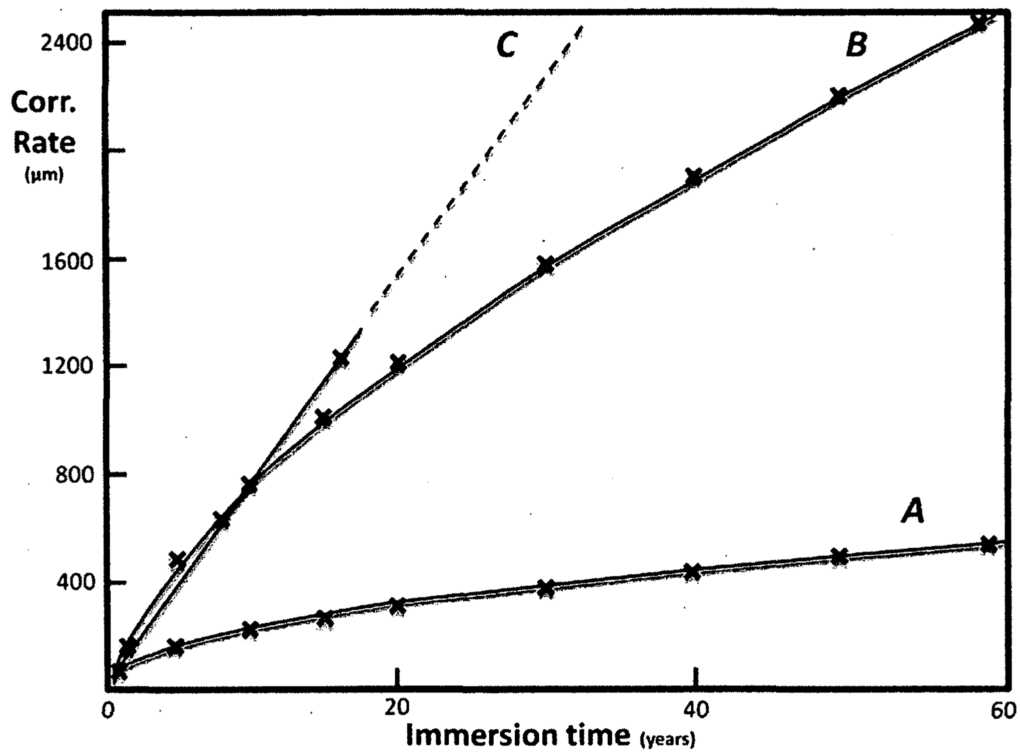


Figure 63 - Corrosion penetration depth as a function of immersion years: curve A is for AISI 1009 steel, curve B for AISI 1090 steel, while curve C represent the corrosion rate of steel obtained from long-term immersion tests. Adapted from (Alcaro et al., 2007).

Figure 64 shows, over a longer period, how an increasingly thick layer of corrosion products inhibits the corrosive process initially (variable based on several parameters). Then, after a period of an almost steady value, a new cathodic process can develop and the corrosion process restarts due to the reduction of sulfate- to sulfide ions (Melchers, 2005).

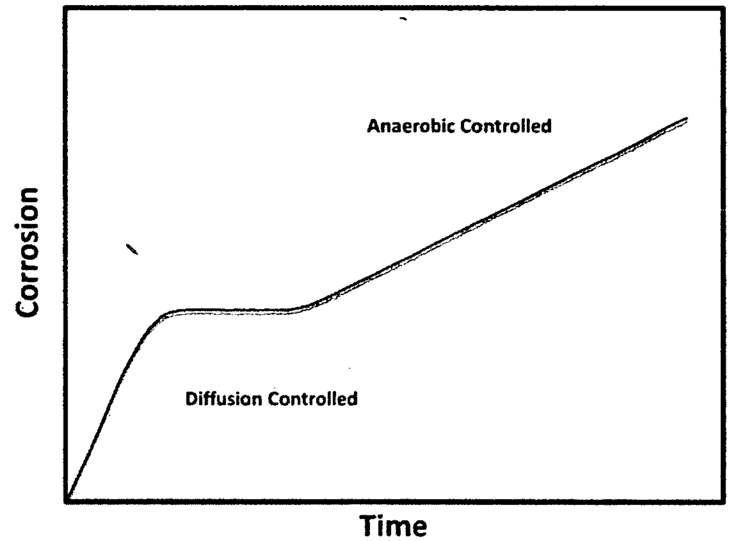


Figure 64 - Diffusion and anaerobic controlled corrosion rate on steels in seawater, adapted from (Melchers, 2005).

Considering that many of the wrecks of ships sunk during World War II have been in contact with seawater for over sixty years, the general corrosion penetration depth – according to curves A, B and C – could be equal to 550 μm, 2500 μm and 4600 μm respectively. The first two values are really too low to explain observed corrosion rates, corresponding to a mean penetration rate of 0.009 and 0.042 mm/year. The last one, with a mean penetration rate of 0.077 mm/year, approaches the 0.1 mm/year value, commonly indicated by naval engineers on the basis of observations made on steel structures working in the sea (Griffin et al., 1989; Paik et al., 2004). This value should be adopted as a rough approximation for corrosion rate where observed values for a given wreck or from features of similar material are not available in the area.

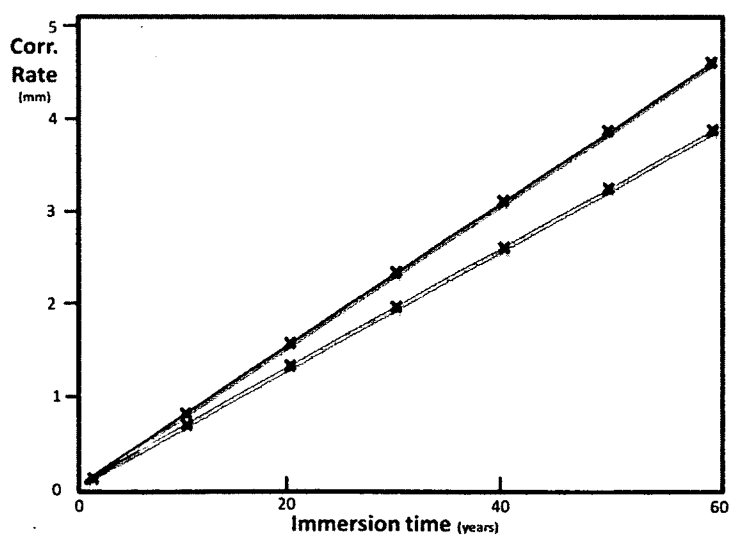


Figure 65 - Corrosion process in seas of middle and high latitudes, in blue, and in south seas, in red. Adapted from (Schumacher, 1979).

Furthermore, according to Figure 65, after an immersion period of sixty years, the penetration of general corrosion is 4.6 mm in tropical seas and to 3.9 mm in seas of the middle and high latitudes (Schumacher, 1979). These values correspond to mean penetration rates of 0.077 and 0.065 mm/year, respectively. Thus, even if relatively minor, the corrosion rate is also affected by the latitude and the related water conditions.

Corrosion penetration concerns not only the general and uniform corrosion that involves more the stability of the steel structure but also the leakage of fluids and/or solids which are contained in its interior (Figure 66). For example, the presence of "fouling" can frequently produce forms of localized corrosion. In these cases the penetration rate is much higher than that of general corrosion, at least during the initial stage of immersion.

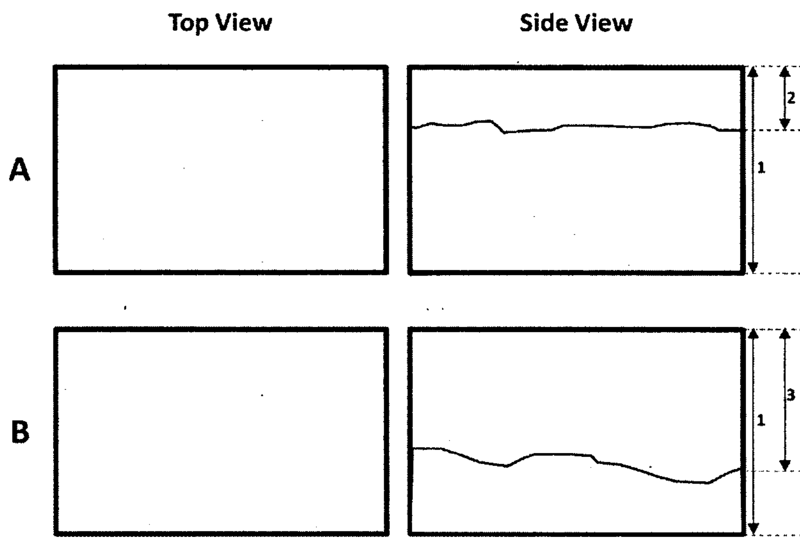


Figure 66 - General corrosion on a metal: A, uniform; B, not uniform; 1, original thickness; 2, mean corrosion depth; 3, maximum corrosion depth. Adapted from (Bianchi and Mazza, 2000).

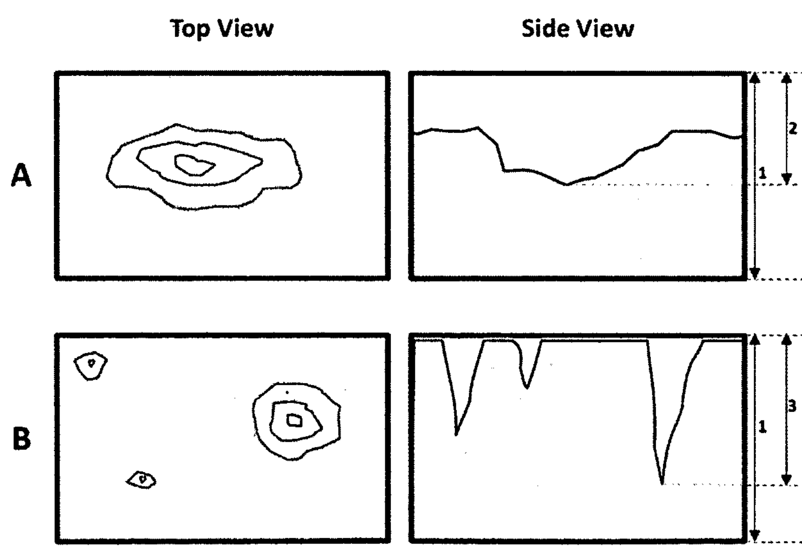


Figure 67 - Localized corrosion: A, crater-shaped; B, pitting; 1, original thickness of metal; 2 and 3: maximum depth. Adapted from (Bianchi and Mazza, 2000).

The localized forms of attack (Figure 67) can produce a continuous moderate leakage of fluids from the tanks much earlier than the collapse of the whole structure. However, a collapse event usually creates a sudden "eruption" towards the sea surface of all the oil contained in the tanks. It is difficult to foresee which of the two scenarios are the best fit for a given wreck, often because the conditions for both scenarios may be present at the same time. However, the effects of small leakages due to localized corrosion are less evident (but this does not mean less harmful for the environment) than the total release of large oil quantity after a structural collapse.

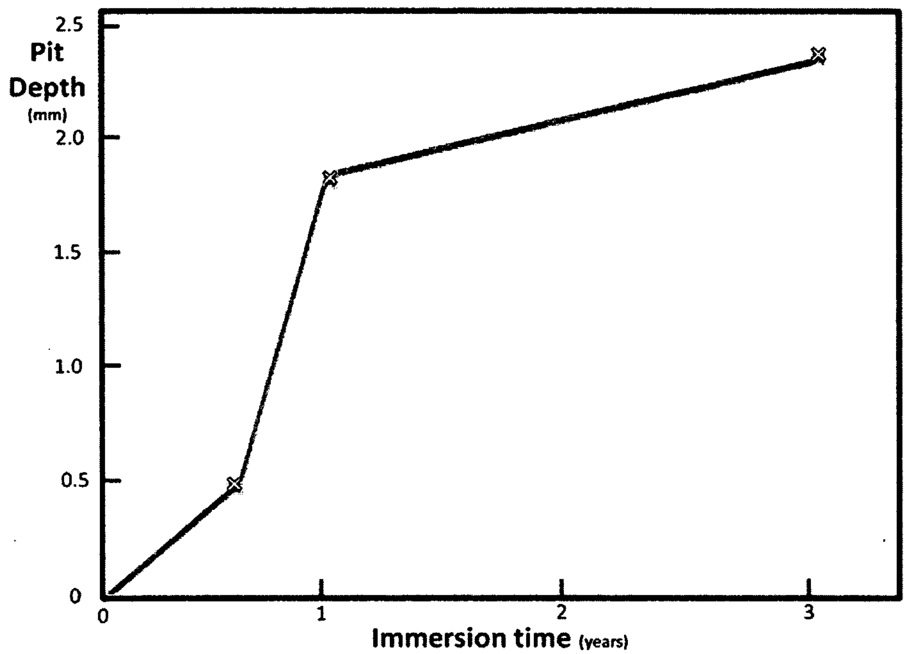


Figure 68 - Penetration depth of pitting corrosion for steel during the first years of immersion in seawater (Griffin et al., 1989).

Figure 68 (Griffin et al., 1989) indicates that the mean depth of pitting on AISI 1019 steel after a 3-years immersion period is about 2.4 mm. Comparing this value with a penetration depth value of 0.2 ÷ 0.3 mm of general corrosion, it is possible to see that – during the first stage of immersion – the penetration depth due to localized corrosion is of more than one order of magnitude greater than general corrosion.

These results are also in agreement with longer test data (Southwell et al., 1976). This 16-years tests show that the corrosion process becomes linear after the first 1-2 years (Figure 69). The extrapolation to 60 years of exposure (broken lines) is useful to calculate the maximum penetration depth at the end of this period. Here, the penetration depth due to general corrosion is of about 3.5 mm, whereas the mean pitting depth has a value of 5 mm.

Corrosion, both general and localized, takes place on a shipwreck not only at the exterior of the hull, but also in its interior, as a consequence of the massive irruption of water during and after the sinking. This fact can nearly double the general corrosion rate of the structure, but only rarely influences the perforation of hull plates due to localized corrosion.

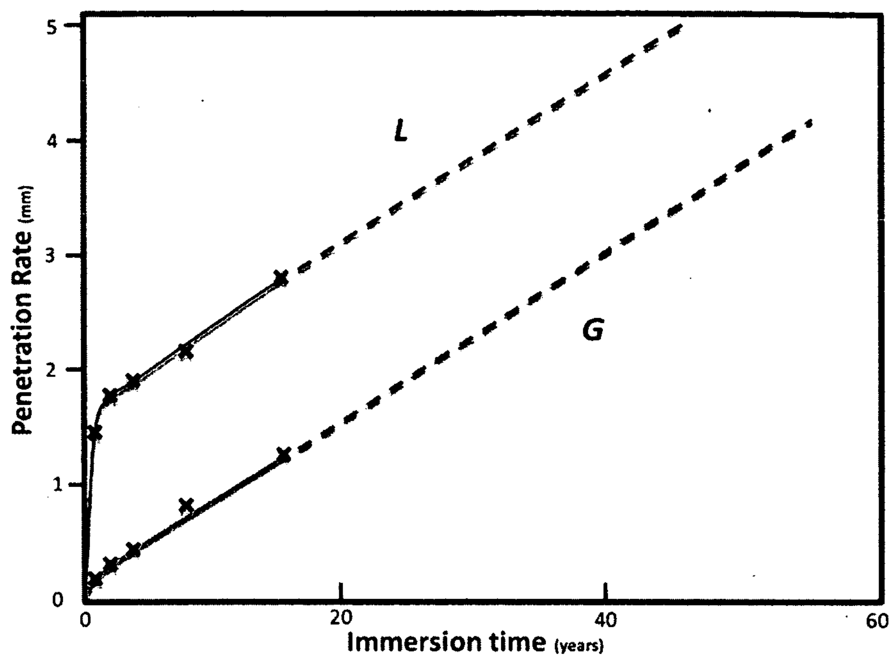


Figure 69 - Penetration rate on steel of localized (L) and general (G) corrosion as a function of immersion time (Southwell et al., 1976).

The fuel tanks for marine engines and crude oil tanks are subject to the same double effect due on the one hand to the exterior attack of seawater, and on the other to the interior action of the fuel, which usually contains substances like water, chlorides, sulfurous compounds (Baikin et al., 1980; Gonik, 2003; OCIMF, 1997) and SRB type micro-organisms (Benka-Coker et al., 1995; Cleland, 1995). In particular, water is present in crude oil in the form of an emulsion and tends gradually to separate and collect down, where corrosion processes can develop under anaerobic conditions.

Another area prone to accelerated damage is the junction of hull plates. Before WWII, the connection was carried out by means of rivets, and the interstices existing between the plates represent a preferential anodic area. The corrosion products built up in the interstices can then exert such a pressure that rivets break off and plates detach one from the other (Birkholz, 1997).

On the contrary, the hull plates of ships built during and after WWII were welded together, but this method also presents some disadvantages. In fact, the seam, especially if hand-made, suffers often from galvanic corrosion. Both these types of damage represent a risk with respect to the structural stability of the ship.

Case Studies

Exhaustive and quantitative information about actual corrosion rate on shipwrecks does not exist. Mainly this is due to the costs of such measurements and to the fact that this kind of measurements is not usually the aim of submarine explorations.

A limited documentation exists on shipwrecks of historical significance – i.e., the *Titanic* (1912) and the battleship *Arizona* (1941) – and of some others (Macleod, 2006; Schwemmer, 1996).

For the *Titanic* – which lies in the Atlantic Ocean at a depth of about 4000 meters – explorations on the wreck by means of submarines estimate that the structure will collapse within the next 50-90 years. Taking into account an original thickness of the hull plates between 18 and 27 millimeters and a mean general corrosion rate of 0.065 millimeter per year, it is possible to calculate that within such a time the overall corrosion penetration depth will reach about 50 % of the original plates thickness (Stephenson, 2005).

The wreck of the battleship *Arizona*, which lies on the seabed of Pearl Harbor at a depth of 10 meters, still contains about 4000 tons of fuel oil. It is slowly coming out of the tanks at a rate of 1 gallon/day. On the basis of a finite element model, the structure runs the risk of collapsing within the next 10-20 years, with a sudden "eruption" towards the sea surface of all the oil (Ruane, 2006). In the case of *Arizona*, as for the *Titanic*, only limited information, not based on effective measurements, of the residual thickness of the steel hull plates is available. These data are mainly only qualitative data on their present situation.

Figure 70 - The *Titanic*'s foremast: the forward half collapsed in the trough created by the after half because it was no longer capable of withstanding its own weight.

More exhaustive observations, based on corrosion potential, pH, as well as direct plate thickness measurements, were carried out on the iron shipwreck *City of Launceston*, sunk in Port Phillip Bay (Australia) in 1865. Moreover, the environment surrounding the wreck was very well characterized as regards the physical-chemical parameters, like dissolved oxygen, pH value and water temperature (Macleod, 2005). The wrought iron plates of the hull had an original thickness of about 16 millimeters, and, since the present thickness does not exceed 2 millimeters, this means that the corrosion has already destroyed nearly 90 % of the total original thickness of the plates. On the basis of these data, the study's author believes that the shipwreck is near to collapse. Moreover, it can be deduced that the mean general corrosion penetration rate during the 136 years of immersion had a value of about 0.103 millimeters per year, which satisfactorily agrees with the value (0.100 millimeters per year) commonly used by nautical engineers in the life calculations of ships still in service.

In the case of the site of the *James Matthew* – a 107 MT snow brig wrecked on 22nd July 1841 about 100 meters offshore in Cockburn Sound (Australia) – the application of zinc sacrificial anodes to effect *in-situ* conservation of the objects has been shown to be an effective method of site stabilization (Heldtberg et al., 2004). So in late June 2009 five 80kg zinc anodes were attached to the *City of Launceston* in attempt to slow the rate of corrosion of the hull and structure, and prevent the collapse and dispersal of the site (MHU, 2009).

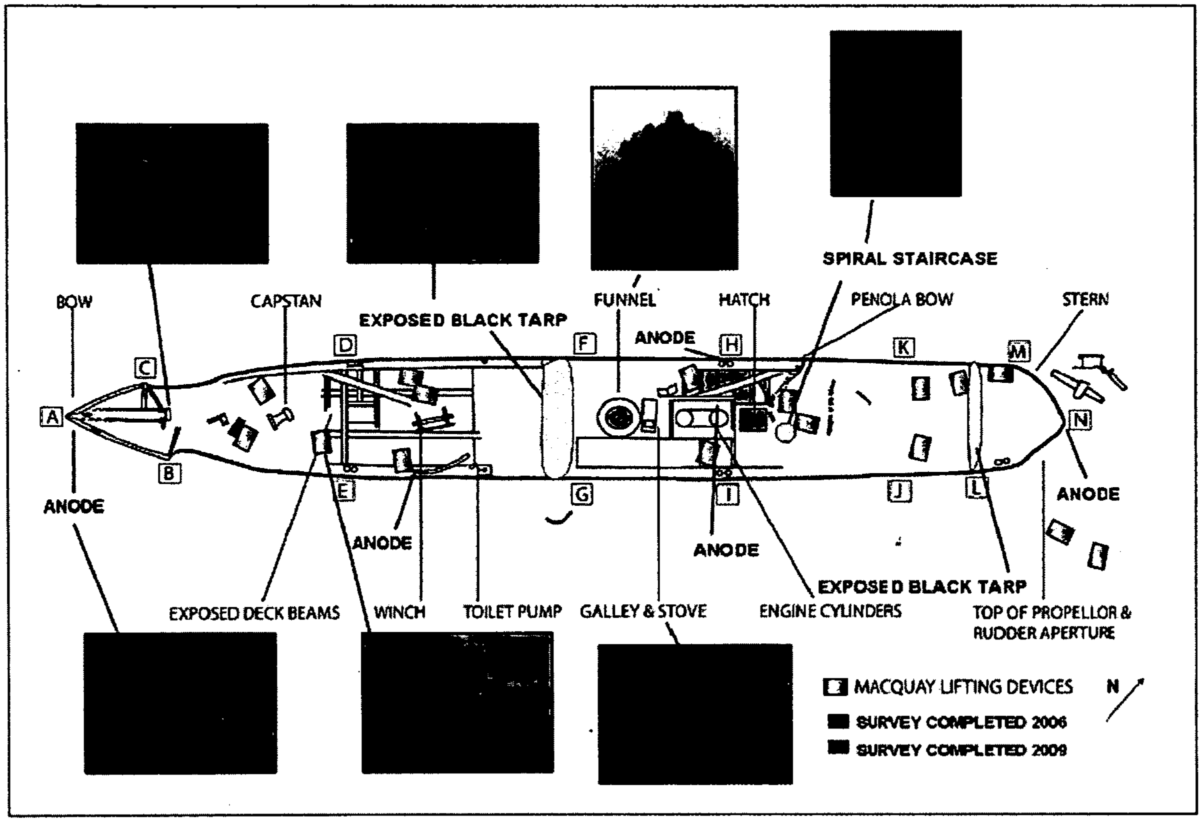


Figure 71 - Site plan of the SS City of Launceston (MHU, 2009).

APPENDIX B

LEGAL ASPECTS

Introduction

When the United Nations Organization was established at the end of WWII, a specialized Agency - called the Intergovernmental Maritime Consultative Organization (IMCO) - was created with the aim of harmonizing national legislation and to avoid freedom of navigation being restricted by State intervention. In particular, the task of improving the standards of building and operating a ship and promoting international co-operation in maritime affairs was given to the Maritime Safety Committee (MSC).

When the movement of crude oil by sea grew enormously, there was a general agreement for improving the operational standards of tankers due to a general concern for the hazards related to possible collisions. The MSC favored the adoption of separation-of-traffic schemes – that is, the definition of one-way traffic routes on either side of a central safety zone – in high-density areas. The limited acceptance of the separation-of-traffic schemes is one of the cause of the tragedy the liner *Andrea Doria*.

The IMCO framework permitted the creation of some primary Conventions such as, for instance:

- The International Convention for the Prevention of Pollution of the Sea by Oil (ICPPSO) of 1954 which with its amendments of 1958 entered into force in 1962;
- The International Convention for Safety of Life at Sea (SOLAS) of 1960.

The rapid evolution of shipbuilding techniques and operating standards required a continuous amendment procedure to the text of the original Conventions. Since the procedure in use, as well as the increasing number of States, complicated the adoption of any amendment to International Conventions in force, a new system was designed in order to secure a quick reception of the amendment by all contracting parties. Thus, the tacit acceptance of amendments to Maritime Conventions circulated to all member-States was evaluated as the best way to speed up the procedure. This new procedure referred exclusively to technical amendments, while the non-technical amendments continued to follow the traditional acceptance procedure of the approval and ratification by two thirds of the Assembly's members.

The International Convention for the Prevention of Pollution from Ships (MARPOL), adopted in 1973, incorporated this new procedure. Furthermore, a new text of SOLAS was elaborated in 1974 in order to up-date the original one and to incorporate all amendments adopted in previous years (1966, 1967, 1968, 1969), but not yet entered into force due to the old laborious procedure.

The practice of discharging oil residuals at sea was banned in most of the Mediterranean Sea with the amendments of the ICPPSO and special duties and certificates were introduced for the correct operation of a tanker in international waters.

The development of oil extraction in the area of the Persian Gulf and the temporary closure of the Suez Canal, due to the political crisis of 1956 in the Middle East, brought a new trend towards building very large crude oil carriers to absorb the additional costs for the alternative route by the Cape of Good Hope. The exposure of the sea to risks of pollution grew due to the development of giant-size tankers (Barada, 2007).

An International Fund as Consequence of the *Torrey Canyon* Disaster

When the tanker *Torrey Canyon* grounded off the Cornish coast on the Seven Stones Rocks on 18th of March 1967, it started to release its cargo of 117,000 tonnes of crude oil on a coastal area which was a tourist attraction. This event represents a legal turning point regarding the freedom of navigation and the right of the flag State to a ship in international waters (Tanikawa, 2003). The British Government, pressed by the exceptional circumstances, used a napalm bomb to set afire the wreck and to avoid the stranding of oil on the coast. This unilateral decision was taken although the *Torrey Canyon* was outside British national waters.

After the *Torrey Canyon* disaster, public opinion demanded urgent measures to avoid similar recurrences. At that time, international civil maritime law did not provide effective legal grounds to adopt unilateral action except the right of pursuit for criminal actions. Only in the USA did anti-trust legislation claim jurisdiction on practices, activities and events carried on outside the national territory, if it affected the economic interest of any US private or public subject.

The awareness of the heavy consequences of a spillage of crude oil from a grounded or sunken tanker in a restricted area, such as the Mediterranean Sea, urged the calling of an International Conference on Oil Pollution of the Sea (ICOPS) in Rome to debate this problem. During this conference, held between 7th and 9th October 1968, eight Resolutions were adopted. Among those one – the 6th Resolution – urged IMCO to re-consider the ICPPSO *"in the light of the knowledge and experience gained in recent years"*.

Furthermore, the Resolutions stressed the need of Governments:

- To cooperate effectively to reduce the effects of sea pollution by oil, recommending the creation of a Mediterranean Organization to deal with this type of disaster;
- To encourage scientific research on the biological effects of pollution and its possible curatives;
- To promote a rapid universal adoption of separation of traffic schemes in all parts of the world where the density of traffic warrants their introduction (i.e., the approaches to main harbors);
- To give special attention to the threat to resources and amenities caused by any form of sea pollution.

Also on the basis of these ICOPS Resolutions, the International Maritime Committee elaborated in 1969 the text of the Civil Liability Convention for Oil Pollution Damage (CLCOPD) and the Convention on the right of intervention in High Sea by the coastal State affected or threatened by an incident occurring outside its territorial waters.

In these cases the liability must be based on the risk determined by the activity (strict liability), not on the fault ascribed to a person (conventional liability). In case of strict liability, it is necessary to prove the link between cause and effect, i.e., the risk created by one person and the damage suffered by another person. However, the insurance of a very large crude oil carrier may not be enough to cover the potential damage of a big leakage, in spite of the introduction of the strict liability principle. In this case, the ship-owner could always invoke liability limitation according to the usual maritime practice.

In 1971 a Convention established an International Fund for compensation of Oil pollution damages to supplement the above mentioned Civil Liability Convention of 1969. In fact, the CLCOPD establishes that the ship-owner is entitled to limit his liability to an amount of $183 per ton of the ship tonnage or $19.3 million whichever is less.

The Fund, which entered into force in 1978, provides supplementary compensation to those who cannot obtain full compensation. It is financed by contributions levied on subjects who have received in one calendar year more than 150,000 tonnes of crude oil and heavy fuel oil. The levy is based on reports of oil receipts in respect of individual contributors. The Reports are submitted by the Governments but the contributions are paid directly to the Fund by the contributors (Jacobson, 2003). The Fund is relieved of its obligation if it proves that the damage resulted from the willful misconduct of the owner or by lack of compliance with the International Conventions requirements.

At the time of the *Haven* incident, the maximum compensation payable by the Fund was $83 million. It has been raised with the Protocol of 1992 to $186 million and now even higher. The spillage or the threat of spillage of heavy oil and crude oil are the basis of the compensation for costs to limit or eliminate effects. The compensation is also used to indemnify the fishing industry, tourist operators and any other subject who has suffered damage due to the oil spill.

In the 1992 text of the Fund Convention there is no distinction between laden or un-laden tankers. This particular distinction has important consequences as far as the eventual cost for the removal of a wreck may be concerned. Furthermore, the geographical scope of application of the Conventions extends to cover the Economic Exclusive Zone. The 1992 text also remarks that cost for preventive measures adopted in order to cope with a grave and imminent danger of pollution damage are recoverable even if an oil spill does not actually occur (James, 2003).

The Montego Bay Convention (MBC) and the Marine Areas

To define rights and obligations of a Country about a shipwreck it is essential to define its correct geographic location. In the international field, the Montego Bay Convention (MBC) – adopted in 1982 – is an important regulatory point of reference about marine areas. The threefold aim of the Montego Bay Convention is to update the regulations provided by previous conventions' codification, to integrate the regulations which are partial or incoherent and to propose solutions for the new issues related to global technological development (Carbone et al., 2006). Moreover, the Montego Bay discipline is acquiring general relevance: in a few years, conventional rules with equivalent contents, also between non-contracting Countries, have been consolidated.

Into the text of MBC, territorial waters are fixed at twelve nautical miles from the baseline, while the contiguous zone is twenty four nautical miles. In this zone the coastal State could continue to enforce laws in four specific areas: pollution, taxation, customs and immigration (these powers are not exclusive as they could belong to several Countries at the same time). Unlike territorial waters, the contiguous zone must be declared by the coastal State to exist.

Even if they have the same extension and partially similar uses, the contiguous zone must not be confused with the archaeological zone. In this latter, on the basis of Article 303 of the MBC, the coastal State can exercise powers of control and repression of offences to National law about discovery and commercialization of historical and archaeological goods. Unlike the contiguous zone, it is required that neighboring coastal States circumscribe the spatial area subject to their own jurisdiction so that control over the recovery of goods lying on the seabed could be exclusive to a single Country. In the same article of the Convention, the third and fourth paragraphs provide for the preservation of owner rights (if identifiable) and reference possible international dispositions on shipwreck salvage (Scovazzi, 2004).

Among these dispositions, we indicate the recent approval of the UNESCO Convention on the Protection of Underwater Cultural Heritage (CPUCH) which outlines structured and exhaustive rules on the protection of historical and archeological goods present in the seas. This Convention became necessary because, as a consequence of recent technological progress, many private enterprises are going to attempt to exploit the many treasures still lying on the bottom of the sea, for economic advantage. Another important Convention on this issue is the Nairobi International Convention on the Removal of Wrecks (NICRW) adopted in Kenya in 2007. It will provide the legal basis for States to remove, or have removed, shipwrecks that may have the potential to affect adversely the safety of lives, goods and property at sea, as well as the marine environment.

Beyond the archaeological zone, Article 149 of the MBC establishes that every historical and archeological object found within 200 nautical miles must be preserved or ceded in the interests of all mankind, but taking in account the preferential right of the Country of origin or the Country of cultural, historical or archaeological origin.

In addition to the abovementioned regulations, the MBC establishes duties and responsibilities about marine environmental safeguarding of the Countries. Article 192, indeed, establishes that each Country has the obligation to protect and preserve the marine environment.

In the subsequent Article 193 the MBC specifies that the abovementioned obligation in not unconditional, but it must be balanced with the real technical and economical possibilities of each Country. However, each Country must commit itself to balance its own policy with the environmental protection. There is therefore a common responsibility of developed Countries to employ all the resources required to perform shipwreck environmental risk management.

The Nairobi International Convention for Wreck Removal

In 1982 the IMCO changed its name to the International Maritime Organization (IMO). Within IMO, a Maritime Environment Protection Committee (MEPC) was established to prevent pollution of the sea and degradation of the natural marine environment, even though the resistance to admit indemnification for environmental damage was outright. This creation proved that the sensitivity toward environmental problems was spreading within IMO.

New pressing rules were soon adopted for transport of chemical substances by sea and to limit the use of toxic substances in paints and varnishes employed on ships. The Dumping Convention of 1972 and the Bunker Oil Pollution Convention of 2001 were directed to avoid persistent pollution of the sea due to maritime transport.

If we consider the merchant fleet in the last half century, we should note a steady increase in consistency, but its renewal has not at the same pace. The abandonment of the single hull for tankers with a deadweight over 20,000 tons (about 20,300 tonnes) approved by the MEPC in 2001 has still to become operational. The increased use of very large crude oil carriers reduces the cost of operating a larger number of carriers, and also diminishes the risk of accidents and wrecks. However a navigational problem for one of them – as in the case of the *Torrey Canyon* – can have heavy consequences.

In many countries it is generally accepted that industry is responsible for the safe disposal of its waste products. In this view, the maritime industry should not escape this general trend, but the sea is still too often regarded as a boundless dumping ground. One indication of a possible change in this policy has been the elaboration on a Convention for Wreck Removal (CWR) that answers the concern about abandoned wrecks. At the same time, it is relevant to underline that the oil industry is not the only contributor to this type of pollution. Sea and coastal pollution may result from collision or stranding of any type of ship. Thus the problem of wreck removal has to be faced independent of whether there was or there was not a cargo on board and from the nature of the cargo itself.

Recovery or removal of wrecks could take place according to the terms of the Salvage Convention of 1989, if there was good will of the owner or the salvage company. Unluckily, the salvor has often no real interest in removing a wreck that is considered a negative value. The heavy cost of the operation of dismantling and disposal of the wreck has the utmost attention. Thus, when the removal of the *Haven* wreck was taken into consideration, the high cost led to the final decision by the Administrative Authority to strip the wreck of all hazardous and noxious substances and to secure free passage over it though the demolition of the funnel and the top part of the wreck. The *Haven* disaster had a dimension which could not be covered by the Civil Liability Convention or by the Fund Convention since the 1976 Fund Protocol had substituted the reference to the original gold Franc with the Special Drawing Right (SDR) of the International Monetary Fund. Because of this substitution the roof of the insurance cover for the Haven disaster has been seven times lower than it was.

In case of an environmental disaster there is typically no room to discuss the opportunity of proceeding to wreck removal. For this, the CWR would impose an additional duty for the removal of a wreck on the owner or operator of the distressed ship: a duty backed by an insurance certificate. This additional insurance certificate could allow a Governmental Authority to recover at least partially the expenses for removing a wreck. On the other hand, any new burden placed on ship-owners and ship operators may in turn be an obstacle to the freedom of navigation and hamper sea trade. Therefore the owners of the cargo, the shippers and the receivers should be jointly bound in the obligation to remove the wreck and its cargo (Barada, 2007).

This Convention does not apply to the territorial waters of any Country if not by express consent, but only to its EEZ or, when a zone has not been established, within an area *"extending not more than 200 miles from the baselines from which the breadth of its territorial sea is measured"* (art. 1). Furthermore, measures taken under the International Convention relating to intervention on the High Sea in cases of Oil Pollution Casualties of 1969 or under the Protocol relating to Intervention on the High Sea in Cases of Pollution by Substances other than Oil of 1973 are also not part of the CWR scope.

War Graves and Sovereign Immunity of Sunken Military Vessels

The maintenance of the sovereign immunity of military vessels sunk by any means is the object of international debate. Even if in the field of international law it is not possible to define a routine procedure, it is commonly held that a sunk warship represents an inviolable war grave. However this principle was not recognized by the UNESCO Convention on the Protection of the Underwater Cultural Heritage where, at article 2 paragraph 9, it only states that *"States Parties shall ensure that proper respect is given to all human remains located in maritime waters"*, without making any distinction between the remains of soldier within warships and sailors of merchant vessels.

During the working procedure of the UNESCO Convention, a proposal was advanced to add into the text a special case for warships' wrecks (UNESCO, 2001). This proposal met the opposition of a large number of states, in particular Africans, with the aim to avoid any distinction among the remains of deaths at sea, remembering the huge number of slaves dead during the travel from Africa to the Americas (UNESCO, 2001).

Nothing more specific can be traced within the Montego Bay Convention (1982). This convention, at article 303, states two very general duties for the States about historical and archaeological goods, wherever they are discovered:

- States *"have the duty to protect objects of an archaeological and historical nature found at sea"*;
- States shall co-operate for this purpose.

Furthermore, even if the text of the Convention provides indications about the territorial waters, the contiguous zone and high seas, there is a lack of provision for the continental shelf (Dromgoole, 2005).

Even if there is an absence of customary laws or treaties that state the maintenance of sovereign immunity for sunken military vessels, a series of arguments can be cited towards this principle.

First of all, these wrecks may contain sensitive documents, weapon systems and other devices that the Flag state wants to avoid coming into possession of other states (Mainetti, 2002). Furthermore, the presence of crew remains is an additional reason for the maintenance of sovereign immunity (Harris, 2001). In fact, it is a common opinion that particular attention has to be reserved for the wrecks of warships due to the presence of the remains of soldiers. In addition, it is also possible to cite article 34 of the First Protocol of the Geneva Convention about the graves of people killed in a foreign country during hostilities. At article 34 paragraph 1 this Protocol states that *"The remains of [...] persons not nationals of the country in which they have died as a result of hostilities shall be respected, and the gravesites of all such persons shall be respected, maintained and marked as provided for in Article 130 of the Fourth Convention, where their remains or gravesites would not receive more favorable consideration under the Conventions and this Protocol"*.

The declared position of a large number of maritime powers - present and past - agree with the principle that a State cannot interact with the wreck of a military vessel without the authorization of the Flag state. For instance, for the US, these wrecks - even if apparently abandoned - are the property of the Government and the Department of the Navy which *"retains custody of all of its naval vessels and aircraft, whether lost within U.S., foreign, or international boundaries"*(see *http://www.history.navy.mil/faqs/faq28-1.htm*). In the same institutional web page of the Department of the Navy, there are references to *"past court cases supporting this doctrine include litigation in Hatteras Inc., v. the USS Hatteras (1984) and U.S. v. Richard Steinmetz (1992, also known as the "Alabama bell case")"*. It is also stated that *"the treatment of historic naval aircraft throughout the world's oceans"* follow the same laws (see *http://www.history.navy.mil/faqs/faq28-1.htm*). The principle of sovereign immunity was already present in the *Abandoned Shipwreck Act* (1990).

The principle of imprescriptibility for the rights of a Flag state (i.e., that rights to the remains cannot be taken away, or lost through time) is also proclaimed within the *Statement on United States Policy for the protection of Sunken Warships* of 19 January 2001. This Statement also contain the following statement: *"The United States will use its authority to protect and preserve sunken State craft of the United States and other nations, whether located in the waters of the United States, a foreign nation, or in international waters"*. A particular evaluation about this Statement is that it could represent a "shooting itself in the foot" since the United States represents the state that has most technological capabilities in the international field (Harris, 2001). The principle of sovereign immunity was underlined again from the US with the *Sunken Military Craft act* (2004).

Similar arguments about the sovereign immunity of sunken warships come from France, that states the *"imprescriptibilité de l'immunité des navires d'Etat"* (France, 1999), as well as the United Kingdom with the *Protection of Military Remains Act* (1986). Furthermore, the United States, France and United Kingdom reached an agreement with Germany and Japan in 1995 for the *"Joint Statement on Sunken State Vessels and Aircraft"*.

The positions of these states are in conflict with the 2001 UNESCO Convention where it states that there are circumstances in which the coastal state can decide the future of a state vessel without the opinion of the Flag state. In fact, article 3 paragraph 7 states that a coastal state *"should inform"* the Flag state about wrecks present in the territorial or archipelagic waters; while, if the wreck is in the EEZ or in the continental shelf, has the right to decide if it should *"prohibit or authorize any activity directed at such heritage to prevent interference with its sovereign rights or jurisdiction"* (article 10 paragraph 2).

On the sovereign immunity of sunken warships, it is of particular interest to consider both the discovery of relic shipwrecks for which it is not easy to prove the continuity of possession from the Flag state, and the wreckage of warships that happened in the last one hundred years in international waters or in the territorial waters of another state. Both these situations are briefly discussed in the next two paragraphs.

Wrecks of Relic Warships

A fundamental starting point is represented by the definition of 'warship'. In fact, while it is easy enough to identify a modern warship, this operation becomes enough complex in the case of the discovery of very old vessels such as, for instance, Viking ships or Spanish galleons. In particular, the criterion of the presence of weapons onboard can provide some interpretative doubts.

Article 29 of the Montego Bay Convention states that "warship" is a ship that belongs to the armed forces of a State *"bearing the external marks distinguishing such ships of its nationality, under the command of an officer duly commissioned by the government of the State and whose name appears in the appropriate service list or its equivalent"*, and manned by *"a crew which is under regular armed forces discipline"*.

It is complicated to check these criteria: for instance, to verify the concept of armed forces discipline among the Dutch navy of the 17th century, as well as to check if the commander of a certain vessel was present, at a given time, under the roles of officers of a certain navy. More generally, it is particularly hard to verify the criteria of article 29 for all those ships used for private aims before the Paris Declaration Respecting Maritime Law of 16 April 1856, issued to abolish privateering. Specifically, the extension of the definition of warships to all the vessels employed by the maritime powers for trading with the Americas' and Oriental Indies' colonies is controversial (Bederman, 2000).

The definition of "warships" was also debated during the working procedures of the 2001 UNESCO Convention. In the final text, article 1 paragraph 8 states that as "State vessels" both the warships and other types of vessels *"owned or operated by a State"* that, at the time of the sinking, were employed *"only for government non-commercial purposes, that are identified as such and that meet the definition of underwater cultural heritage"*.

An example of this interpretative problem are the events related to the Spanish galleons *Juno* and *La Galga* discovered in 1997 by an American salvage company - the Sea Hunt Inc. - and claimed by Spain. From one side, the arguments provided by the company were based on the facts that the galleons were not sunk in war, that they were not employed *"for strictly non-commercial government service"* - carrying, for instance, tobacco, gold, passengers, etc. - and that the recovered goods do not come from Spain, but they were of African and Indian origins. On the other side, Spain defended its own rights to the galleons on the basis of their belonging to the Spanish navy, being in the service of the Spanish state at the time of the sinking and never being removed from the Spanish naval register (Shapreau, 2001).

The Spanish position in this situation was totally different from the one sustained in 1965 with the United States for some galleons sunk in American territorial waters. In fact, in a letter of 11 January 1965, the Spanish ambassador to the United States affirmed that no rights could be advanced by Spain on the galleons sunk in Florida's waters since *"if the discovery is considered marine salvage the owner of the ship and/or merchandise would have lost all rights because he abandoned any attempt of recovery; if the discovery is considered a discovery of a treasure in the territory under the jurisdiction of a state, the laws of this state will determine title to the treasure; in most cases extinctive prescription would act against any right possibly reclaim by a previous owner"* (Anon., 2000).

However, the court decision excluded the possibility that private citizens could try to salvage the galleons' wrecks without the authorization of the Flag state. Even if the Spanish position is common to an increasing number of States, it cannot be considered a principle of customary law both for the non-uniform acceptance from all the States and for its application often applied on a basis of political opportunities.

On the lack of uniformity, it is enough to remember the Declaration of Santo Domingo of 1998. In this declaration may states of South America stated that the Spanish galleons and the vessels of other European countries sunken in their territorial waters are abandoned and, thus, they are under the exclusive jurisdiction of the Coastal state (Brea-Franco, 2000).

In relation to the application of the definition of 'warship' based on political opportunities, the case brought by the French government over the *Orient* sunk on 2 August 1798 by Admiral Nelson in the Egyptian bay of Abukir is exemplificative. France, even if it initially asserted rights over the vessel's wreck, had then renounced the rights, officially due to the shortage of funds to assign to the wreck salvage (Garabello, 2004). In the same period, France declared its intention to recover seven vessels discovered in Venezuelan waters (France, 1999). A possible interpretation of the French behavior for the Orient can be found in the principle of reciprocity. In fact, this principle, even if it does not produce an obligation, represents a courtesy form that consists in recognizing the same rights of the other State over the goods of archaeological interests in its own territory. Thus, the French willingness to resign rights was to avoid the question of the restitution of the many Egyptian archaeological goods present in French museums.

Another interesting event, although ongoing, is related to the English vessel *HMS Sussex*, sunk in 1946 during a strong storm while it was convoying a merchant fleet near the Straits of Gibraltar. After the discovery of a diplomatic letter indicating that the English vessel was carrying a large amount of gold and silver, an American company - *Odyssey Marine Exploration* - obtained from the United Kingdom and Spain the necessary authorizations and started investigation over an area of about 400 square nautical miles that, in 2001, identified by side scan sonar the possible wreck at a depth of about 820 meters. Although the coordinates were not disclosed, *Odyssey* reported in its Preliminary Report that the *HMS Sussex* wreck was in International waters (Cunningham Dobson, 2009). This localization partially simplified the case since, on the basis of the "Protection of Military Remains Act", the United Kingdom has ownership of all the military wrecks except the ones within the territorial waters of other states. However, in the case that the wreck was closer to the coast, Spain should assign to the United Kingdom the right of property, as such rights were reclaimed by Spain for the galleons *Juno* and *La Galga*, avoiding the involvement of this case in the dispute related to the absence of territorial water for Gibraltar based on Article 10 of the Treaty of Utrecht (1713).

The epilogue of the case involving the *Juno* and *La Galga* galleons pushed *Odyssey* to look for an agreement with the government of the United Kingdom as Flag state of the *HMS Sussex,* before starting any salvage operation. Although not all of the details are disclosed, *Odyssey* signed an agreement that recognized the rights of the United Kingdom, as vessel owner, to approve the recovery plan for the *HMS Sussex*, as well as to a proportional part of the revenues coming from selling the recovered goods.

Furthermore, particular attention was reserved in case of the eventual discovery of human remains. In this case, even if judged highly improbable since prior experience of other deep water shipwrecks suggests that the presence of human remains is unlikely, the company would notify the British government and proceed so as to avoid unwarranted disruption to the remains, leaving them *in situ* where possible.

Since December 2005, *Odyssey* has started to execute the final phases of the project. However, the operations, suspended on 15 January 2006 after the deterioration of relationships between the company and the Spanish authorities, have still not resumed. This happened even though the Spain, with an agreement reached on 23 March 2007, obtained an agreement to participate with its own archaeologists in the operations, and to be provided with information about other wrecks present in the area.

The legitimacy of the agreement between *Odyssey* and the UK government has been subject to debate since it appears to neatly violate the Annex of the cited 2001 UNESCO Convention that was approved by *consensus* by the UK. In particular, rule 2 of this Annex states that *"[t]he commercial exploitation of underwater cultural heritage for trade or speculation or its irretrievable dispersal is fundamentally incompatible with the protection and proper management of underwater cultural heritage"*, further clarifying that *"[u]nderwater cultural heritage shall not be traded, sold, bought or bartered as commercial goods"*.

Thus, the agreement between the UK and *Odyssey*, even if positive in the parts that foresee the presence of archaeologists and the respect of scientific standards during the operations, do not respect the 2001 UNESCO Convention because it foresees the commercialization of archaeological goods (Garabello, 2004). This could be the reason for the absence of any reference to the Convention in the agreement.

Finally, it has to be remembered that the 2001 UNESCO Convention approach has been often criticized as fruit of an *"unfortunate, misguided and counterproductive"* prejudice against *"salvage law"* (Dorsey, 2000). Furthermore, it was also negatively judged since, being similar to part XI of the Montego Bay Convention for exploration and exploitation of the Area, it becomes unreasonable for private companies to *"mobilize vast amounts of time, capital, and technology to locate, study, and recover under water cultural heritage"* (Bederman, 1999).

Warships Sunk Less Than One Hundred Years

The 2001 UNESCO Convention cannot be applied to warships sunk less than one century ago, since Article 1 states that the Convention is valid for *"all traces of human existence having a cultural, historical or archaeological character"* that were underwater *"for at least 100 years"*.

For more recent wrecks, it is common that States sign bilateral agreement from which both the Flag state and the coastal state can obtain advantages. Some examples in this direction are:

- The agreement in 1989 between the UK and South Africa for a British warship - *HMS Birkenhead* - sunk in South African waters (Anon., 1990; Matera, 2002; Staniland, 1999);
- The agreement in 1997 between the UK and Canada for the *HMS Erebus* and *HMS Terror*, two warships sunk, but not localized, within Canadian waters (Garabello, 2004);
- The contacts between Australia and Turkey for an Australian submarine sunk in Turkey during the First World War (Australia, 1996).

One of the first example of this solution was the agreement signed on 6 November 1952 between Italy and the United Kingdom over the wreck of *HMS Spartan*, a British warship sunk in Italian waters. On the basis of this bilateral agreement, half of the revenue obtained selling the recovered parts of the wreck should go to the UK government. Furthermore, this latter should be exonerated of any responsibility related to the wreck, and any document, cipher, money or remains of British citizens should be given back to the British embassy (USA, 1953).

From this agreement, it is evident that Italy recognized that the UK had both sovereign immunity and ownership of the *HMS Spartan*'s wreck. Furthermore, the same agreement also contains some larger provisions for the future: the notification from the UK to the Italian government of the names and the positions of all the British warships present in Italian territorial waters; and, in the case that some of these wrecks were required to be removed, the duty for Italy to inform the UK government for subsequent agreements in case this latter decides to proceed to salvage the wreck itself.

A case in which sovereign immunity was recognized for an Italian wreck is related to the submergible *Sciré*, sunk on 10 August 1942 by the British navy near Haifa Harbor in very shallow waters (about 32 meters). In the sinking more than seventy Italian soldiers passed away and only two bodies reached the coast to be buried at Haifa (and then returned to Italy in 1984).

Some years after the sinking, the wreck was identified by an Israeli researcher at about nine kilometers offshore. The Italian government was informed about the discovery and it organized three expeditions to the wreck in 1960, 1963 and 1984. During the last expedition, the Italian navy ship *Anteo* recovered the remains of part of the crew and some pieces of the wreck (Anon., 1984). Furthermore, the expedition bolted the hatches of the wreck to avoid the war grave being easily accessible.

During all these event the Israeli government recognized the sovereign immunity of Italy to the sunken submarine, but also it stimulated Italy to take decisions about the recovery of the human remains that were in the hull and still easily visible, opposing any private initiative to salvage the wreck.

Unfortunately, in September 2002, during a joint exercise between the American and Israeli navies, the *USNS Apache* came in contact with the submergible wreck, damaging and partially reopening the entrances. The gravity of this event was sharpened by some remains still present within the wreck and this prompted the presentation of official apologies to Italy from the Israeli government (Anon., 2002; Franceschini, 2002). In the field of international law, this sad event was deplored since the behavior of the American vessel was totally incompatible with the declarations released in the mentioned *Statement on United States Policy for the Protection of Sunken Warships* of 2001.

This event recalls a similar case in 1974 when the US proceeded to the partial recovery of a Soviet submarine sunk in International waters, about 75 miles northwest of the Hawaiian Archipelago. The recovery operations, during the 'Jennifer Project', was executed only seven years after that the Soviet Union reported the submarine sunk. It is evident that, on this occasion, the United States did not recognize the sovereign immunity on the Soviet submarine, even if there were no explicit acts of abandonment and that the operations caused protests from the Soviet Union (Bederman, 2000).

APPENDIX C

OIL CHARACTERISTICS AND RELATED SPILL CONSEQUENCES

Oils mainly differ from each other in viscosity, volatility and toxicity:

- Viscosity can be described as the oil's resistance to flow;
- Volatility gives indications about the time necessary for oil evaporation into the air;
- Toxicity mainly refers to how an oil type is toxic to either human beings or other organisms.

This last characteristic is often difficult to evaluate since oil is a mixture of various chemicals, with different proportions between chemicals within the same oil type (e.g., Alaska North Slope has a different mixture than an Arabian crude oil). Furthermore, an oil spill may have diverse concurrent toxic effects:

- Birds and mammals may die when oil reduces their insulation, fouling fur and feathers;
- Small organisms may be smothered by an oily layer washing ashore;
- Small amounts of hydrocarbons may prejudice the development of larvae and eggs, etc.

The combination of viscosity, volatility, and toxicity determine the different effects of an oil type on the environment, and on the difficulties related to clean up. A common classification is based on four main categories which are described in Table 26 along with possible effects on the shoreline.

In case of a release of a light crude oil, the lighter and more volatile components will be lost by evaporation, dissolution and biodegradation. The water-soluble fraction, which principally contains aromatic hydrocarbons and polar compounds, will be responsible for the acute toxicity effects on organisms. The remaining heavy fraction will become attached to the substrate or sequestered in the sediments. The tar-like residue will persist for many years (Talouli et al., 2009).

Since war ships were potentially required to operate in cold as well as tropical waters – and in all seasons – the fuel oil was a blend that allowed it to remain fluid in arctic waters without heating. Fuel Oil Number 5 is typically about 75-80% Fuel Oil No. 6. Navy Special Fuel Oil (NSFO) and is cut with a lighter marine diesel up to 20% so that in some cases it has a final viscosity of 170 centistokes at ambient temperature. The light oil component will entrain into the water column and weather quite quickly when spilled in tropical waters whereas the residual oil will be more persistent and likely to be transported long distances by wind and currents until it comes ashore or impacts on fringing reef systems. A sample of oil leaking from the *USS Mississinewa* was subjected to room temperature weathering and lost only 15% of its mass within one day (Gilbert, 2001). This demonstrated the lack of light hydrocarbons and the persistent nature of the oil.

Table 26 - Characteristics and effects on the coastal area for category of oil types, adapted from (NOAA, 2012).

Category	Oil Types	Characteristics	Effects on coastal area
1 Very Light Oils	Jet Fuels, Gasoline	Highly volatile (1-2 days). High concentrations of toxic (soluble) compounds.	Localized, severe impacts to water column and intertidal resources. No cleanup possible.
2 Light Oils	Diesel, No. 2 Fuel Oil, Light Crudes	Moderate volatility, with residual (up to one-third of spill amount) after a few days. Moderate concentrations of toxic (soluble) compounds.	Will "oil" intertidal resources with potential of a long-term contamination. Cleanup can be very effective.
3 Medium Oils	Most Crude Oils	About one-third will evaporate within 24 hours.	Oil contamination of intertidal areas can be severe and long-term. Oil impacts to waterfowl and fur-bearing mammals can be severe. Cleanup most effective if conducted quickly.
4 Heavy Oils	Heavy Crude Oils, No. 6 Fuel Oil, Bunker C	Little or no evaporation or dissolution. Very slow weathering.	Little or no evaporation or dissolution. Probable heavy contamination of intertidal areas. Severe impacts to waterfowl and fur-bearing mammals (coating and ingestion). Possible long-term contamination of sediments. Shoreline cleanup difficult under all conditions.

A possible definition for a non-persistent oil is the one used in legislation by the US Environment Protection Agency and US Coast Guard. Following this definition, an oil is defined as non-persistent when spilt on water:

- At least 50% of which by volume, distils at a temperature of 340°C (645°F); and
- At least 95% of which by volume, distils at a temperature of 370°C (700°F).

The above definition has also been adopted by the International Oil Pollution Compensation Fund. Oils and bunker fuels not in compliance with this definition are defined as 'environmentally persistent when spilt'. The determination of an oil persistence is significant since it involves:

- Spill movement, weathering extent and fate of the oil;
- Determining travel distances from spill site and therefore planning distances;
- Determination of on-water recovery techniques, logistics and placement of assets; and
- Scale of shoreline cleanup capacity and worst case planning volumes.

Oil Fingerprinting

Oil fingerprinting is the process of determining the original oil source through the analysis of the chemical composition of oil residues found in the environment. This complex process relies both on the results of ratios between certain discriminating hydrocarbon components and on the experience of the analytical chemist.

The U.S. Coast Guard adopts this process to determine a responsible party when an oil spill with no known source washes up on a shoreline. Furthermore, in case of large oil spills, fingerprinting can be used to verify if oil in the environment has also other inputs like natural oil seeps. For instance, in the NOAA Prince William Sound study related to the Exxon Valdez spill, fingerprinting identified other sources such as diesel fuel from boats working in the area, non-specific combustion sources (that could include anything from wood stoves to vessel exhaust), and spilled oil not linked to the Exxon Valdez (NOAA, 2012).

In some areas, like the Gulf of Mexico, a large number of oil seeps release important amount of petroleum and natural gas into the water.

The concurrence of different oil sources underlines the importance of the chemical fingerprinting process.

Oil Weathering

The overall behavior of an oil type is based on its chemical mixture. Each component interacts with the surrounding habitat that, thus, largely influences its behavior and changes its physical characteristics (e.g., density). For instance, some components tend to be quickly volatile; some others are more easily broken down by microbes present on the beaches, or be degraded by sunlight.

The oil weathering includes all the processes modifying over time chemical composition and physical characteristics of oil. The comparison between the actual chemical composition of the oil residue and the original source can be used to roughly determine the age of the spill event. Nevertheless, some NOAA studies reports remaining oil from the Exxon Valdez spill (1989) ranges from very weathered to relatively fresh (NOAA, 2012). The less exposed to the elements the oil is as in the case of buried residues, the more slowly it weathers.

Oil weathering in the marine environment is mainly related to the oil type and environmental conditions. The main environmental factors are:

- Area of slick exposed;
- Wind speed and sea surface roughness;
- Air temperature;
- Exposure to sunlight (solar radiation); and
- Formation of emulsions (since it slows down the evaporation).

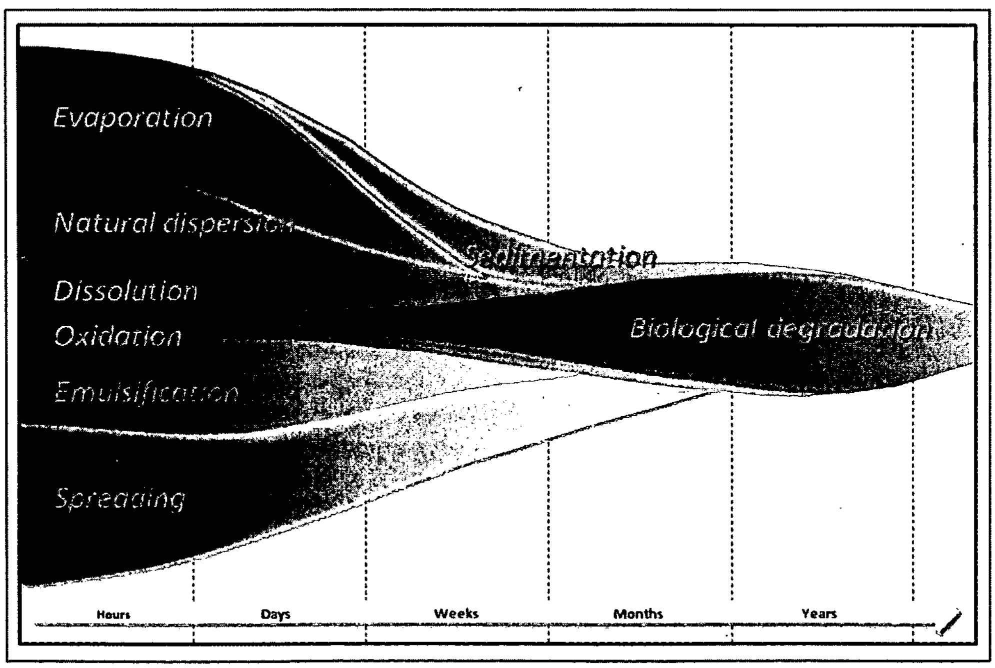


Figure 72 - Physical processes following an oil spill. Adapted from (Hassellöv, 2007).

The water largely alters the chemical and physical properties of oil. The main processes involved are (Figure 72):

- Evaporation, the lighter fractions disappear from the sea surface leaving a product of higher viscosity;
- Dispersion, the increasing contact surface between oil and water accelerates natural degradation of oil in water;
- Spreading, related to oil properties, but also on the prevailing wind and current conditions;
- Dissolution, generally only applicable to a small amount of the oil (even if, when dissolved in the water, the oil is no longer a threat for the shoreline, the toxicity to aquatic organisms is enhanced;
- Oxidation, usually limited to the surficial layer since it is primarily due to photo oxidation; and
- Emulsification, oil hydration results in enlarged volumes of oil pollution.

Spreading

Any shear on the sea surface causes a thickening of the slick in the downwind direction. A characteristics shape is similar to a comet, with a small region of black oil containing most of the pollutant and a much larger but thin sheen. Oil spreading is usually modeled using a numerical approach: the slick is divided into separate Lagrangian elements and each of them may be tracked taking into account wind stress, surface currents, random turbulence, Fay gravity-viscous forces, etc. (Fay, 1971). Among them, wind stress represents the most important cause of long term oil spreading. A rule of thumb is that an oil slick moves at approximately 3 % of the wind speed measured at ten meters above the water surface (Lehr et al., 2002).

Furthermore, a part of the oil slick is carried, as droplets of different size, into the water column through the effect of breaking waves. An estimation of the depth of this process is one and a half times the significant wave height (Delvigne and Sweeney, 1988). While the larger droplets come back quickly to the surface, the smaller ones remain under the sea surface for longer periods, trailing the moving main slick. With time, an oil slick may be split into multiple patches by wave action and by the troughs created on the water surface by Langmuir effects (Thorpe, 1995). In the long term, an oil slick usually forms small tarballs, with centimetric diameters as order of magnitude, spread out over a wide area.

Evaporation

Since any oil is a mixture of different compounds and the oil slick tends to lose first the more volatile components, the evaporation rate for an oil slick decreases with time. This process is mainly influenced by wind speed and water temperature (Lehr et al., 2002).

Two main approaches for modeling the evaporation process are to treat the oil as a uniform substance whose properties change as the slick weathers (Stiver and Mackay, 1984); or as a relatively small number of discrete, non-interacting components (Jones, 1997). In this last approach, the total evaporation rate is the sum on the rates related to each component.

The inhalation of volatile components of freshly spilled oil may be a significant health hazard for cleanup workers.

Emulsification

The emulsification process, driven by turbulent energy in the surrounding water, increases the volume of the oil slick. In fact, a fully "emulsified" and stable emulsion may contain 80 to 90 % water (Lehr et al., 2002).

This process is difficult to quantify since not all oils emulsify, while many crude oil starts to be affected by this process only after a certain amount of evaporation and photo-oxidation. Some studies pointed out that some oil components play a fundamental role: asphaltenes, resins, waxes (Bobra, 1991; Fingas et al., 1996). The process proceeds rapidly once the emulsification begins (Mackay et al., 1980).

Dispersion

This process is related to the natural turbulence in the water column that prevents small oil droplets from coming back to the sea surface. Dispersion, directly related to the oil viscosity, is sometimes enhanced by spraying chemical surfactants on oil slicks.

Models

Some integrated mathematical models have been developed recently to simulate the behavior of spilled oil such as the Automated Data Inquiry for Oil Spills (ADIOS) software package that was first released in 1994 (Reed et al., 1999).

This software developed by the NOAA Hazardous Material Response Division (HAZMAT) to aid responders in oil spill cleanup by forecasting the weathering processes and characteristics of oil slicks (Lehr et al., 2002). The software permits simulation of different types and ranges of release scenarios, providing the resulting uncertainty of the model output. The weathering processes included in the actual version (called 'ADIOS2') are spreading, evaporation, dispersion, sedimentation, and emulsification; while dispersants, in-situ burning, and skimming are the user cleanup options.

The validity of ADIOS and other similar models is usually based on comparisons with experimental data and observations of real spills, on the solidity of their theoretical bases, and on general acceptance by the spill community. Since oil spills are emergency events, the personnel involved are more interested in addressing the environmental damage than collecting data for future oil spill modeling. Thus, the weathering models are usually based on laboratory or small spill experiments. The lack of data about real and large accidental spills makes the calibration of oil weathering models a difficult task (Lehr et al., 2002).

Environmental Effects of an Oil Spill

The effects of oil spill on a coastal marine environment are a function of several factors, such as (Hassellöv, 2007):

- Oil type;
- Amount of oil;
- Type of shoreline; and
- Wind and current conditions.

The type of oil spill involved represents a key parameter for the extent of the environmental impact. The oil composition is dominated, in different proportions, by three classes of hydrocarbons: alkanes (paraffins), cyclo-alkanes (naphtenes) and aromatics (Smil, 2008). Cyclo-alkanes are the most abundant compounds in crude oil (about half of the weight). Light oils are usually considered more acutely toxic compared to heavy fuel oil, but with the advantage that, since the lighter oil quickly evaporates, they are not transported as far in the marine environment as heavier oils. Both oils can de-insulate bird's plumage, degreasing or besmearing feathers. Especially in cold climates, small amounts of oil can also become dangerous for bird life (Hassellöv, 2007).

In large WWII wreck sites, the majority of oil deposits are likely to have heavy fuel oil along with diesel, lubrication oils and some aviation fuels and gasoline. Submarines were mostly diesel driven with smaller quantities of fuel on board whereas large carriers, battle ships, and destroyers should have large quantities of heavier fuel oil. For instance, larger vessels sunk during the Pacific war have as fuel oil a blend of non-persistent as well as persistent oils: bunker oil (No #6 fuel oil) and marine diesel (No #2 fuel oil) (Talouli et al., 2009).

The consequences and the rate of recovery related to an oil spill reaching the shoreline largely vary as a function of the natural shoreline characteristics. Even when initial reports from an oil spill often describe something similar to a complete and unrecoverable ecological disaster, recovery will follow according to three main phases:

- Colonization of opportunistic species;
- Intermediate phase (between established colonization and the third phase); and
- Recovered ecosystem.

A study of the Exxon Valdez spill in Alaska compared, over a period of up to five years after, the different recovery of seafloors characterized by kelp and eelgrass communities (Dean and Jewett, 2001). The conclusions pointed out that the recovery was much slower in the eelgrass beds, which partly can be explained by less exposed localities and by community composition (dominated by relatively more sensitive species). These data corroborate the hypothesis that impacts from large oil spills are more persistent in soft-sediment environments (5-6 years) than for exposed rocky shore (1-2 years) (Hassellöv, 2007). More controversial is the question of whether the acute toxic effects in the water column decline rapidly due to dilution, or are delayed in the order of months (de la Huz et al., 2005; Dueri et al., 2009).

Diesel oil would weather relatively quickly, either being dispersed into the water column or evaporated within 12-24 hours. This does not mean there would not be ecological impacts on aquatic life, coral reefs or possible wildlife impacts but that it would be 'removed' from the water surface within a short time after release. Once dispersed or dissolved within the water column diesel oil could still have significant impacts on inter-tidal life and fisheries.

Heavier fuel oil mixtures released into the marine environment partially disperse and dissipate as the slick spreads, while some components dissolve into the water or settle into inter-tidal zones (if near shore). Finally, the heavier persistent components may form emulsions in rough seas or end up as tar balls and pats on shorelines, sometimes travelling long distances at sea.

Mitigation Operations

The cost of removing oil varies greatly depending upon the amount of oil, the depth of the wreck, and the availability of equipment. In the case of the *USS Mississinewa*, the operation had a total cost of $4.5 million corresponding to $2.25 a gallon (Navy, 2004). The *Luckenbach* cost $20 million to offload approximately 100,000 gallons, or $200/gallon (MLAUS, 2009). The wreck of the *SS Catala*, a grounded wreck in Washington State, cost approximately $225/gallon, while the *Princess Kathleen*, a passenger vessel that sank in Alaska, cost approximately $100/gallon (Grennan, 2010).

Entombing represents an alternative option. It consists of covering the wrecks in sand and concrete in order to contain dangerous or toxic pollutants. For instance, unexploded ordnance is a serious consideration that could justify entombing. This option has been suggested for the *USS Montebello* in Monterrey Bay (Basta and Kennedy, 2004) as well as for the wreck of *U-864*, a German U-boat. This latter has released several kilograms of mercury a year into important Norwegian fishing grounds (MAMMOET, 2008; Olsvik et al., 2011). This appears to be the least likely option as public opinion usually prefers a complete removal (Cowell and Gibbs, 2007).

Oil Clean Up

The three most common cleanup processes are: the use of chemical dispersants, mechanical skimming, and in-situ burning.

Mechanical skimming may be dangerous due to inhalation of toxic chemicals (e.g., benzene) by the vessel crew. For in-situ burning, the major concern is usually related to the resulting smoke plume (Lehr and Overstreet, 1998).

The remaining part of an oil slick that is not evaporated, dispersed or cleaned up may be deposited onto nearby shorelines. The prediction of the oil trajectories requires modeling of site specific winds and currents (Fallah and Stark, 1976).

Some attempts to clean up the oil from the shoreline may themselves harm the coastal and marine resources. For instance, the use of hot water and chemicals can be noxious for plants and animals, and teams of clean-up workers may increase the existing damage by trampling on sensitive organisms or mixing oil residuals deeply into the beach. These information come mainly from some recent large oil disasters, such as the *Exxon Valdez*, which have substantially increased knowledge about oil spill impacts.

APPENDIX D

METADATA

Recent awareness of the importance of how things relate spatially and the advancement of electronic technology has permitted an expansion in the use of digital geographic information worldwide. Individuals from a wide range of disciplines have started to massively produce, enhance, and modify digital geographic information. As a direct consequence, there has been a growth in importance of methods capable of assisting in understanding the complexity and diversity of geographic datasets.

The use of geospatial data implies modeling the real world for use in computer analysis and graphic display of information. Any description of reality is always an abstraction, mainly partial and just one of an infinite number of models. Each model is not an exact duplication of the real world: some characteristics are greatly approximated, others simplified, and some ignored. Thus, there is significant risk for a model of the real world to be misused if the assumptions and modeling rules adopted are not fully documented using metadata.

For instance, to exchange data between services, they have to share common schemas or to translate between them. If a Sensor Observation Service (SOS) represents a wind direction with a "wind blows from" conceptualization, and a Web Processing Service (WPS) computes the dispersion of a plume based on a "wind blows to" conceptualization, the simulated dispersion plume would point in the opposite direction. Thus the challenge is to establish semantic interoperability to exchange data in a meaningful way, without human intervention (Manso and Wachowicz, 2009).

Misunderstanding and incorrectly using geographic data can usually be traced back to missing or unclear descriptions of their intended interpretation (Guarino, 1998). In fact, most semantic conflicts during a workflow appear if source data has not been sufficiently specified at the beginning. A typical activity includes the discovery of relevant geospatial data, their pre-processing, the application of appropriate analysis methods, and finally rendering the results on a map. Catalogues can be used to discover resources published in a Spatial Data Infrastructure (SDI). Access to geospatial data depends on the underlying format: coverages are provided by Open Geospatial Consortium's Web Coverage Service (WCS), feature datasets are managed and offered by the OGC's Web Feature Service (WFS), sensor observations by the SOS that is a specialization of the WFS. Processing data, e.g. running a spatial analysis or interpolation algorithm, is accomplished by Web Processing Services (WPS). The result of the process, either a feature set or a coverage, are rendered by a Web Mapping Service (WMS) as a map in a common image format. The ambiguities from lack of data description are propagated through the whole workflow (Janowicz et al., 2010).

The application schema alone is not sufficient to grasp the meaning of the underlying data model: for instance, the labels identifying different entities are often ambiguous, and application-specific knowledge and semantic heterogeneities are a common source of misinterpretation (Maue and Schade, 2009). Metadata has the dual duty to allow a full description of a dataset by a data producer, and to permit an end-user to understand its applicability for any intended use.

The metadata must allow the end-user to:

- Assess if a dataset is suitable or not for a specific use (and how it matches the requirements);
- Evaluate data quality of the entire dataset, of specific feature classes, and sometimes also of attributes;
- Obtain technical details useful to import and convert the information;
- Obtain information about legal and ethical constraints on use; and
- Acquire knowledge about sources of further information (Goodchild, 2008).

Current Standard Metadata Formats

The actual process of standard creation is conservative. In fact, since the definition works by consensus, when a standard is accepted it usually does not contain the most recent research. The main task of the various ISO Technical Committees is the preparation of draft International Standards. After that the draft circulates to the member bodies for voting. The publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote. At the same time, the efforts required to write a new standard, as well as the usually high cost of applying a new one, make modifications difficult.

This section attempts to evaluate the current metadata standards with respect to spatial data quality. Since the metadata are succinct information, the choice of which information to embed must be oriented to the main typology of end-users and purposes.

Recently, some efforts have been made to make metadata standard as part of data formats and GIS software, allowing metadata to be stored and visualized with the data set itself (e.g., ESRI's Geoportal Extension supports the Dublin Core standard and several variants of both FGDC and ISO standards).

Content Standard for Digital Geospatial Metadata (CSDGM)

One of the earliest attempts to create a standard for geospatial metadata was the Content Standard for Digital Geospatial Metadata (CSDGM) adopted in 1994 by the U.S. Federal Geographic Committee. The standard also provides a common set of terminology for concepts related to the metadata.

The quality part of this standard was based on 5 elements:

- Attribute accuracy, that is the accuracy of the attributes by which geographic features are characterized;
- Logical consistency, which represents the degree to which the contents of the data set follow the rules by which they were specified;
- Completeness, the degree to which the data set reports every relevant feature present;
- Positional accuracy, which estimates how each location reported in the data set matches the true location in the real world; and
- Lineage, which contains details of the processes by which the data set was acquired and compiled.

A second version of the CSDGM standard was adopted in 1998 with a sixth optional component to allow cloud cover to be described for remotely sensed data sets.

ISO 19139 / 19115 Dataset

This metadata pair of standards will be discussed in the next paragraph since it is the standard adopted for the PPMS GeoDB.

ISO 19139 / 19119 Services

This pair of standards is used specifically for a web service, a software component accessible over the World Wide Web for use in other applications. Web services are independent on any particular operating system or programming language (e.g., Open GeoSpatial Consortium (OGC) Web Map Service). The presence of a *<SV_ServiceIdentification>* tag permits differentiation of an ISO 19119 Dataset document from an ISO 19115 Dataset document.

North American Profile (Dataset and Services)

This standard is based on the two previous standards, but implements additional mandatory values that are not mandatory in the original ISO ones. Before this standard, U.S. government agencies were mandated to apply the CSDGM for their metadata.

INSPIRE Metadata

The Spatial Information in European project (INSPIRE) has a metadata catalog in order to offer quality geographic information through a single access point for the European Community. The metadata are based on ISO 19115 (dataset) and ISO 19119 (services) standards with some extended elements and some different mandatory elements than the original standards.

Dublin Core

The Dublin Core Metadata Initiative (DCMI) developed an interoperable online metadata standard supporting a broad range of purposes and business models called the Dublin Core standard. This standard presents a Simple Level and a Qualified Level. The Simple Dublin Core describes 15 elements; the Qualified Dublin Core has some additional element and refinements. This standard has '*dc*' in the metadata tags (e.g. *<dc:title>*).

ISO 19139 / 19115

Technical Committee 211 of the International Standards Organization (ISO/TC 211) is responsible for the development of international standards in the area of digital Geographic Information. These standards are oriented to specify methods, tools, and services for data management, acquiring, processing, analyzing, accessing, presenting and transferring such data between different users, systems and locations.

In 2003, the ISO 19115 standard for geospatial metadata was adopted, with the possibility that some standards in member States could eventually be brought into compliance. This standard strongly follows the CSDGM standard, but with two main changes: temporal accuracy was added (previously this was part of completeness), and attribute accuracy was renamed thematic accuracy.

Since a primary goal in the management of metadata for geographic data is the ability to access the metadata and the related spatial data it describes, a software implementation is required to use common encoding methods to achieve operational use of the metadata for geographic data. To support this, the ISO 19139 standard is an implementation of the ISO 19115 standard. ISO 19139 standard defines the Geographic MetaData XML (gmd) encoding with the aim of providing common implementation guidelines by XML Schemas for describing, validating and exchanging metadata about geographic datasets.

The ISO 19115 standard defines what elements should be included for describing a geographic resource, and the ISO 19139 provides details. A document in ISO 19139 can be identified by looking for the *<MD_Metadata>* tag.

Before the publication of ISO 19139, some interpretations of the ISO 19115 schema were created, using different tags. An example is the ESRI-ISO that also includes some special *<ESRI>* tags for ESRI software.

The ISO 19139 standard defines:

- Mandatory and conditional metadata sections, metadata entities, and metadata elements;
- The minimum set of metadata required to serve the full range of metadata applications (e.g., determining data fitness for use, data transfer, etc.);
- Optional metadata elements; and
- A method for extending metadata to fit specialized needs.

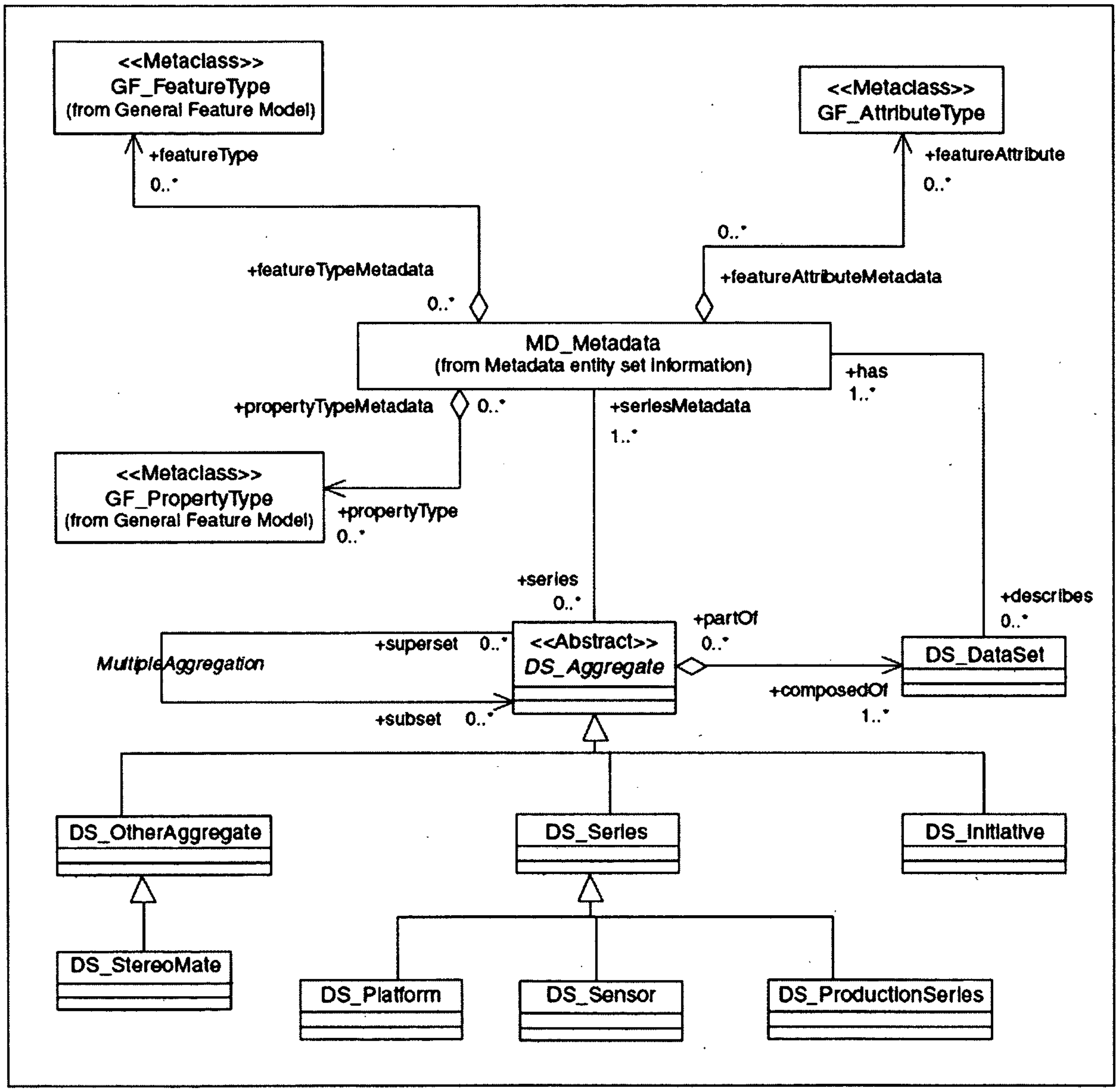


Figure 73 - Metadata application

This metadata standard is applicable to independent datasets, aggregations of datasets, individual geographic features, and the various classes of objects that compose a feature.

Metadata should be provided for geographic datasets and may, optionally, be provided for aggregations of datasets, features, and attributes of features. This standard is composed of one or more Metadata Sections (UML Packages) containing one or more Metadata Entities (UML classes).

In Figure 73, a UML class diagram defines the classes of geographic information to which metadata applies. It specifies that a dataset (*DS_DataSet*) must have one or more related Metadata entity sets (*MD_Metadata*). Metadata may optionally relate to a Feature, Feature Attribute, Feature Type, Feature Property Type (a Metaclass instantiated by Feature association role, Feature attribute type, and Feature operation), and aggregations of datasets (*DS_Aggregate*). Dataset aggregations may be specified (subclassed) as a general association (*DS_OtherAggregate*), a dataset series (*DS_Series*), or a special activity (*DS_Initiative*).

In other words, a dataset may be a member of a data series, and may be composed of a set of identified feature types and instances, and attribute types and instances (Figure 74). At first sight, it can appear that there are many levels of metadata to be maintained. In the majority of cases this is not so, since only metadata exceptions are defined at lower levels. If the metadata values are not changed, then the metadata is aggregated at a higher level. By creating several levels of abstraction, a linked hierarchy can assist in filtering or targeting user queries to the requested level of detail. When a lower level of metadata hierarchy is populated only the revised metadata values are recorded.

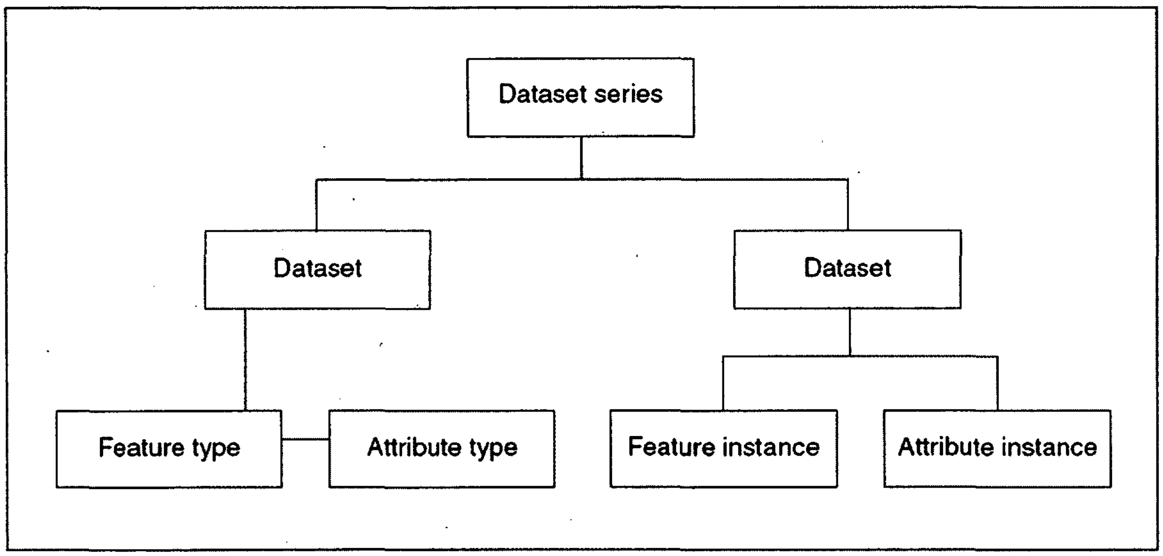


Figure 74- Metadata hierarchy

Metadata for geographic data are presented in UML Packages (Figure 75). Each package contains one or more entities (UML Classes), which can be specialized (sub-classed) or generalized (super-classed). Entities contain elements (UML class attributes) which identify the discrete units of metadata. Entities may be related to one or more other entities; they can also be aggregated and repeated as necessary.

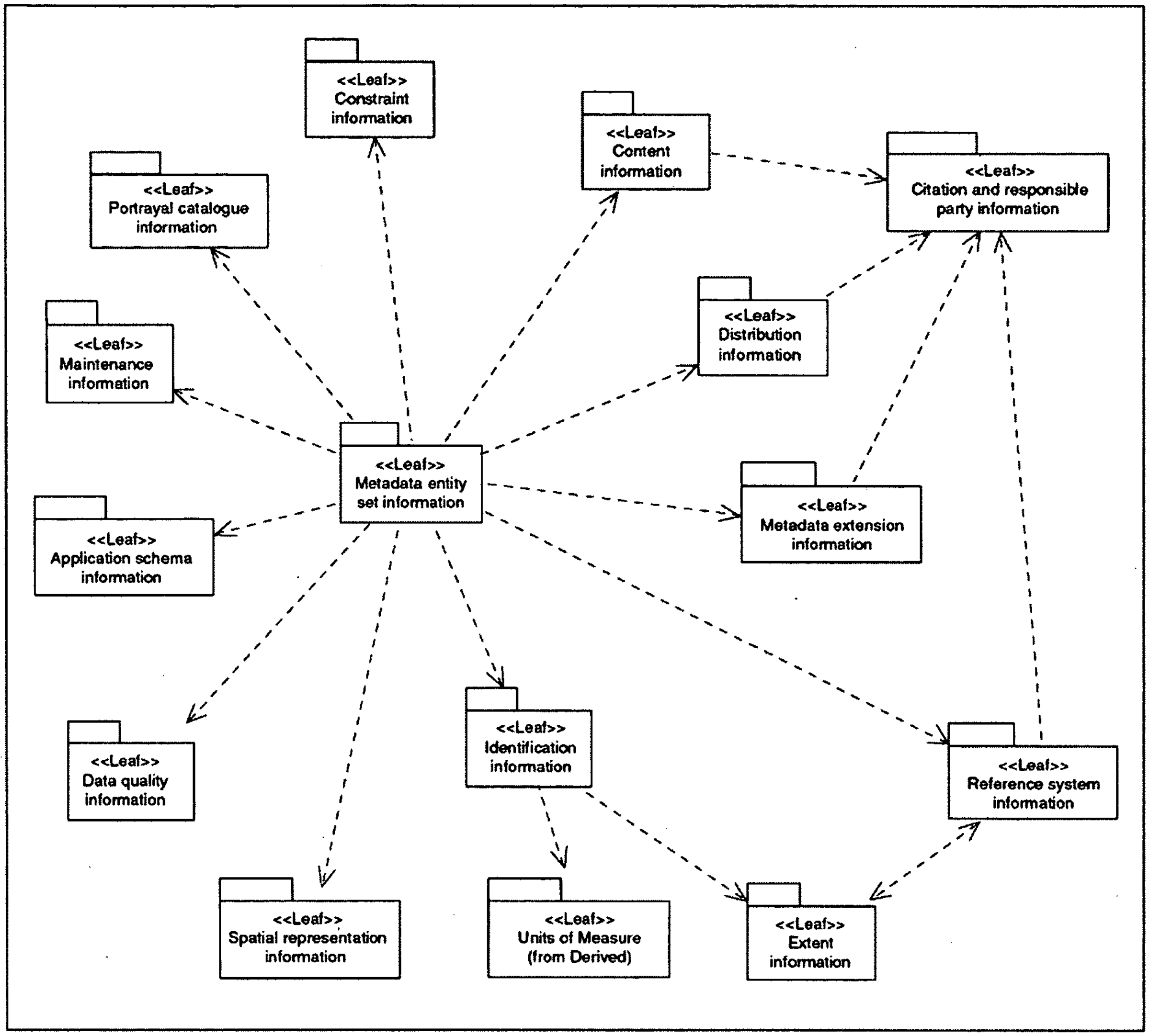


Figure 75 - Metadata packages

In Table 27, the relationships between packages of metadata and metadata entities are presented. The metadata standard is fully specified in UML model diagrams and a data dictionary for each package, which can be found in ISO 19115 Annexes A and B respectively.

Table 27 - Relationships between packages of metadata and metadata entities.

Package	Entity
Metadata entity set information	MD_Metadata
Identification information	MD_Identification
Constraint information	MD_Constraints
Data quality information	DQ_DataQuality
Maintenance information	MD_MaintenanceInformation
Spatial representation information	MD_SpatialRepresentation
Reference system information	MD_ReferenceSystem
Content information	MD_ContentInformation
Portrayal catalogue information	MD_PortrayalCatalogueReference
Distribution information	MD_Distribution
Metadata extension information	MD_MetadataExtensionInformation
Application schema information	MD_ApplicationSchemaInformation
Extent information	EX_Extent
Citation and responsible party information	CI_Citation CI_ResponsibleParty

Metadata entity set information consists of the mandatory entity *MD_Metadata*. This entity contains both mandatory and optional metadata elements (UML attributes).

Core Metadata for Geographic Datasets

Only a subset of the full number of elements defined in ISO 19115 is usually used. Since it is essential that a basic minimum number of metadata elements be maintained for a dataset, ISO 19115 identifies 22 core metadata elements required to identify a dataset, typically for catalogue purposes (Table 28).

The basic idea is to provide metadata elements answering the following questions:

- What? (Does a dataset on a specific topic exist?);
- Where? (For a specific place?);
- When? (For a specific date or period?); and
- Who? (A point of contact to learn more about or order the dataset).

Using the recommended optional elements in addition to the mandatory elements will increase interoperability, allowing users to understand without ambiguity the geographic data and the related metadata provided by either the producer or the distributor.

Table 28 - ISO 1915 Core Metadata: An "M" indicates that the element is mandatory; an "O" indicates that the element is optional; and a "C" indicates that the element is mandatory under certain conditions.

Metadata element	Content
Dataset title (M)	The name by which the cited resource is known
Dataset reference date (M)	A reference date for the cited resource
Dataset responsible party (O)	Name and position information for an individual or organization that is responsible for the resource
Geographic location of the dataset (C)	Geographic position or a description of the geographic area for the dataset
Dataset language (M)	Language(s) used within the dataset
Dataset character set (C)	Full name of the character coding standard used for the dataset
Dataset topic category (M)	Main theme(s) of the dataset
Spatial resolution of the dataset (O)	Level of detail (expressed as the scale of a comparable hardcopy map or chart) or the ground sample distance
Abstract describing the dataset (M)	Brief narrative summary of the content of the resource(s)
Distribution format (O)	Name and version of the data transfer format(s)
Additional extent information for the dataset (vertical and temporal) (O)	Temporal and vertical components of the extent of the referring object
Spatial representation type (O)	Method used to represent geographic information in the dataset
Reference system (O)	Information about the reference system
Lineage (O)	Non-quantitative quality information about the lineage of the data specified by the scope
On-line resource (O)	On-line information that can be used to contact the individual or organization
Metadata file identifier (O)	Unique identifier for this metadata file
Metadata standard name (O)	Name of the metadata standard (including profile name) used
Metadata standard version (O)	Version (profile) of the metadata standard used
Metadata language (C)	Language used for documenting metadata
Metadata character set (C)	Full name of the character coding standard used for the metadata set
Metadata point of contact (M)	Name and position information for an individual or organization that is responsible for the metadata
Metadata date stamp (M)	A reference date for the metadata

Metadata Extensions

Due to the diversity of data, the generic ISO metadata standard may not satisfy all applications. Annex F of ISO 19115 provides the rules for extending metadata by:

- Adding new metadata elements, entity, or sections;
- Creating (from a free text element) or expanding (if already existing) a codelist; or
- Imposing a more stringent domain or obligation on an existing metadata element.

When the information for a discipline or application field to be added is extensive, a creation of a community profile is recommended. A community profile should contain the core metadata components, either a part or all of the other metadata components, and extended metadata (Figure 76). The development of the S-1XX series of Product Specification represents the introduction of a hydrographic profile that will be followed by the PPMS GeoDB Product Specifications.

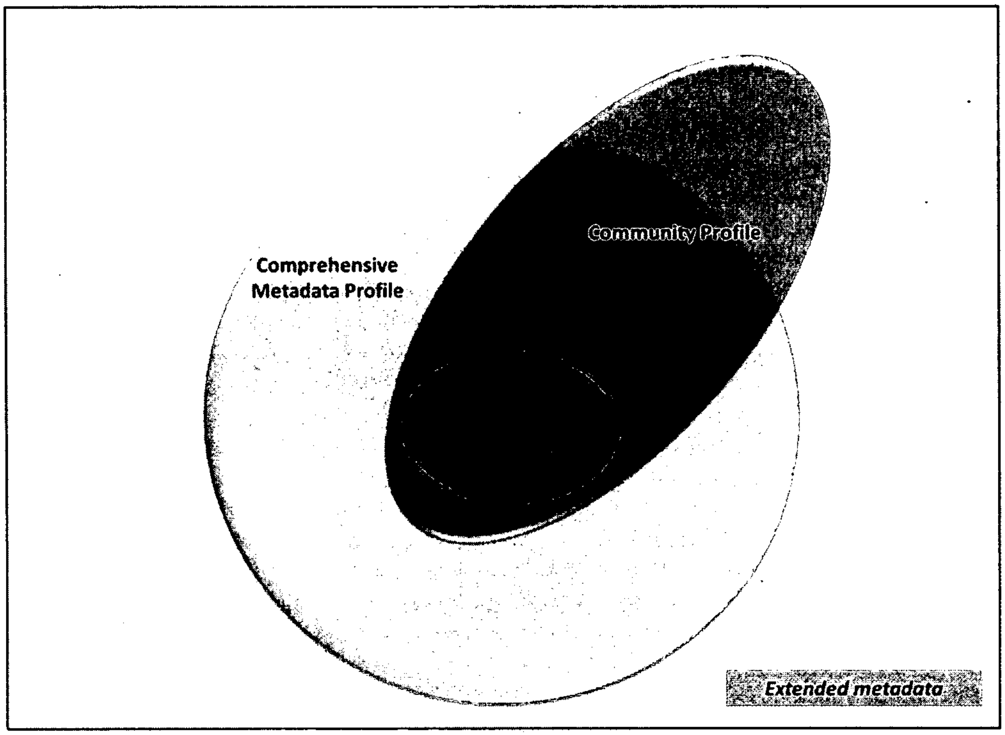


Figure 76 - Metadata components

Issues Related to the Application of the Current ISO standard

The geospatial metadata standards present several issues (Goodchild, 2008). The main problems are briefly presented in the following sub-paragraphs.

Decoupling

Paper maps are characterized by a scaling ratio, called the Representative Fraction (RF) and defined as the ratio of distances on the adopted representation to their corresponding distances in the real world. The Representative Fraction gives useful information about:

- The map's contents (e.g., maps with a small RF portray only the larger features of the Earth's surface);

- The spatial resolution (giving indications about the lower limit to the sizes of symbols that can be drawn and read on a map, and to the real-world features that can be portrayed);
- The positional accuracy, since national mapping agencies usually prescribe the positional accuracy of features on maps as functions of RF.

In the 1980s almost all geographic data sets had been created by digitizing or scanning paper documents. The digitizing process introduces errors and uncertainties added to those present in the paper document. Thus, the resulting positional errors in data sets derived by digitizing paper maps are themselves directly related to the RF of the original map. Since the 1990s, the coupling between content, spatial resolution, and positional accuracy under an RF value has broken down. For instance, data sets created from digital orthophotos never had an RF to inherit.

There are very few references about this issue in the ISO standard: RF is not mentioned in the ISO standard; both absolute and relative positional accuracies are not allowed for; spatial resolution is only mentioned as an optional element of "Core Metadata" (with no further details provided). In summary, the decoupled approach was not fully adopted in the ISO standard.

Uncertainty

After some efforts to apply the theory of errors to the compilation of digital representations of features (Goodchild and Gopal, 1989), since the 1990s it has been clear that there is a need to define better the data quality than simply the error and accuracy. In fact, for many types of geographic data it is very difficult to think of the data as observations reflecting some real-world truth modified by the process of measurement (e.g., vegetation-cover classification). At the same time, the classic probabilistic approach also presents some difficulties (i.e., define a probability $p_i(i^*)$ that the true class found at a point would be i if the observed class was i^*) since the definition of a class used in many mapping programs is usually open to varying interpretation. Some progress was made using concepts of vagueness and probability in theoretical frameworks such as fuzzy and rough sets (Fisher and Unwin, 2005).

The terms error and accuracy are now generally avoided in the research community, which tends instead to favor uncertainty, along with imprecision and vagueness. In the ISO standard, the term accuracy occurs 7 times, while uncertainty does not occur.

Separability

The ISO standard distinguishes between the accuracies of attributes and positions.

It is difficult to separate these two concepts in the case of geographic variation conceptualized as a continuous field (e.g., a continuous field of topographic elevation). Similar arguments exist also for area-class maps, where it is impossible to separate errors of boundary positioning from errors of class determination, except when boundaries follow well-defined features (e.g., roads or rivers).

Granularity

In the paper map world, a strong association exists between a map's contents and its marginalia. In the digital world, this concept is usually broken. The digitized contents of maps can be separated into layers, partitioned spatially or integrated into larger, apparently seamless data sets. It becomes difficult to define a single positional accuracy for a combination of two or more layers of data. In fact, each set of features may have very different data quality characteristics, difficult to capture in a single data quality statement prepared according to a metadata standard.

Today, integrated databases contain representations of many different types of features, linked by various types of relationship (specialization, association, aggregation, and composition). A partial solution is to use metadata to describe the database at the level of the class or some defined collections of features. But the definition of metadata at the level of the entire database is difficult, and there is also the need to describe the quality of class relationships.

Collection-level Metadata

Object-level metadata describe the properties of individual data sets and they focus on a single information object within the larger framework of an entire collection (e.g., geospatial data warehouses, digital libraries, or geo-libraries) containing potentially thousands of separate data sets.

Collection-level metadata (CLM) are defined as data about the contents of an entire collection, describing such characteristics as geographic and temporal coverage, the set of themes that dominate the collection, and the general level of data quality (Goodchild and Zhou, 2003). However, the task of describing collections is far more complex than the task of describing individual data sets. This type of metadata is not present in the ISO standard.

Auto-correlation

Tobler's First Law describes the tendency for nearby things to be more similar than distant ones (Sui, 2004; Tobler, 1970), and it was formalized by geo-statistics as the theory of regionalized variables. For instance, errors in elevation in a digital elevation model are strongly autocorrelated, such that nearby errors tend to be similar (Hunter and Goodchild, 1997).

Autocorrelation has many common causes since any geographic data set inherits errors and uncertainties from many parts of its compilation process (e.g., misregistration of an image affects the positions of all of the features extracted from that image).

In summary, it is known that the correlations or covariances of errors of attributes and positions are as important, if not more important, than their variances. Knowledge of covariance of errors is critical to any analysis of the propagation of uncertainties during manipulation of spatial data sets. Thus appropriately defined parameters should be an essential part of any attempt to describe data quality in metadata. The current ISO standard focuses entirely on marginal properties such as mean positional error.

Cross-correlation

The possibility to overlay various layers is one of the major advantages of a GIS. However, this possibility creates the problems of misfit that almost always occur. If the positional uncertainties in two layers are not perfectly correlated, the result of overlay will be a large number of small slivers. On the other hand if the two layers were both obtained from the same root, then uncertainties may be perfectly correlated and no misfits will occur.

Even if metadata well describe independently the uncertainties of different data sets, the results of overlay cannot be obtained from this information. In fact, misfit is a joint property of a pair of data sets, not a property of either of them.

Binary metadata has been proposed in order to describe the ability of two data sets to interoperate (Goodchild, 2008), since such metadata cannot be obtained from the separate metadata descriptions of each data set (although indications might be inferred by comparing lineage). The ISO standard approach is only unary.

APPENDIX E

THE UNIFIED MODELING LANGUAGE (UML) AND THE IHO S-100 PROFILE

The Unified Modeling Language (UML) is a standardized general-purpose modeling language mainly used for expressing and documenting designs. It became a standard for technical exchange of models after adoption by the Object Management Group (OMG) in November 1997.

UML is predominantly a visual language and largely self-documented. Basically, it consists of a number of graphical elements variously combined to form diagrams, according to several rules. Each diagram represents a particular view of a system, and a set of diagrams composes a UML model.

UML Graphical Elements

The main graphical elements are:

- Class;
- Package;
- Note; and
- Stereotype.

An object is defined as a *"discrete entity with a well-defined boundary and identity that encapsulates state and behavior"*, while a class is "the descriptor for a set of objects that share the same attributes, operations, methods, relationships and behavior" (Rumbaugh et al., 1999). Figure 77 presents an example of notation for a class:

- Name (`MultibeamEchosounder`);
- A list of attributes (e.g., `manufacturerName`, `modelName`, `serialNumber`); and
- A list of operations (e.g., `startAcquisition`, `stopAcquisition`, `setParameters`).

Additional information for attributes can be the type (e.g. string, floating-point number) and the default value. Sometimes, a note is used to add some explanatory text into a diagram to avoid ambiguities.

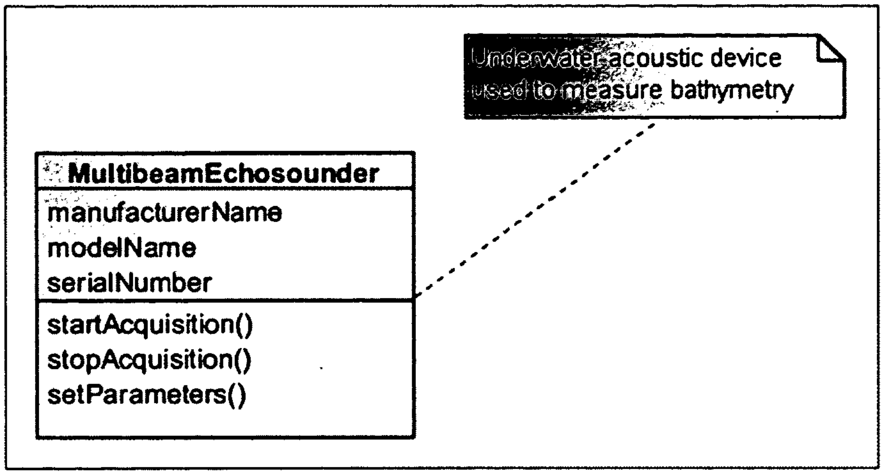


Figure 77 - Example of UML Class icon with explanatory note.

A package is designed to group elements that belong to a particular subsystem (Figure 78). It provides a namespace for the grouped elements (e.g., `SurveyTools::GPSReceiver`). A package can relate to another in three ways: generalization, dependency, and refinement. One package refines another if it has the same elements, but with more details.

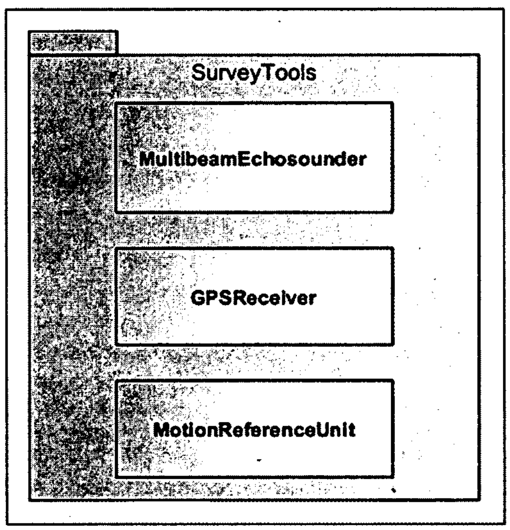


Figure 78 - Example of UML Package with Classes as children.

A stereotype is used to turn an existing UML element into a new one. In fact, it is a model element that is used to classify (or mark) other UML elements so that they in some respect behave as if they were instances of new virtual (or pseudo) metamodel classes whose form is based on existing base metamodel classes.

An extensive set of ready-made stereotypes are provided by UML. Some commonly-used instances of stereotypes are:

- «`Interface`» is a class with only operations (Figure 79);
- «`Import`» dependency adds the contents of the supplier to the client's namespace;
- «`Metaclass`» is a stereotype used for a class whose instances are also classes;

- «Type» is a class that specifies a domain of objects along with attributes, operations, and associations;
- «DataType» is a descriptor of a set of values that lack identity and whose operations do not have side effects. Datatypes include primitive pre-defined types and user-definable types. Pre-defined types include numbers, string, and time, while user-definable types include enumerations;
- «Enumeration» is a data type whose instances form a list of named literal values. Both the enumeration name and its literal values are declared. Enumeration means a short list of well-understood potential values within a class;
- «CodeList» used to describe a more open and flexible enumeration. Code lists are useful for expressing a long list of potential values. If the elements of the list are completely known, an enumeration should be used; if the only likely values of the elements are known, a code list is more appropriate;
- «Union» describes a selection of one of the specified types. This is useful to specify a set of alternative classes/types that can be used, without the need to create a common super-type/class;
- «Abstract» class (or other classifier) that cannot be directly instantiated. UML notation for this prescribes to show the name in italics.

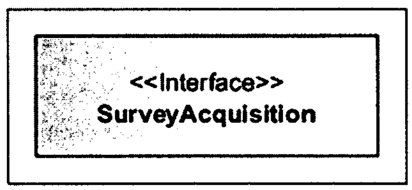


Figure 79 - Example of a stereotype applied to a class.

UML Diagrams

UML provides a large number of diagram types, showing system points of view for different stakeholders. The most common diagrams are:

- Class Diagram: its purpose is to describe how classes relate to one another. It may highlight a particular part of the model, or provide a general visualization;
- Package Diagram, used to explain the organization of packages and their elements.
- Object Diagram, similar to a Class Diagram but with class instances and specific values;
- Use Case Diagram, describing a system's behavior from the point of view of an user (called an 'actor');
- State Diagram, representing an object in particular states; and
- Sequence Diagram, showing the time-based dynamics of the interaction among objects.

These diagram types can be clustered as static, dynamic, and user's point of view.

UML Implementation of Object-Oriented Concepts

UML largely implements object-orientation concepts such as:

- Abstraction;
- Inheritance;
- Polymorphism;
- Encapsulation;
- Association; and
- Instantiation/Dependency.

Abstraction permits reduction of properties and operations of an element to only the ones necessary for a particular application and given users. For instance, a survey operator does not need the serial number of each MBES component, while they become fundamental for a technician in case of a failure.

Inheritance is used to point out those attributes and operations common to all the elements as part of a given super-class (generalization). The super-class is called the generalized class, while the sub-classes are specialized classes. An example could be the super-class `AcousticSystem` with attributes like `modelName` common to subclasses such as `MultibeamEchosounder` and `SideScanSonar`. Related to inheritance is the concept of visibility, used to specify which attributes and operations can be used only by the original class (private level), by each class that inherits from it (protected level), or by any class (public level).

Polymorphism is the ability of objects belonging to different UML elements to differently respond to an operation with the same name. For instance, the operation `setParameters()` has different meanings depending on the type of acoustic system.

Encapsulation permits hiding of operations of an object from the other objects. Coming back to the example of the survey operator, he does not need to know all the operations internal to the object of the `MultibeamEchosounder` class.

Association is commonly used to describe how classes are related to one another. Some of these associations are bidirectional, while others are unidirectional. Multiplicity provides the number of objects between two elements which can be related: e.g., many survey operators can use the same acquisition system, and a given survey operator can work with all the acquisition systems for which he was trained. If unspecified, the default multiplicity for associations is `0..*` (meaning from zero to an infinite number), and the default multiplicity for attributes is `1`. Any association is also characterized by a direction. If the direction is not specified, it is assumed to be a two-way association. If one-way associations are intended, the direction of the association can be marked by an arrow at the end of the line.

If an association is navigable in a particular direction, the model shall supply a 'role name' that is appropriate for the role of the target object in relation to the source object. Thus in a two-way association, two role names will be supplied. In Figure 80, role names between classes and cardinalities are represented as they are expressed in UML diagrams.

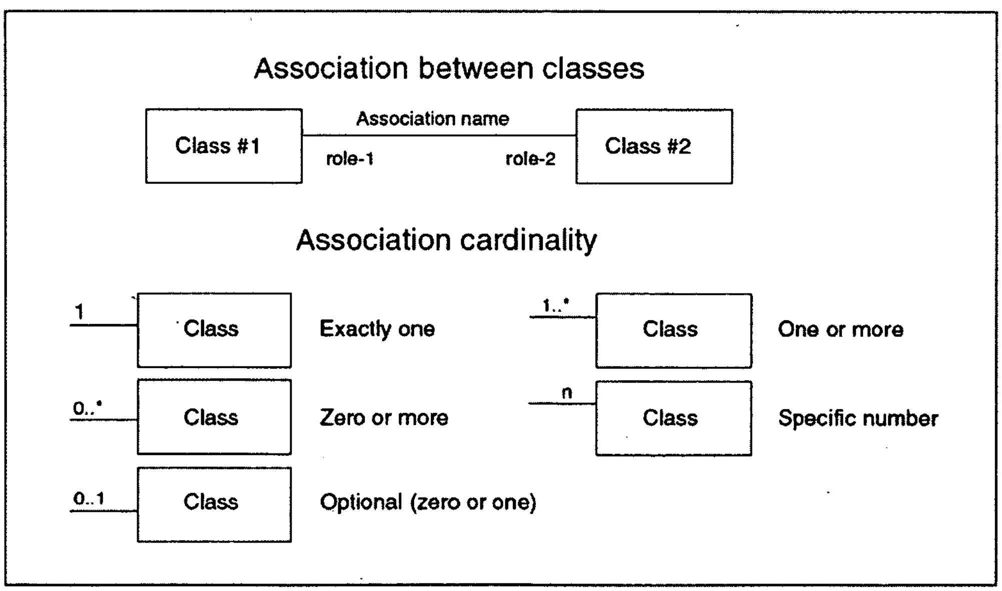


Figure 80 - Association between classes and association cardinality.

Aggregation is a particular type of association, describing the case when a class object is not only associated with another classes object, but is made from a number of them (part-whole association). For instance, the class `MultibeamEchosounder` can be described as an aggregation of `Transducer`, `ProcessingUnit`, and `OperatorStation` classes.

A composition association is a strong aggregation. In a composition association, if a container object is deleted, then all of its contained objects are deleted as well. The composition association shall be used when the objects representing the parts of a container object cannot exist without the container object (ISO, 2003).

A dependency is a different kind of relationship where one class uses another. A dependency relationship shows that the client class depends on the supplier class/interface to provide certain services, such as:

- Client class accesses a value (constant or variable) defined in the supplier class/interface;
- Operations of the client class invoke operations of the supplier class/interface;
- Operations of the client class have signatures whose return class or arguments are instances of the supplier class/interface.

An instantiated relationship represents the act of substituting actual values for the parameters of a parameterized class or parameterized class utility to create a specialized version of the more general item.

S-100 UML Profile

The Conceptual Schema Language adopted in IHO S-100 is composed by a subset of UML (a profile) and several basic data types.

The main characteristics of the S-100 UML profile are:

- It is based on static structure diagrams: Class Diagrams and Package Diagrams;
- Other UML diagram-types are used only informatively;
- All normative models shall contain complete definitions of attributes, associations, and data types; and
- Operations are not used.

Furthermore, the S-100 UML Profile provide several clauses about classes, attributes, types, relationship and associations, stereotypes, naming, notes, packages and documentation of models. The more important are here briefly summarized:

- A class is viewed in S-100 as a specification, not as an implementation;
- The use of multiple inheritances shall be minimized, since it increases model complexity;
- Since a one-way association is equivalent to an attribute definition, S-100 recommends the use of the association notation for all cases except for those involving attributes of basic data types; and
- Generalization, dependency and refinement are used according to the standard UML notation and usage.

The UML primitive types supported are: `Integer`, `PositiveInteger`, `NonNegativeInteger`, `Real`, `Boolean`, `CharacterString`, `Date`, `Time`, and `DateTime`. As complex types, the S-100 Profile provides: `UnlimitedInteger`, `Matrix`, `S100_Multiplicity`, `S100_NumericRange`, `S100_UnitOfMeasure`, `S100_Measure`, `S100_Length`, and `S100_Angle`. The stereotypes used in S-100 are: «`Interface`», «`Type`», «`Enumeration`», «`MetaClass`», and «`DataType`».

In addition to the UML diagrams, S-100 requires to document the semantics of the model (attributes, associations, operations and constraints) by context tables (Table 29). Thus, each class has a context table with the following columns:

1 Role Name: Class, Attribute, Association, Enumeration, or Literal;
2 Name (for an association, it is the role name used for the given class);
3 Description: the semantics of the property;

4 Multiplicity: the number of occurrence of the property in the class (indirectly, it describes if a property is mandatory or optional);
5 Data Type; and
6 Remarks: any additional information about the property as constraints or conditions.

For the documentation of enumerated types the Multiplicity and Data Type column are not used.

Table 29 - Example of context table.

Role Name	Name	Description	Multiplicity	Data Type	Remarks

APPENDIX F

THE EXTENSIBLE MARKUP LANGUAGE (XML)

Why Use a Markup Language?

All information must be encoded to be used in software applications. With binary encoding, information is stored in sequences of binary numbers. With text-based encoding, information is encoded in text form (even if the low level storage is still binary). Binary encoding has some advantages such as being compact and very efficient in data processing. Text-based encoding is flexible, extensible and easy to understand (reduced need of external documentation).

A markup language is a relatively modern system to annotate a document in such a way that a computer can distinguish them and use the information provided by these annotations to manipulate the normal text. It has advantages of binary file formats (easy to understand by a computer, compact, the ability to add metadata), as well as advantages of text files (universally interchangeable).

An early attempt to create a universal data format with rich information storage capabilities was the Standard Generalized Markup Language (SGML, ISO 8879:1986). Even if it is self-describing, SGML is such a complicated language that, for instance, it is not well suited for data interchange over the web.

A very well-known language based on the SGML work is the HyperText Markup Language (HTML). Although HTML has been incredibly successful, it is only intended for displaying documents in a browser. The tags it makes available do not provide any information about the content they encompass, only instructions about how to display that content.

The Extensible Markup Language (XML) represents a successful answer that was created to address these issues (Hunter et al., 2007).

Basic Characteristics of XML

XML is not really a language at all, but a standard for creating languages that meet the XML criteria. Thus, XML is a meta language that describes a syntax used to create other languages. It is a subset of SGML with which it shares the same goal (i.e., markup of any type of data), but with the complexity reduced as much as possible (W3C, 2008). As a direct consequence, any document that follows XML's syntax rules is by definition also following SGML's syntax rules, and can therefore be read by existing SGML tools.

XML encoding is a special case of text-based encoding that solves many limitations of arbitrary text-based encodings. It has a well-defined grammar, thus parsers and querying tools can be used to process XML data. It is a widely used standard with many available tools for editing, querying, transforming and presenting data. XML editors can check data integrity, verifying constraints on content and structure. XML encoding is extensible since it can be used to write other markup languages (e.g., Geography Markup Language). It is largely self-documenting.

In fact, it provides a formal schema language that is more or less human readable. Thus, programmers do not need to provide extensive external documentation comments to explain the adopted coding.

XML encoding has a natural hierarchical structure. Each element of this structure is enclosed in a pair of angle brackets (opening tag and closing tag). In the following example, an instance document for a reference encoding is shown:

```
<Book>
  <Title>Gone with the wind</Title>
  <Author>
    <First>Margaret</First>
    <Last>Mitchell</Last>
  </Author>
  <Year>1936</Year>
  <Publisher>Macmillan Publishers</Publisher>
  <ISBN>978-0446365383</ISBN>
</Book>
```

The main components of XML are:

- XML 1.0: the base recommendation upon which the XML family is built. It describes the syntax that XML documents have to follow, the rules that XML parsers have to follow, and anything else you need to know to read or write an XML document. It also defines document type definitions (DTDs), although they sometimes are treated as a separate technology;
- Namespaces: they provide a means to distinguish one XML vocabulary from another, which enables creation of richer documents by combining multiple vocabularies into one document type;
- XPath: describes a querying language for addressing parts of an XML document;
- XHTML: an XML version of HTML, the syntax used is more rigid and is readable by XML parsers (since XHTML is XML);
- XQuery Recommendation: designed to provide a means of querying data directly from XML documents on the web;
- Two basic means for traditional applications to interface with XML documents: a document object model (DOM) and the Simple API for XML (SAX); and
- The Scalable Vector Graphics (SVG) specification, used to describe two-dimensional graphics.

XML groups information in hierarchies: the items in an XML documents are related to each other in parent/child and sibling/sibling relationships. This structure is also called a tree with branches (any parts of the tree that contain children) and leaves (without children). The element that contains the entire XML document is called the root element.

One important characteristic of XML is being potentially self-describing since the tags may have essentially any meaningful name. However, the XML version of information has the relative disadvantage of being larger than a plain-text equivalent and much larger than binary versions. In fact, marking up data increases the information size, sometimes enormously, but achieving small file sizes is not one of the goals of XML (Hunter et al., 2007). In fact, one of the design goals for XML is that *"[t]erseness in XML markup is of minimal importance"* (W3C, 2008) since the characteristics oriented to be optimal as exchanged format are preferred to being a processing language.

The main reason for the success of XML is related to making easier writing software that accesses the information by a customizable data structure, while it is always possible to compress XML documents before sending them to address the size problem. Since it is platform and language independent, XML is a potential fit for an exchange format.

XML developers sacrifice the small file sizes of proprietary formats for the flexibility of universal data access. XML has no specific application; it is designed for whatever use you need. However, before XML can be considered useful, applications or libraries must be written to understand it. There are programs called XML parsers that can read XML syntax and extract the information. Several XML parsers are available:

- Microsoft's XML parser called MSXML;
- Apache Xerces which is a free XML parser in Java and C++;
- Expat is a free XML 1.0 parser toolkit written in C, etc.

These parsers can be used within a program in such a way that it will never have to look at the XML directly so that a large part of the workload is already done. Furthermore, the parser does not have to know anything about the application that is using it, nor about the types of XML documents the application works with. Basically, an XML parser just reads through the characters in the document, determining which characters are part of the document's markup and which are part of the data.

Because it is so flexible, XML is often targeted to be the basis for defining data exchange languages (e.g., it is widely used by Internet applications). XML is also largely extensible since it is possible to structure the same data in different ways that suit the requirements of an application or category of applications. It is also easy to convert between different vocabularies when required.

XML does not contain information at all about its visualization. XML styling is accomplished through another document dedicated to the task, called a stylesheet. Among the variety of languages that may be used, a popular one is the Extensible Stylesheet Language Transformations (XSLT). This technology could potentially make any kind of transformations very simple. For instance, XML can be transformed into HTML via XSLT, or displayed directly in browsers via CSS. This technology was used in the PPMS GeoDB Product Specification to create user-friendly products to access the Feature Catalogue.

Well-formed XML

Well-formedness is a property of an XML document that meets certain syntactical rules outlined in the XML 1.0 recommendation. If an XML document is well-formed, it can be more easily read by a computer program.

An XML tag is a text starting with a '<' character and ending with a '>' character. The tags are paired: any opening tag (e.g., `<name>`) must have a closing tag (`</name>`). In the XML world, these constructs are called start-tags and end-tags. An element is represented by all of the information from the beginning of a start-tag to the end of an end-tag, including everything in between. Specifically, the text between the start-tag and end-tag is called the element content. When an element is empty (e.g., `<Address2></Address2>`), it can be also written as `<Address2/>`.This tag is called a self-closing tag.

An XML document starts by providing the so called 'XML declaration' used to label documents as being XML and to provide to the parsers a few other pieces of information. For this reason, it must be right at the beginning of an XML file. The XML declaration starts with the characters '`<?xml`' and ends with the characters '`?>`'. A declaration must include the XML version, while the encoding and standalone attributes are optional (and must be in that order). The encoding allows specification for the XML parser what character encoding is used in the XML document. If no encoding is specified, UTF-8 or UTF-16 is assumed. Thus, if no encoding is specified and the document is not in one of the two Unicode encodings, the parser raises an error.

If the standalone attribute is included in the XML declaration, it must be set to either:

- "`yes`" to specify that the document exists entirely on its own (without depending on any other files);
- "`no`" to indicate that the document may depend on an external DTD.

In addition to these, XML documents must adhere to a certain number of other rules to be well formed:

- Tags cannot overlap (elements must be properly nested);
- XML documents can have only one root element;
- Element names must obey XML naming conventions (XML is case sensitive and will keep whitespace in text content), etc.

XML Naming Conventions

Although there are no reserved words, XML names can start with letters (including non-Latin characters) or the dash '-' character, but not numbers or other punctuation characters. There cannot be a space after the opening '<' character. After the first character, numbers, hyphens, and periods are allowed, but not spaces. Furthermore, names cannot contain the colon ':' character. Strictly speaking, this character is allowed, but the XML specification says that it is reserved for working with namespaces. Names cannot start with the letters `xml`, in uppercase, lowercase, or mixed.

In addition to tags and elements, XML documents can also include attributes as simple name/value pairs associated with an element. Attribute value may be expressed either in single quotes or double quotes (but they have to match), and they are attached to the start-tag (not to the end-tag):

```
<name nickname="Jonny">
  <first>John</first>
  <last>Doe</last>
</name>
```

The order in which attributes are included on an element is not considered significant. Therefore, if information in an XML document must appear in a certain order, they should be added as elements rather than attributes.

It is also possible to insert comments into an XML document as part of the text intended for people who are reading the XML markup itself. They start with the string '<!--' and end with the string '-->'.

XML Namespaces

XML namespaces are used to differentiate elements and attributes of different XML document types when combined into other documents. In other words, any element with the same prefix belongs to a namespace. Thus, a namespace is a purely abstract entity that represents a group of names that belong with each other conceptually.

The XML Namespaces Recommendation introduces a standard syntax for declaring namespaces and identifying the namespace for a given element or attribute in an XML document (e.g., `<ppms:shoreline xmlns:ppms=http://ccom.unh.edu/ppms>`). Qualified names (QName) consist of two parts:

- The local part (e.g. `shoreline`), which is the same as the names we have been giving elements all along;
- The namespace prefix (e.g., `ppms`), which specifies to which namespace this name belongs.

The namespace prefix, unless it is `xml` or `xmlns`, must have been declared in a namespace declaration attribute (e.g., `xmlns:ppms=http://ccom.unh.edu/ppms`) in either the start-tag of the element where the prefix is used or in an ancestor element. The scope of a namespace declaration declaring a prefix "extends from the beginning of the start-tag in which it appears to the end of the corresponding end-tag"(W3C, 2009).

When an XML document is parsed, the parser simply replaces any namespace prefixes (e.g. `ppms`) with the namespace itself (e.g. `http://ccom.unh.edu/ppms`). If the `xmlns` attribute does not specify any prefix name to use for a given namespace, it is intended to be used as a default namespace and it applies to all unprefixed element names within its scope. Furthermore, the prefixes `xml`, `xmlns` and all other prefixes beginning with the three-letter sequence `xml` are reserved, and thus users should not use them (W3C, 2009).

A given resource is uniquely identified by a string of characters called a URI (Uniform Resource Identifier), and it can be provided in the form of a URL (Uniform Resource Locator) or a URN (Universal Resource Name). The two dominant technologies to programmatically get information out of XML documents, the Document Object Model (DOM) and Simple API for XML (SAX), provide methods that enable, for a given QName, to get not only the namespace URI, but also the prefix (for those applications that need it). However, the XML Namespaces specification states *"it is not a goal that it [the namespace URI] be directly useable for retrieval of a schema (if any exists)"* (W3C, 2009).

Attributes usually do not have namespaces the way elements do. They are just 'associated' with the elements to which they belong. Attributes that are specifically declared to be in a namespace are called global attributes. A common example of a global attribute is the XHTML class attribute, which might be used on any XML element, XHTML or not. This makes things easier when using Cascading Style Sheets (CSS) to display an XML document.

Valid XML

In addition to being well-formed, an XML document may be valid. This means that:

- It contains a reference to another document, called a schema, that formally describes the grammar of an XML document type, and
- Its content matches the schema definition of allowable elements, attributes, and other grammar parts.

All the elements and attributes making up a document type are called the document's vocabulary. It is common that creators of XML documents included one or more schemas as a method for checking validity in the XML Recommendation. Even if several languages are available for creating schemas, Document Type Definitions (DTDs) and the XML Schema Document (XSD) languages, both from the W3C, are the ones more widely used together with the simpler RELAX NG.

Document Type Definitions (DTDs)

The grammar known as Document Type Definitions (DTDs) is described in the XML 1.0 Recommendation (W3C, 2008). It is the oldest existing schema language (it is included in the XML 1.0 standard) and inherited from SGML. DTDs provides a non-XML mechanism for defining the structure of an XML document.

With the extensive use of XML, DTDs revealed several limitations:

- The DTD syntax is totally different from the generic XML syntax (even if it was a benefit during initial migration from SGML to XML);
- Poor support for XML namespaces (e.g., namespace declarations are treated as attributes);
- Only support for rudimentary datatypes;
- Limited description for the content model (Lee and Chu, 2000);
- Lack of readability, etc.

However, DTDs are still part of the XML Recommendation, and they will be used in many diverse situations for a long time, even if other grammar-description methods emerge.

XML Schema Document (XSD)

In 1999, the W3C started the development of a different approach to validate XML documents that could go beyond the DTD limitations. In 2001, XML Schema Document (XSD) was published as one of several XML schema language, combining together features from DTDs and other previous attempts at XML Schema languages (W3C, 2004). Today, it is a mature technology that is widely adopted.

An XML Schema is much more complex than a DTD, but it is written in XML and it uses the same grammar. Almost all current XML languages, including GML, are written using XML Schema. An XML Schema provides elements that define other elements.

```
<element name="Author">
<complexType>
  <sequence>
    <element name="First" type="string"/>
    <element name="Last" type="string"/>
  <sequence>
</complexType>
</element>
```

The XML Schema language provides an extensive set of built-in data types and, at the same time, mechanisms to build additional data types from the existing built-in ones.

Since it uses an XML-based format, XSDs can be processed using ordinary XML tools as well, so users do not have to learn a new proprietary syntax (Lee and Chu, 2000). The most relevant features provided by XSD compared to DTDs are (Walmsley, 2002):

- Namespace awareness (it is possible to define an element by reusing element names defined elsewhere);
- Include and import clauses (in case of a large schema, it allows more modular schema definitions for better readability and maintenance);
- Better datatype support (e.g., user-defined type);
- Inheritance (by extending or restricting a base type).

In XML Schema, datatypes are categorized as simple (e.g., strings, dates, etc.) or complex. The main difference is that only a complex type can have element content or carry attributes. Attributes always have simple types.

The schema language specification supports a certain number of built-in types (e.g. `xsd:string`, `xsd:decimal`) that cover most of the types being used in general programming languages (while DTDs only allow the ten XML-related primitive types). New simple types can be created by deriving from built-in or derived types via inheritance (Lee and Chu, 2000). Furthermore, XSD supports a multitude of constructs (e.g., range, mask) to limit the valid domain for datatypes.

The behavior of inheritance is different for simple and complex types. For simple types, inheritance is only allowed by restriction (e.g., enumeration). For complex types, XSD supports inheritance both by extension and by restriction. In inheritance by extension, newly added elements are always appended at the end, while inheritance by restriction requires to have at least one sub-element from the original type.

XSD provides `documentation` and `appinfo` elements in order to support description for both human readers and application programs.

Some limitations of XSD 1.0 are (Bradley, 2004):

- The W3C Recommendation itself is not easy reading;
- Absence of dynamic constraints (e.g., the Schematron language allows for selectively turning on and off constraints) (ISO, 2006);

- It is not possible to have different versions of the same attribute of an element (XSD only provides a construct 'version' for schema definition);
- The declaration of a global element also enables the element to appear at the top-level of an instance document (there is no simple way to avoid this).

Recently (April 2012) XSD 1.1 became a W3C Recommendation (W3C, 2012). This new version focuses on simplicity of design (without loss of backward compatibility) and introduces several significant new features as:

- Use of XPath expressions to define assertions against the document content;
- Wildcard for elements and attributes that apply to all types in the schema;
- Conditional type assignment (the type against which an element is validated can be based on the values of the element's attributes), etc.

APPENDIX G

THE GEOGRAPHY MARKUP LANGUAGE (GML)

Background

Geography Markup Language (GML) is a mark-up language that is used to describe geographic objects in the world around us (Lake, 2004). It is built on the Extensible Mark-up Language (XML) and was developed by the Open GIS Consortium (OGC), an international industry consortium with the mission to deliver spatial interface specifications that are openly available for global use. Since it is accepted by a mass of industrial companies and research institutions, GML acts as a *de facto* standard in spatial data processing and exchange.

In 2002, the OGC approached the International Standard Organization (ISO) about making GML an ISO standard. Although the joint ISO/OGC project team added some additional components, ISO 19136 did not significantly alter the GML 3.0 specification.

The remaining part of this Appendix is divided in two parts:

- The first one presents an overview about GML rules and philosophy;
- The last part is focused on ISO 19136 since it is adopted as an example of an encoding format in the PPMS GeoDB Product Specification.

GML Overview

Relations with the XML World

Since GML is a geographic dialect of XML, it has many relations with the XML world and related languages. For instance, XSD has been used to create GML schemas in all version of GML since version 2.0. Furthermore, database administrators can create additional domain-specific GML application schemas by building on the GML 3.0 core schemas. At the same time, GML primarily uses XML technologies such as XLink, XSLT and SVG.

XLink is an XML Linking Language that permits creation of associations and relationships between multiple resources (XML elements). These associations may be navigated in more than one direction. XPointer provides XLink with a means of pointing at XML resources. The combination of XLink and XPointer permits expression of abstract links between geographic features and other GML objects, such as geometries and topologies.

The Extensible Stylesheet Language Transformations (XSLT) is a language for transforming XML from one XML grammar to another or to non-XML languages like HTML. XSLT can perform many useful geospatial tasks with GML data such as mapping schemas, transforming coordinates, generating graphical maps, etc.

Scalable Vector Graphic (SVG) is an XML language for encoding two-dimensional graphical drawings. It can be used from simple geometric shapes to complex maps with annotation and animation. In addition, SVG supports raster graphics such as GIF and JPEG.

While GML describes the content of geographic features, SVG provides the means for graphical representation.

Core and Application Schemas

GML Application Schemas are essentially vocabularies that define the structure of GML data for a given application. They are both human and machine-readable.

A GML Application Schema is based on one or many GML core schemas. In GML 2.0 there are only 3 Core Schemas (Feature, Geometry and XLinks), while GML 3.0 provides 25 additional Core Schemas. All together they provide a large and flexible framework to create GML Application Schemas for specific domains.

In the creation of an application schema, some rules must be followed, such as for instance the import of required GML core schemas. Data modelers usually need to be familiar with only a subset of all the available Core Schemas. A big issue of this application specific approach is that the same object may be classified as a different feature in different domains. Furthermore, even if a common vocabulary is established, different people may disagree about the classification. The results is that same feature names may have a different meaning, depending on the domain, while different feature names may have the same meaning in different schemas. On the basis of this list of issues, it should be clear that the process of creating GML application schemas is inexact (Lake, 2004).

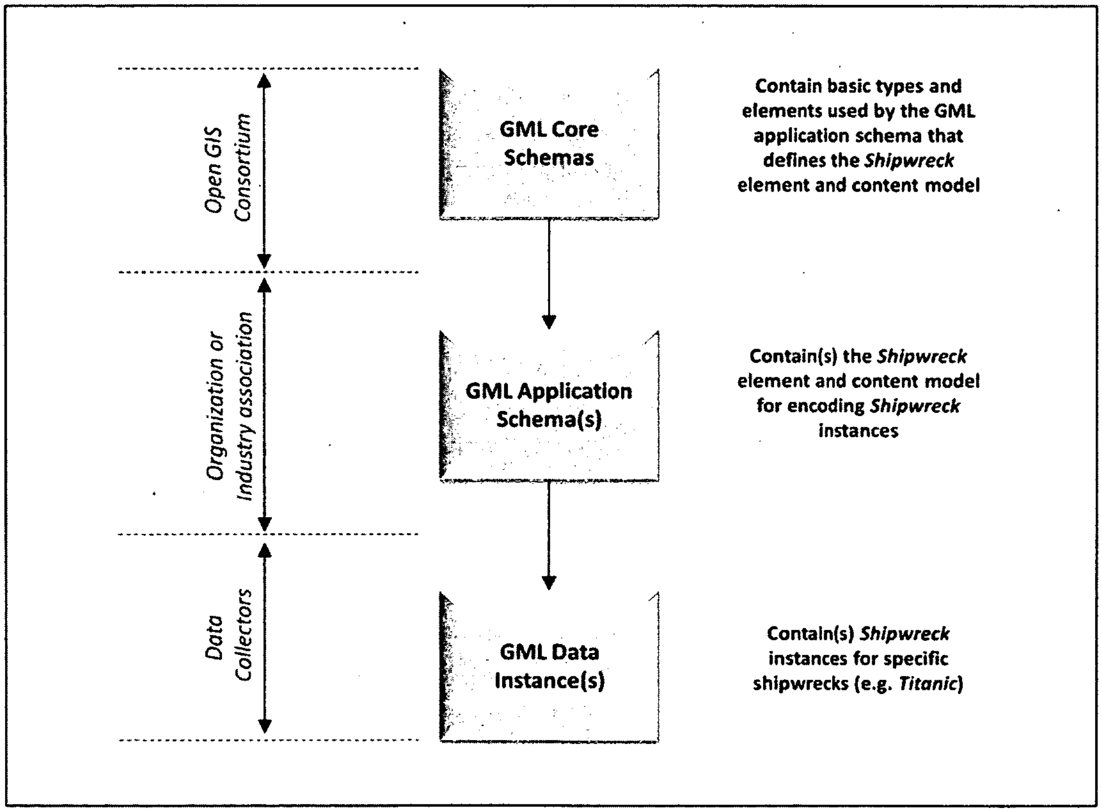


Figure 81 - Relationships between GML data instances, application schemas and core schemas.

GML schemas – based on XSD – describe the structure of GML data, defining elements and attributes used in data instances, while GML instances are files or part of files containing the actual geographic data encoded in GML. The latest OGC GML specification has 28 kernel schemas, including basic GML, geometry, topology, coordinate reference system, etc. GML data belonging to specific geographical spaces is described using a pre-defined application schema based on these kernel schemas (Wang et al., 2011).

In every domain that uses GML to encode geospatial data, domain-specific GML schemas and data instances are created. For instance, to encode data about a shipwreck, a `Shipwreck` instance based on a GML application schema that defines a `Shipwreck` element and its content must be created (Figure 81). For any new feature created, the content model must inherit (directly or indirectly) from `AbstractFeatureType`. A specific feature type is not defined in GML, but in Application Schemas usually created by database administrators. The `Shipwreck` element, as well as all other elements related to the PPMS GeoDB world, are not directly defined in the GML Core Schemas but are introduced through dedicated PPMS-oriented Application Schemas.

<u>Data Model</u>

In GML, real-world objects are called features categorized into particular types. They can be concrete, like a street, or abstract, such as political boundaries. A GML file has a tree structure like any XML file. Each feature may have child nodes, which may be spatial or non-spatial. Every instance must have its own unique identifier, usually based on existing domain standards.

For a shipwreck, a possible instance could be:

```
<Shipwreck gml:id="00001">
  ...
</Shipwreck>
```

Each feature is described using properties. GML properties can be clustered in geometric properties (e.g., location, form, extent) and non-geometric properties (e.g., color, name, etc.). The GML naming convention has the first letter in upper case for features and lower case for properties. For both features and properties, all subsequent words start with an uppercase letter.

A schema fragment for the `Shipwreck` element and content model could be:

```
<element name="Shipwreck" type="ShipwreckType"
  substitutionGroup="gml:_Feature"/>

<complexType name="ShipwreckType">
  <complexContent>
    <extension base="gml:AbstractFeatureType">
          <sequence>
                <element name="vesselName" type="string"/>
          </sequence>
    </extension>
  </complexContent>
</complexType>
```

In the first part, there is the element declaration with the name and its type, while the attribute `substitutionGroup` will be discussed later. The second part represents the type definition with content and structure of the feature.

The `ShipwreckType` extends `AbstractFeatureType`, which is an abstract type for creating new features in GML. Abstract elements and types cannot be included in GML instances, but only used to define new elements and types in GML Core and Application Schemas.

Namespaces

GML Application Schemas always have an associated namespace. A GML namespace is a concept directly coming from XML. It is used to assign unique identities to sets of elements and attributes. In GML, and more generally in XML, namespaces are based on the W3C Namespaces specification.

A namespace is essentially an Internet address that is represented by a prefix placed before a feature or a property to make it globally unique. The `gml` prefix refers directly to the GML Core Schemas and the GML namespace that is `http://www.opengis.net/gml`.

Different organizations can create their own schemas and associate them with a namespace for which they are responsible. GML application schemas can be shared on the Internet and can import or include one another.

For instance, `ppms` may not be the only domain that might define a `Shipwreck` element in a GML application schema (e.g., the `hydro` namespace is another likely source). By placing a prefix (e.g., `ppms`) in front of the `Shipwreck` element, it should become easy to distinguish it from other elements with the same name, but in different domains.

Feature Relationships

In the real world, objects are related to other objects. In the same way, a GML feature can have relationships with other GML features.

Simple relationships are expressed with GML properties where the property name designates the role of the target participant with respect to the source in the relationship.

For instance, if a bridge spans a city, it is possible to write:

```
<app:Bridge gml:id="b1">

      <app:spans>

            <app:City gml:id= "c1"/>

      </app:spans>

      <app:height>50</app:height>

</app:Bridge>
```

In this case, we have the `height` property of the bridge, but also the `spans` property used to indicate the relationship between the `Bridge` and the `City` features.

Remote properties are usually used to write relationships between features. For instance, the above relationship could be expressed in a GML document using different relationships for the two elements:

```
<app:Bridge gml:id="b1">

      <app:spans xlink:href="#c1"/>

      <app:height>50</app:height>
```

```
</app:Bridge>

      ...

<app:City gml:id="c1">

      <app:spanned xlink:href="#b1"/>

</app:City>
```

Geometries

GML provides a certain number of geometry elements in order to describe the geometric aspects of a feature. Some features only have one geometry-valued property, while others have several geometry-valued properties to describe different aspects of a feature. In the case of a `Shipwreck`, it can have a `position` property to describe its location on the Earth's surface and a `shape` property for the actual site definition of the vessel.

Geometries can be stand-alone objects and features are referenced to them by properties. In this way, the same geometry can be shared between different GML features.

In GML 2, there are only linear geometries (Figure 82). Many others have been added to the GML 3.0 Geometry schemas: `Curve`, `Surface`, `Solid`, etc.

Furthermore, GML provides several convenience geometry-valued properties, such as `position`, `location`, etc.

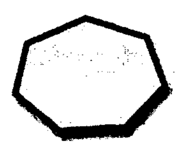
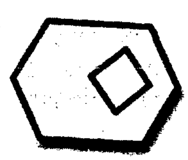
Geometries	Example
`Point`	
`LineString`	
`Polygon`	
`MultiPolygon`	

Figure 82 - GML geometries (GML 2.0).

Internet

GML plays several roles in relation to geospatial web services since it expresses geographic information in an easy way that supports sharing on the Internet.

A web service is an application on a network that accepts and processes requests from other applications present on the same network. It is an exchange of messages between the service and its consumer clients. The communication is between programs (client-server) and not directed to humans (as in a website). Web services rely on standardized web protocols such as the HyperText Transfer Protocol (HTTP). Furthermore, almost all the requests sent to a web service are formatted in XML.

A geospatial web service works with geographic information in order to:

- Provide access to geographic information stored in a database;
- Perform simple (e.g., calculate area) and complex (based on temporal and spatial distribution) geographic computations;
- Return messages that contain geographic information (such as text, numbers, or geographic features).

GML is commonly used in the request and response messages that are sent to and from the service. For instance, GML is used in both the request and response messages of the Web Feature Service (WFS) developed by the Open GIS Consortium.

ISO 19136:2007

Introduction

ISO 19136 is the first version of GML to become an International Standard, and it describes version 3.2.1 of GML. This standard clearly states that GML is *"[..] an XML grammar written in XML Schema for the description of application schemas as well as the transport and storage of geographic information"* (ISO, 2007).

It provides an open, vendor-neutral framework for the description of geographical application schemas for the transport and storage of geographic information in XML, leaving to the implementers to decide whether *"to store directly geographic information in GML, or to use GML only for schema and data transport"* (ISO, 2007).

All components of the GML schema are defined in the namespace with the identifier "`http://www.opengis.net/gml/3.2`", for which the prefix `gml` or the default namespace is used within the ISO 19136 standard.

ISO GML Terms and Concepts

ISO 19136 provides many useful definitions that are applied in a GML application. These term definitions are adopted in the PPMS GeoDB PS. The most important are summarized in Table 30.

Table 30 - PPMS GeoDB GML Dictionary based on ISO 19136

Term	Definition
Geographic Feature	A geographic feature is an abstraction of real world phenomena associated with a location relative to the Earth. The feature type determines the number of properties it may have, together with their names and types. A feature may occur as a type or an instance.
Geographic Properties	A set of properties defines the state of a feature. Each property may be thought of as a triple composed by a name, a type, and a value.
Feature Collection	A collection of features that may itself be regarded as a feature. Thus, a feature collection has a feature type and may have distinct properties of its own, in addition to the features it contains.
Coverage	A subtype of feature that has a coverage function with a spatiotemporal domain and a value set range of homogeneous 1- to n-dimensional tuples. Coverage *"acts as a function to return values from its range for any direct position within its spatiotemporal domain"* (ISO 19123).
Observation	A subtype of feature with a time at which the observation took place, and a value for the observation.
Reference System	Provides a scale of measurement for assigning values to a position, time or other descriptive quantity or quality. A coordinate reference system consists of a set of coordinate system axes that is related to the Earth through a datum that defines the size and shape of the Earth. A coordinate system is a set of mathematical rules for specifying how coordinates are to be assigned to points. A coordinate is one of a sequence of numbers designating the position of a point in n-dimensional space (ISO 19111:2007). A temporal reference system provides standard units for measuring time and describing temporal length or duration. A reference system dictionary provides definitions of reference systems used in spatial or temporal geometries.
Spatial Geometries	The values of spatial feature properties. They indicate the coordinate reference system in which their measurements have been made. The "parent" geometry element of a geometric complex or geometric aggregate makes this indication for its constituent geometries.
Temporal Geometries	The values of temporal feature properties. Like their spatial counterparts, temporal geometries indicate the temporal reference system in which their measurements have been made.
Spatial and Temporal Topologies	Used to express the different topological relationships between features.
Unit of Measure Dictionary	Provides definitions of numerical measures of physical quantities, such as length, temperature and pressure, and of conversions between units.

ISO GML Application Schema and Conceptual Models

An application schema is a conceptual schema for data required by one or more applications (ISO 19101:2002). A GML application schema captures the feature types of an application domain and it is specified in XSD. In ISO 19136, all application schemas are expected to be modeled in accordance with the General Feature Model (as specified in ISO 19109). GML specifies standardized XML encodings of several of the conceptual classes defined in the ISO 19100 series:

- Unit of measure, and basic types from ISO/TS 19103;
- Geometry, and topology objects from ISO 19107;
- Temporal geometry and topology objects, and temporal reference systems from ISO 19108;
- Features from ISO 10109;
- Coordinate reference systems from ISO 19111; and
- Coverage geometry from ISO 19123.

ISO 19136 supports two ways to construct XML Schemas:

- Defining the application schema in UML using the rules specified in ISO 19109 for application schemas, and mapping it to a GML application schema as specified in ISO 19136;
- Writing directly in XSD using the rules for GML application schemas specified in ISO 19136.

Both ways are valid approaches. The PPMS GeoDB application followed the former.

The GML schema consists of W3C XML Schema components that define types and declare:

- XML elements to encode GML objects with identity,
- XML elements to encode GML properties of those objects, and
- XML attributes qualifying those properties.

A GML object is an XML element of a type derived directly or indirectly from `gml:AbstractGMLType`. From this derivation, a GML object shall have a `gml:id` attribute.

A GML property shall not be derived from `gml:AbstractGMLType`, shall not have a `gml:id` attribute, or any other attribute of XML type ID. An element is a GML property if and only if it is a child element of a GML object.

A GML object shall not appear as the immediate child of a GML object. Consequently, no element may be both a GML object and a GML property.

All XML attributes declared in the GML schema are defined without namespace, the only exception is the `gml:id` XML attribute.

ISO GML Instance Documents

A feature is encoded as an XML element with the name of the feature type. Each feature attribute and feature association role is a property of a feature. Feature properties are encoded in an XML element.

A property element may contain its value as content encoded inline, or reference its value with a simple XLink. The value of a property may be simple, or it may be a feature or other complex object. When recorded inline, the value of a simple property is recorded as a literal value with no embedded markup (text), while if the value is complex it appears as a sub-tree using XML markup.

The 'object-property model' is used in the GML encoding model since the first version was adopted by OGC. While in some cases this encoding pattern adds extra levels of elements in instance documents it also provides significant benefits: it helps to make a GML instance document understandable on its own, provides a predictable structure and avoids too heavy reliance on XSD.

Hierarchy of ISO GML Object Classes

In ISO 19136 there are more than feature objects; Figure 83 shows the hierarchy of the GML object classes and abstract element for each class.

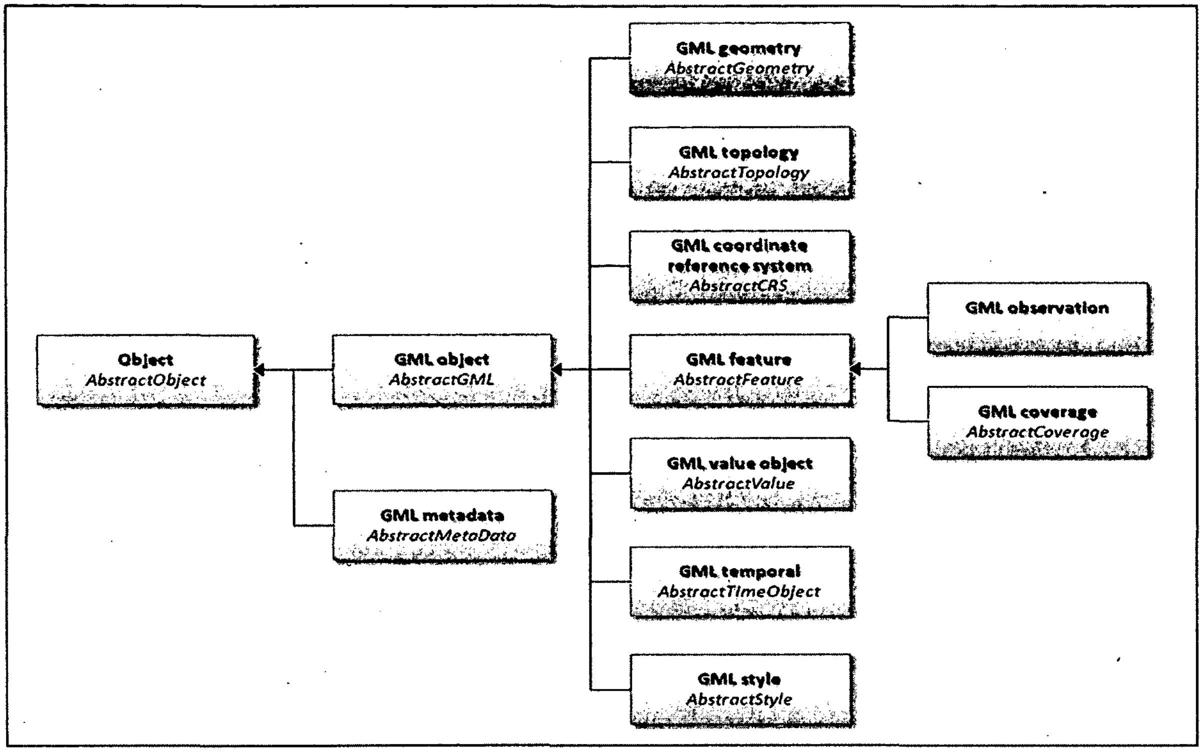


Figure 83 - Hierarchy of GML object classes

The `AbstractObject` has no type defined (thus it is implicitly an XML Schema `anyType`). It is at the top of the XML Schema substitution group, unifying different contents used for datatypes in GML. The `AbstractGML` is any GML object having identity, an abstract superclass for all GML objects. Its complex type is `AbstractGMLType` that has a required attribute `gml:id` (to provide a unique identifier in the XML document within which it occurs) and a group of `StandardObjectProperties`.

In general, if a content model derives by extension from another content model, the new content model also inherits all of the elements in the original content model in addition to the defined new elements. If it derives by restriction, it is necessary to redefine all of the elements that the content model inherits from the original. Any object has to respect the rules of the base types from which it derives.

A GML schema parser must be able to read a GML Application Schema and determine the kind of GML object for every object declared in the schema.

ISO GML Objects, Features, Features Collections and Related Properties

A GML object is an XML element of a type derived from `AbstractGMLType`. A GML property is a child element of a GML object. It corresponds to the feature attribute and feature association roles in ISO 19109. If a GML property of a feature has an `xlink:href` attribute that references a feature, the property represents a feature association role.

A GML feature is a representation of an identifiable real-world object in a selected domain of discourse. Thus, anything that is a GML feature derives from `AbstractFeature`, which inherits properties and attributes from `AbstractGML`. In addition, the complex `AbstractFeatureType` has two specific optional properties suitable for geographic features: the `boundedBy` property describes an envelope enclosing the entire feature instance; and the `location` property describes the extent, position or relative location of the feature.

A new feature may be defined by its declaration as a global element. The content model of this new feature must derive, directly or indirectly, from `AbstractFeatureType` that derives from `AbstractGMLType`.

```
<element name="Shipwreck" type="ppms:ShipwreckType"
  substitutionGroup="gml:AbstractFeature"/>

<complexType name="ShipwreckType">
  <complexContent>
        <extension base="gml:AbstractFeatureType">
              <sequence>
                    <element name="..." type="..."/>
                    <element name="..." type="..."/>
                    ...
              </sequence>
        </extension>
  </complexContent>
</complexType>
```

The `substitutionGroup="gml:AbstractFeature"` allows the inclusion of the `Shipwreck` in an element that has `AbstractFeature` as a value (e.g., the `featureMember` property has an `AbstractFeature` element as its only value). Since the `ShipwreckType` is derived for extension from `AbstractFeatureType`, all GML-aware software can detect that it is a feature.

In ISO 19136 GML, a feature collection is a collection of GML feature instances. It can be any GML feature with a property element derived by extension from `AbstractFeatureMemberType`.

By default, this abstract property type does not imply any ownership of the features in the collection. The `OwnershipAttributeGroup` attribute may be used to assert ownership of a feature in the collection. A collection shall not own a feature already owned by another object.

In addition, the complex type describing the content model of the GML feature collection may also include a reference to the attribute group `AggregationAttributeGroup` to provide additional information about the semantics of the object collection. Derivation permits user-added content by extension or restriction.

In ISO 19136, the Dynamic Feature Core Schema also defines a certain number of types and relationships to represent the time-varying properties of geographic features. The states of a feature are captured by time-stamped instances.

The content model of a `DynamicFeature` object extends the `AbstractFeatureType` with the `dynamicProperties` group. They are representations of the state of a feature at a point in time or during some interval. Since snapshots include also time-invariant properties, successive snapshots may contain redundant information. More generally, the `history` approach can be viewed as event-oriented, whereas a snapshot is more state-oriented, in that it provides a snapshot showing the status of the whole.

It is also possible to create user-defined history properties and time slices. This approach is used in the PPMS GeoDB to describe the dynamic characteristics of `AbstractPPMS` Class.

ISO Units of measure

In ISO GML, the `MeasureType` supports recording an amount encoded as a value of XML Schema `double`, together with a unit of measure indicated by an attribute `uom`, short for 'units of measure'. The value of the `uom` attribute identifies a reference system for the amount, usually a ratio or interval scale.

Lexical Conventions

There are several lexical conventions used to assist in human comprehension of GML instances and schemas:

- Objects are instantiated as XML elements in UpperCamelCase;
- Properties are instantiated as XML elements is in lowerCamelCase;
- Abstract elements have a prefix `Abstract` (objects) or `abstract` (properties) prepended to their name;
- The names of XML Schema complex types are in UpperCamelCase ending in the word `Type`;
- Abstract XML Schema complex types have the word `Abstract` prepended.

APPENDIX H

BUILDING A GML APPLICATION FOR THE PPMS GEODB

A number of steps were followed in the creation of 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; and
- Define Metadata Schemas as a function of the schema-defined objects.

Due to the complexity and the number of objects involved in this application, several Application Schemas have been created:

- *root.xsd*, containing the root element and the main Feature Collections which can be present in an instance document;
- *ppms.xsd*, used to describe different kinds of Potential Polluting Marine Sites;
- *resources.xsd*, for different types of resources that can be threatened by a PPMS;
- *complementaryinfo.xsd*, for various types of information useful in PPMS management;
- *ppmsgeodb.xsd*, a wrapper to include all the above Application Schemas.

Declaration of a target namespace

In general, every Application Schema must declare a target namespace in which the schema types are located (see Appendix G for more details). The identifier is a Uniform Resource Identifier (URI) that is controlled by the Application Schema owner's organization. For the Schemas of this Application, the common declared `targetNamespace` attribute is `"http://www.ccom.unh.edu/ppms"`.

The same `schema` tag is used to:

- Declare XML namespace prefixes for the target namespace itself (`ppms`) and for other namespaces referenced in the schema (`xsd`, `gml`, etc.);
- Define the qualified value for the `elementFormDefault` attribute (a GML requirement is that each element must be associated with a namespace, either by the use of a declared prefix or via a default namespace declaration), and the unqualified value for the `attributeFormDefault` attribute;
- Describe the version of the Application Schema.

The following tag is common to all the Schemas in the PPMS GeoDB Application; more prefixes are added in particular Application Schemas when necessary (e.g. `hydro`):

```
<schema targetNamespace="http://www.ccom.unh.edu/ppms"
  xmlns:ppms="http://www.ccom.unh.edu/ppms"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema"
  xmlns:gml="http://www.opengis.net/gml/3.2"
  elementFormDefault="qualified"
  attributeFormDefault="unqualified"
  version="0.0.1"/>
```

Import of Core Schemas

ISO 19136 GML has 28 Core Schemas; some of them can be clustered into the same object class: feature, geometry, topology, reference systems, coverages, observations, etc.

All GML Core Schemas are defined in the namespace with one of the following two identifiers:

- http://www.opengis.net/gml/3.2; or
- http://www.iso.org/ittf/ISO_19136_Schemas.

An Application Schema does not usually need to import all the available Core Schemas. For instance, there are 5 top-level schemas: the import of one of them automatically imports all the schemas on which it depends. Since some XML Schema parsers recognize only the first schema import, not more than one schema should be added from the GML namespace. The creation of a custom profile to include only certain Schemas and Schema components is also possible.

In the present implementation, the Root Application Schema (*root.xsd*) imports *gml.xsd*, a wrapper Core Schema including all the other GML Core Schemas (Figure 84).

```
<import namespace="http://www.opengis.net/gml/3.2"
  schemaLocation= "http://schemas.opengis.net/gml/3.2.1/gml.xsd"/>
```

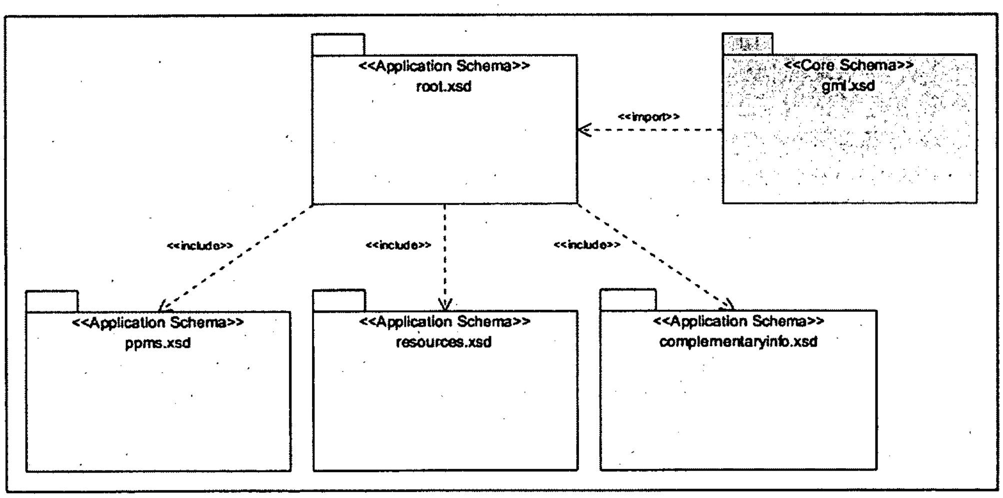


Figure 84 - Import of the GML Core Schemas wrapper within the Application Schemas.

The fundamental difference between `import` and `include` is that the first is used to refer to declarations or definitions that are in a *different* target namespace, the latter in the *same* target namespace.

Derivation of objects and object collections from the corresponding GML abstract types

This section of the Product Specification follows the subdivision in four Application Schemas (and a wrapper).

The *root.xsd* Application Schema is used by any instance document following the PPMS GeoDB Application. The Schema describes the `Root` element and the three possible main Feature Collections it can have as children: `PPMSCollection`, `ResourcesCollection`, and `ComplementaryInfoCollection`. At the same time, it defines the abstract basic Feature Collections that each main Feature Collection can contain: `AbstractPPMS`, `AbstractResources`, and `AbstractComplementaryInfo`. The other Application Schemas derive for extension from these features.

All the objects described are features that, inheriting from the `AbstractFeature` object, have in common the same standard GML properties and the `gml:id` attribute (Figure 85).

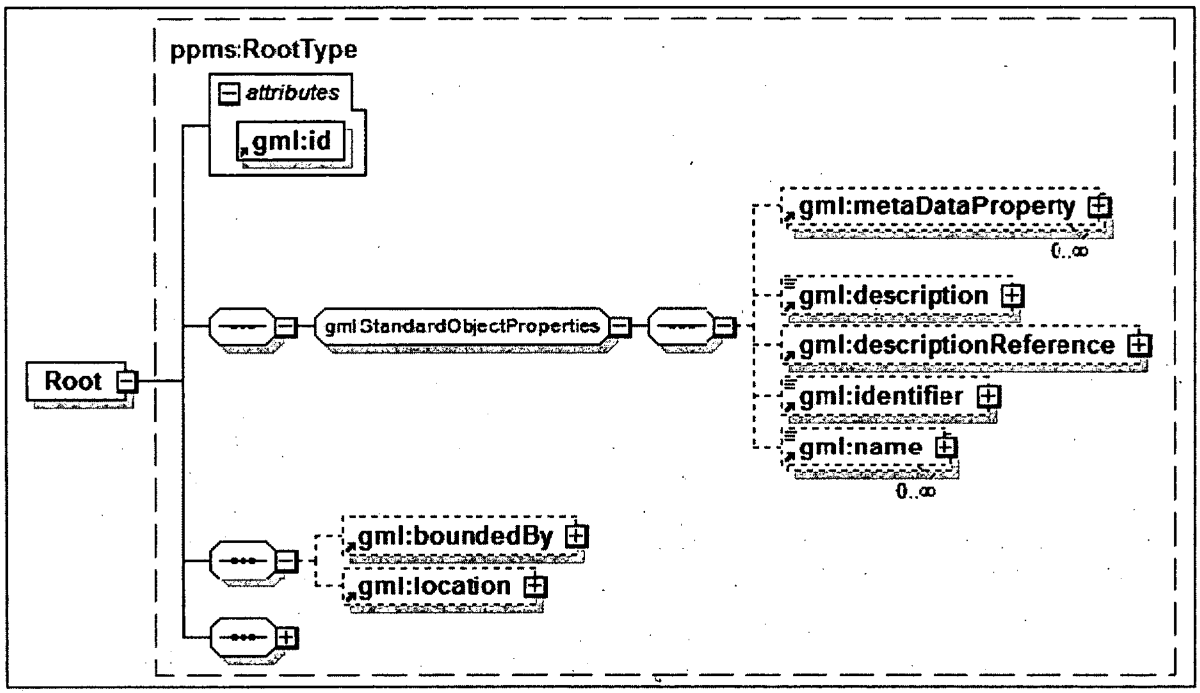


Figure 85 - Root element content model inherited from `AbstractFeatureType`.

The elements in the Root Application schema and their relationships are display in Figure 86.

The optional properties (`hasPPMS`, `hasResources`, and `hasComplementaryInfo`) allow association of the `Root` element and the mentioned main Feature Collections (Figure 87).

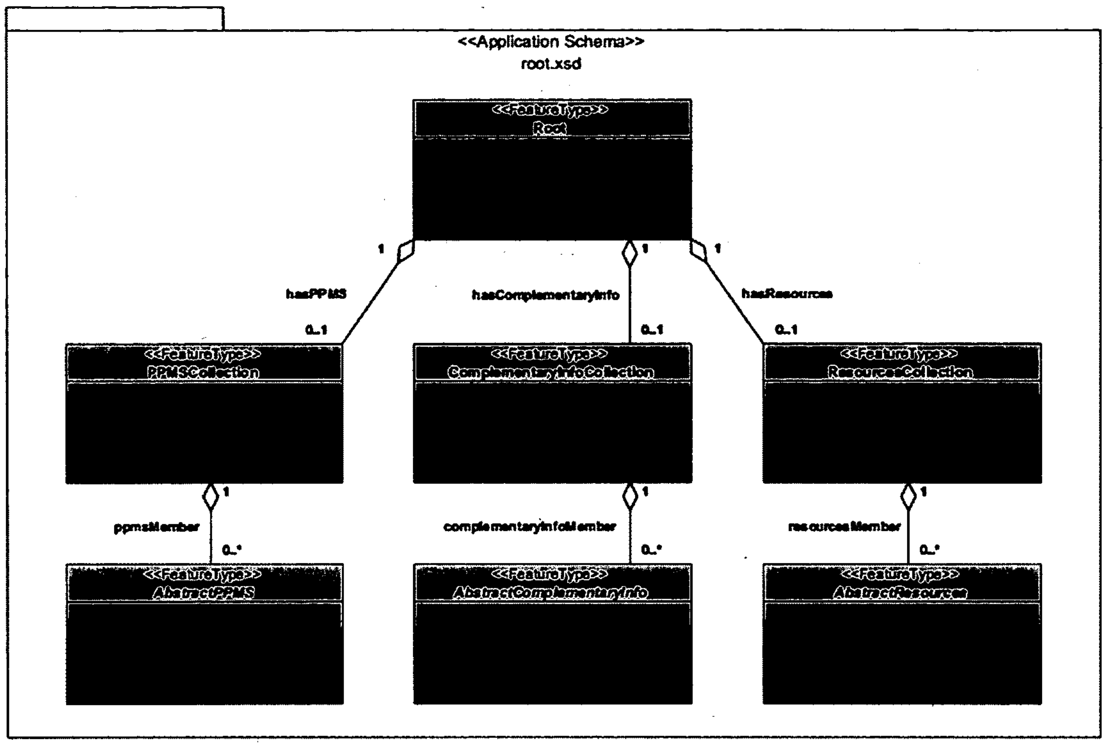


Figure 86 - UML representation of the relationship between the root element and the 3 main feature collections.

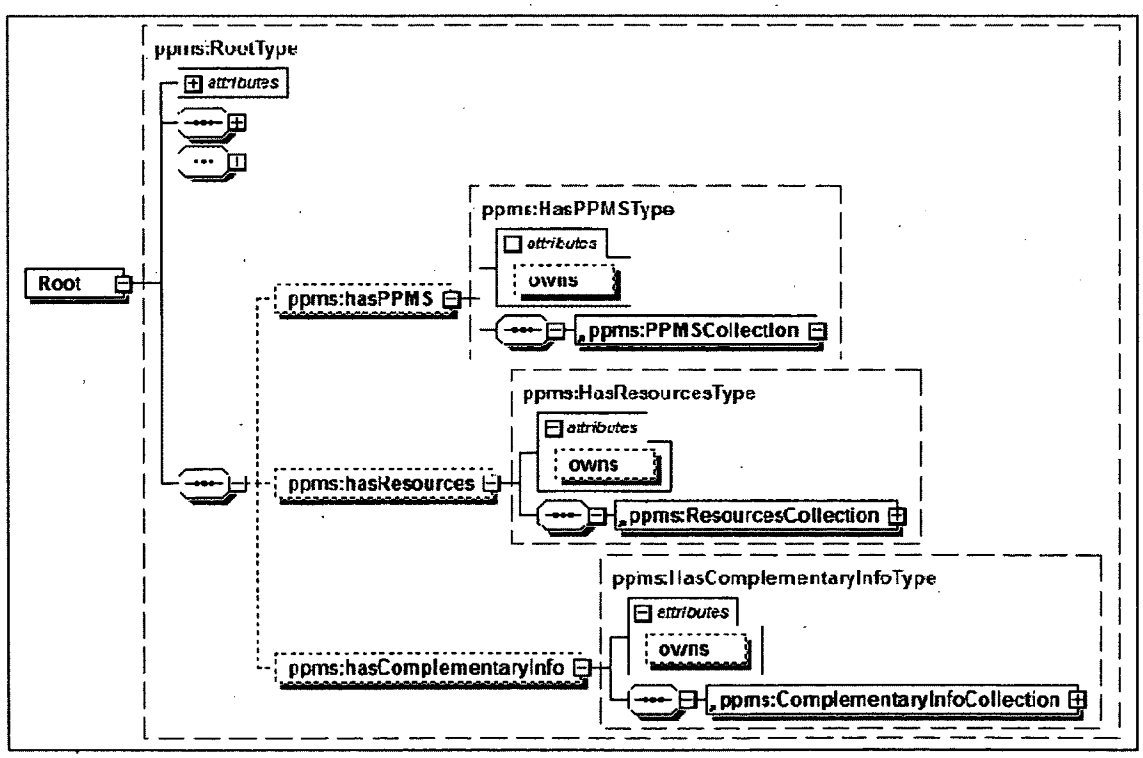


Figure 87 - Optional properties to link the `Root` element and the main Feature Collections.

The three main Feature Collections have as child an abstract element (as shown in Figure 88 for PPMSs, Figure 89 for Marine Resources, and Figure 90 for Complementary Information) to be substituted by the different kind of specific feature collections and features described in the corresponding Application Schema.

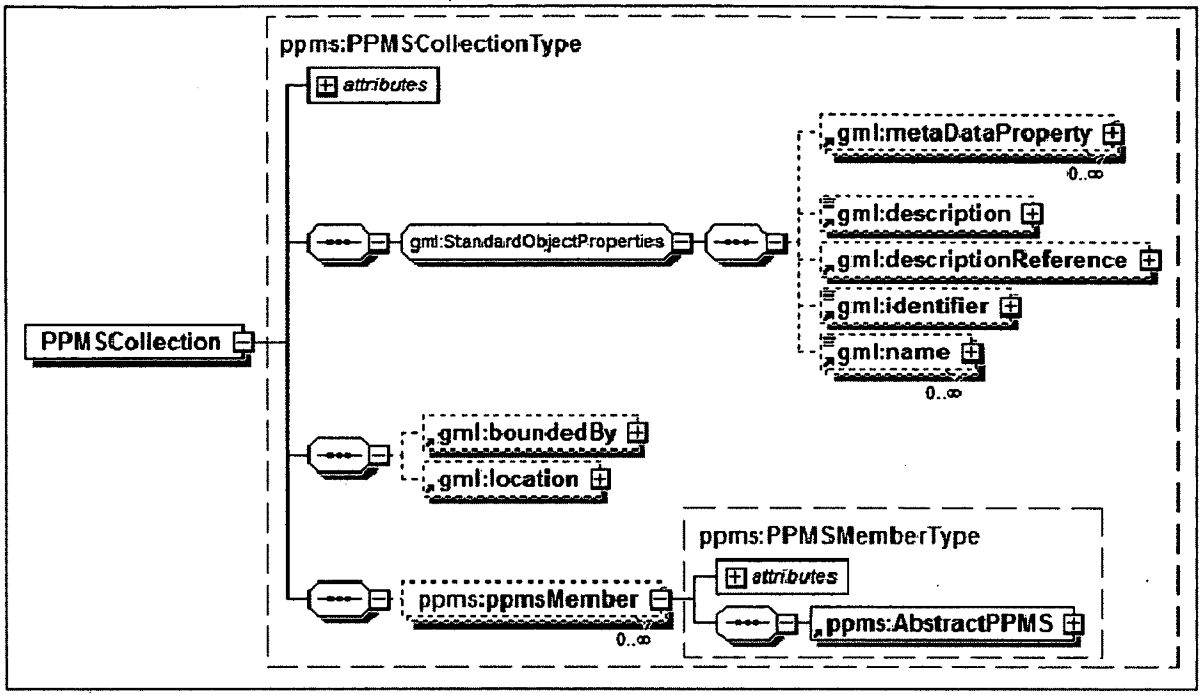


Figure 88 - `PPMSCollection` properties.

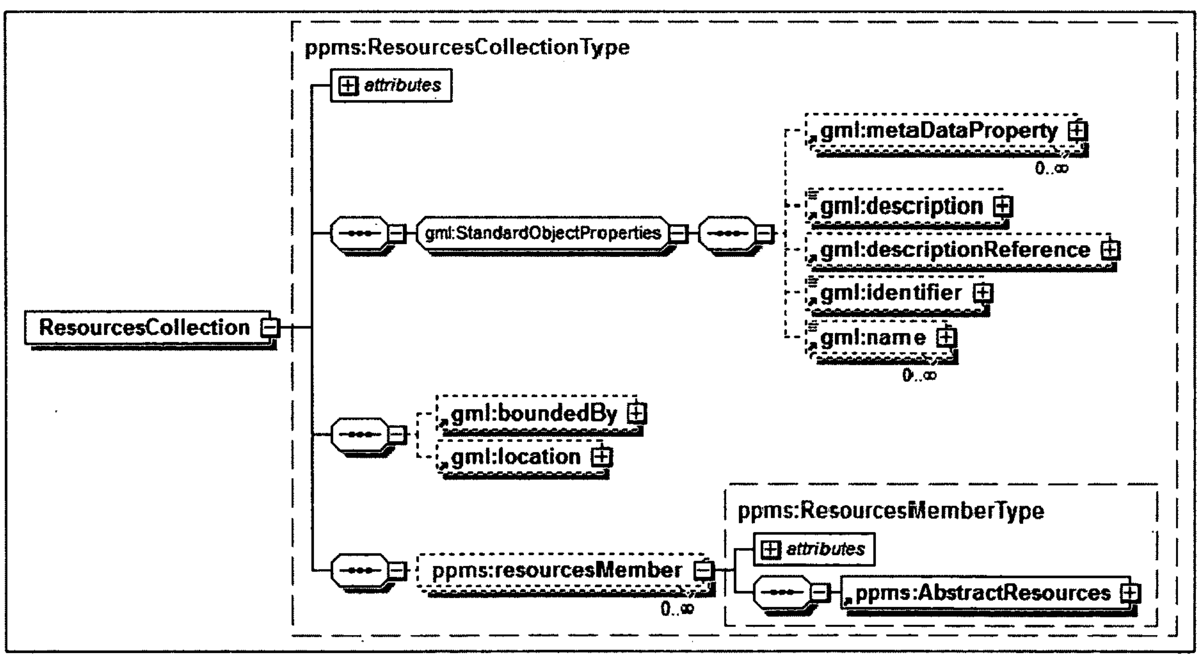


Figure 89 - `ResourcesCollection` properties.

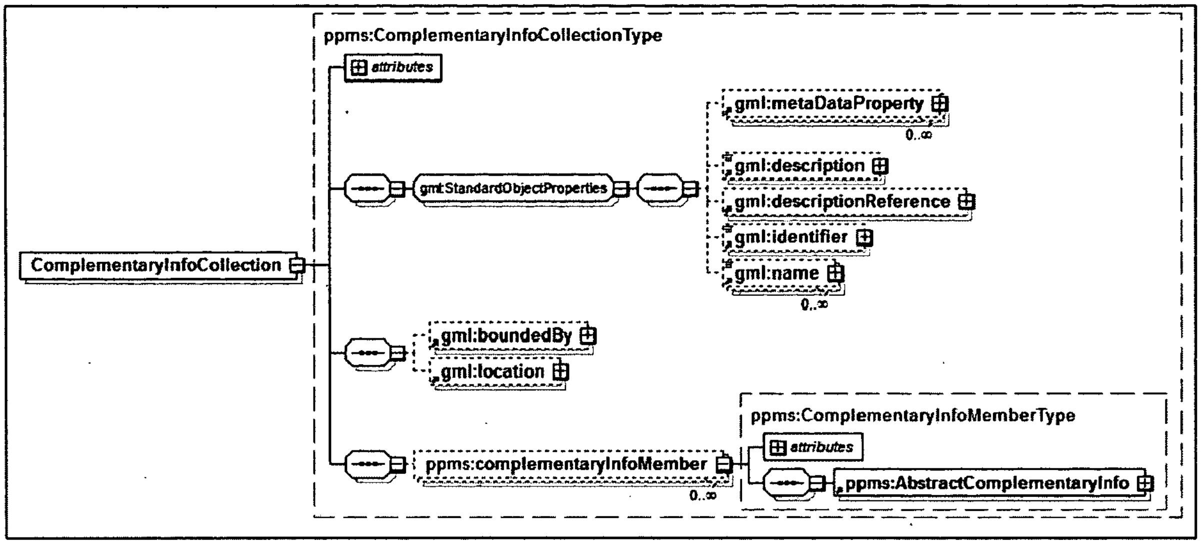


Figure 90 - `ComplementaryInfoCollection` properties.

The *ppms.xsd* Application Schema defines several kinds of Feature Collection, each one deriving for extension from the `AbstractPPMS` feature object (previously defined in the Root Application Schema). They cluster for typology the Potential Polluting Marine Sites (Figure 91).

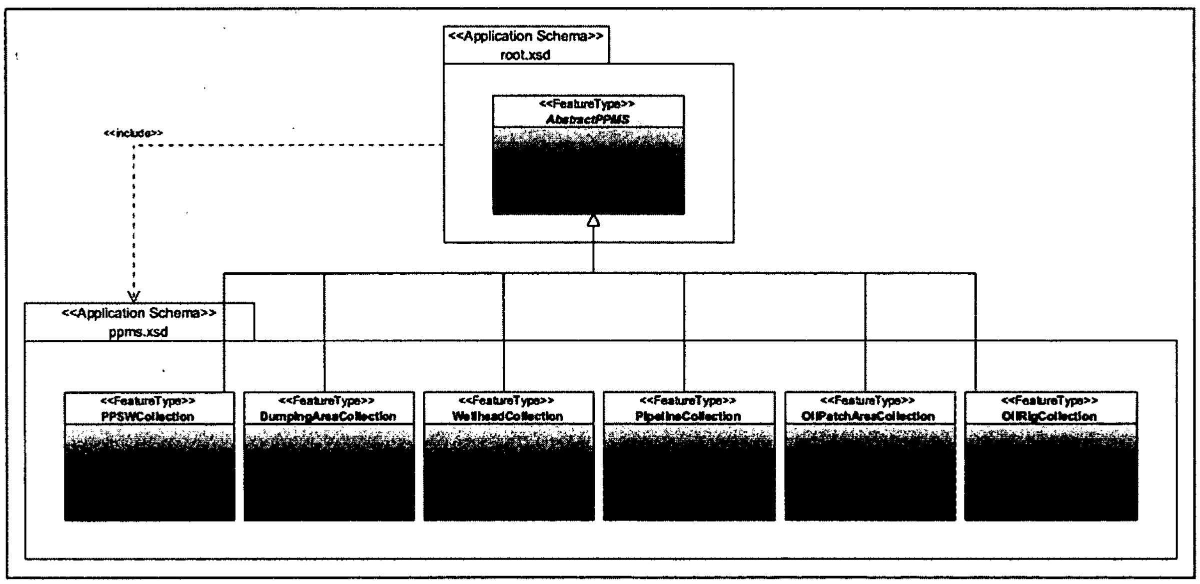


Figure 91 - UML representation of the relationship between the `AbstractPPMS` feature collection and the derived different typology of PPMS Feature Collections.

It is a direct consequence that the main Feature Collection – described in the Root Application Schema – may have as members a variable and unbounded number of basic Feature Collections described in the PPMS Application Schema (Figure 92). One advantage of this approach is the possibility to manage different basic Feature Collection of the same kind (for instance, `PPSWCollection`) created by different Organizations.

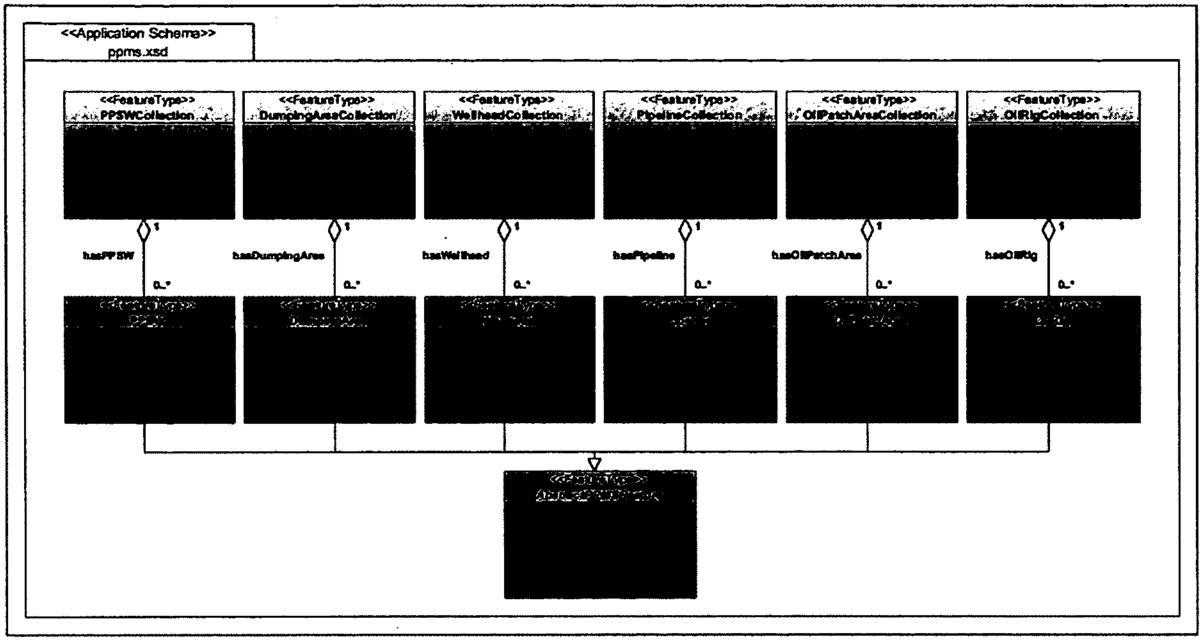


Figure 92 - UML representation of the relationship between the basic PPMS feature collection and their feature member, each one deriving from `AbstractPPMSFeature`.

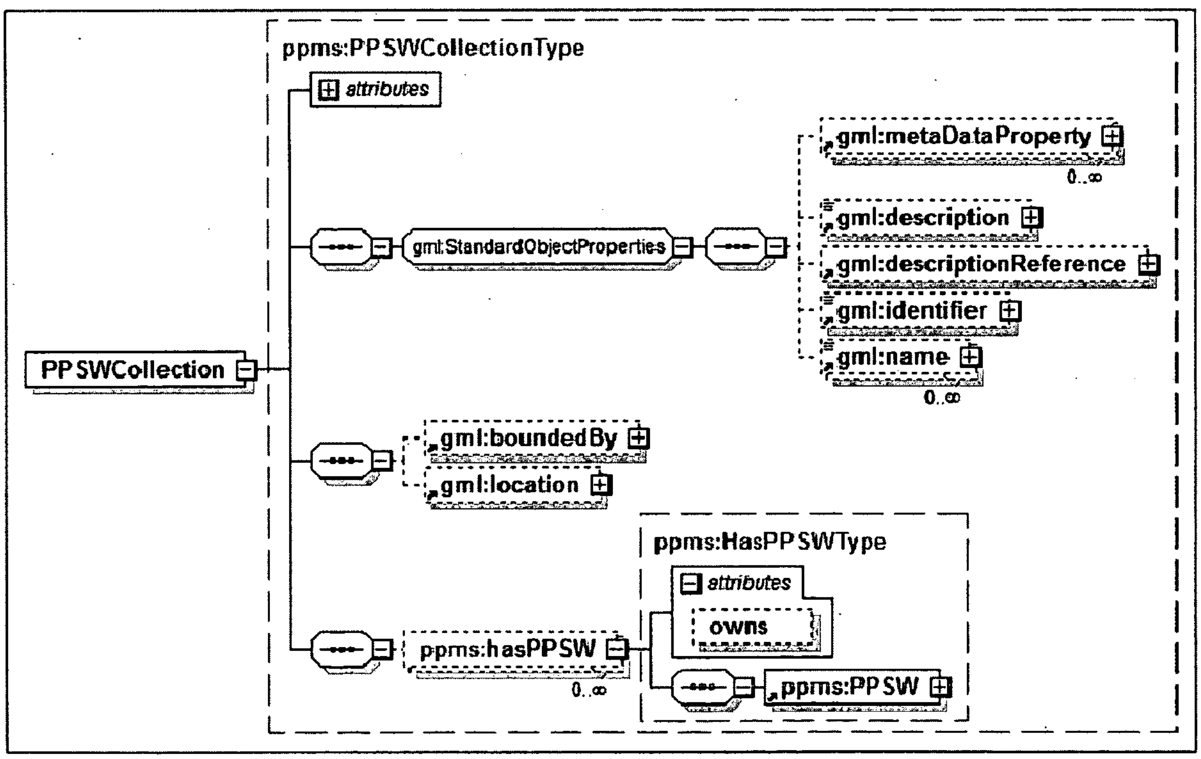


Figure 93 - The `hasPPSW` property of the `PPSWCollection` feature permits to have an unbounded number of `PPSW` features.

In order to permit the definitions of common properties and attributes to all the PPMS features, an `AbstractPPMSFeature` is created. Thus, all the PPMS features are derived for extension from this abstract element: `PPSW`, `DumpingArea`, `Wellhead`, `Pipeline`, `OilPatchArea`, and `OilRig`. An example of the content model for a basic PPMS Feature Collection (`PPSWCollection`) is shown in Figure 93.

A similar approach is used for the Marine Resources Application Scema (*resources.xsd*) and the Complementary Info Application schema (*complementaryinfo.xsd*). The *resources.xsd* Application Schema clusters different typologies of basic Feature Collections representing environmental and economic resources related to the presence of Potential Polluting Marine Sites (Figure 94).

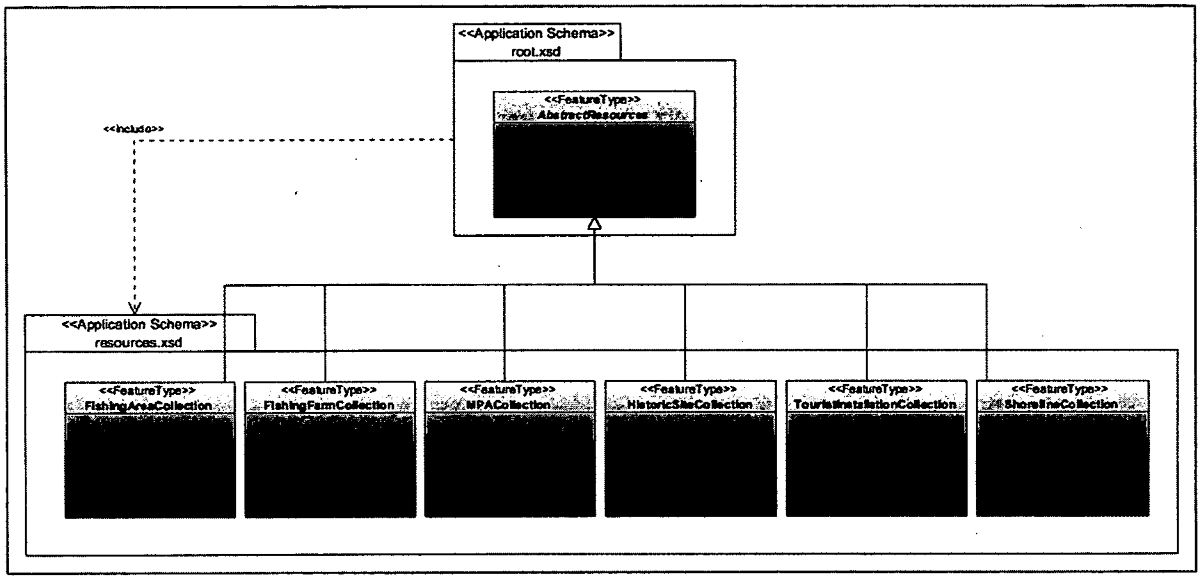


Figure 94 - UML representation of the relationship between the `AbstractResources` feature collection and the derived different typology of Resources Feature Collections.

It is a direct consequence that the main Feature Collection – described in the Root Application Schema – may have as members a variable and unbounded number of basic Feature Collections described in the Resources Application Schema (Figure 95). One advantage of this approach is the possibility to manage different basic Feature Collection of the same kind (for instance, `ShorelineCollection`) created by different Organizations.

Figure 96 shows an example of the content model for a basic Resources Feature Collection (`ShorelineCollection`). The relationships between basic Feature Collection, belonging to the main `ComplementaryInfoCollection`, and their Feature members are displayed in Figure 97. Figure 98 shows an example of the content model for a basic Complementary Info Feature Collection (`SurveyCollection`).

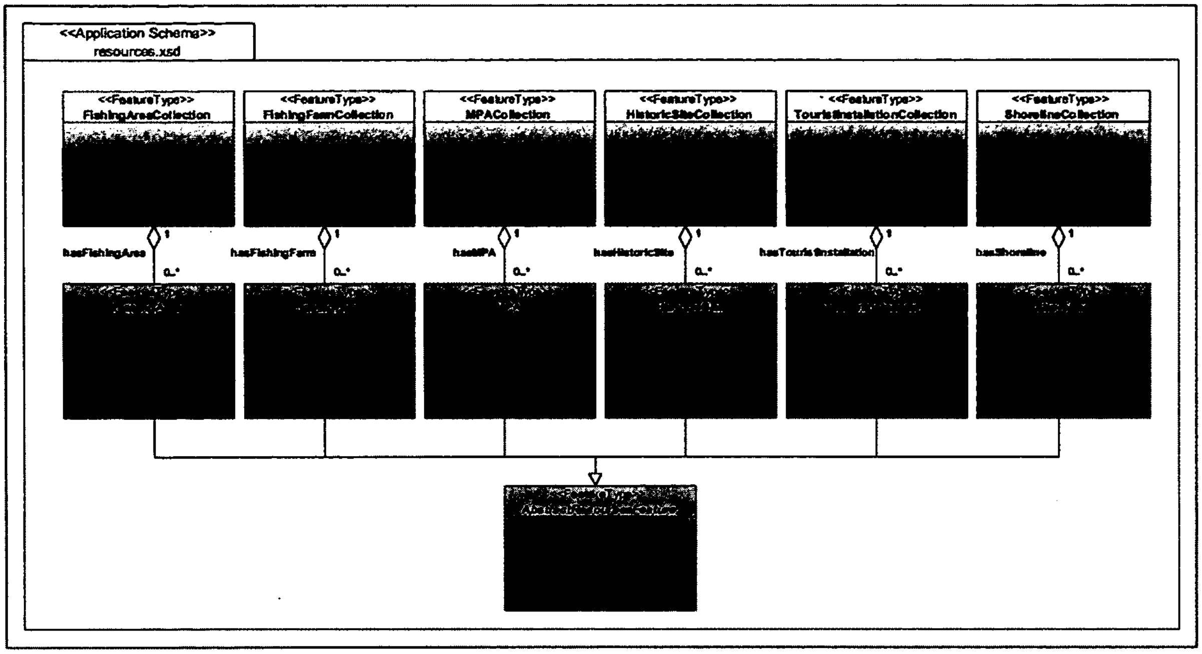


Figure 95 - UML representation of the relationship between the basic Resources Feature Collections and their feature member, each one deriving from `AbstractResourcesFeature`.

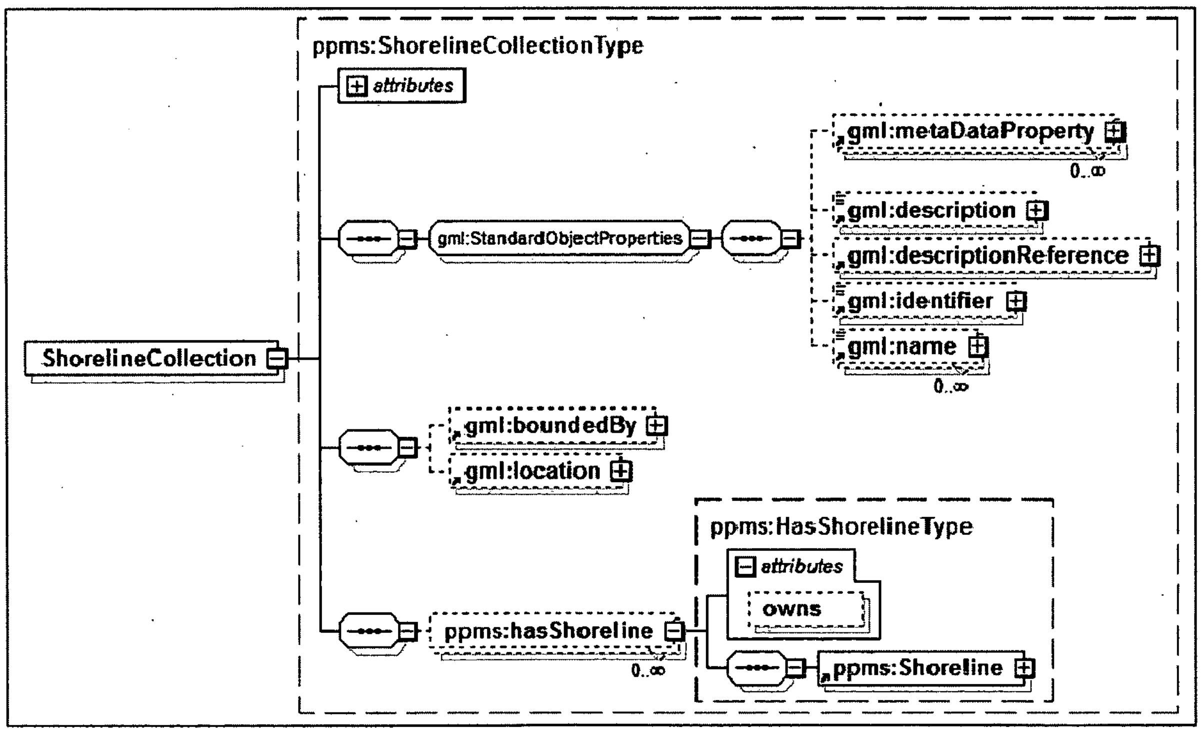


Figure 96 - The `hasShoreline` property of the `ShorelineCollection` feature permits to have an unbounded number of `Shoreline` features.

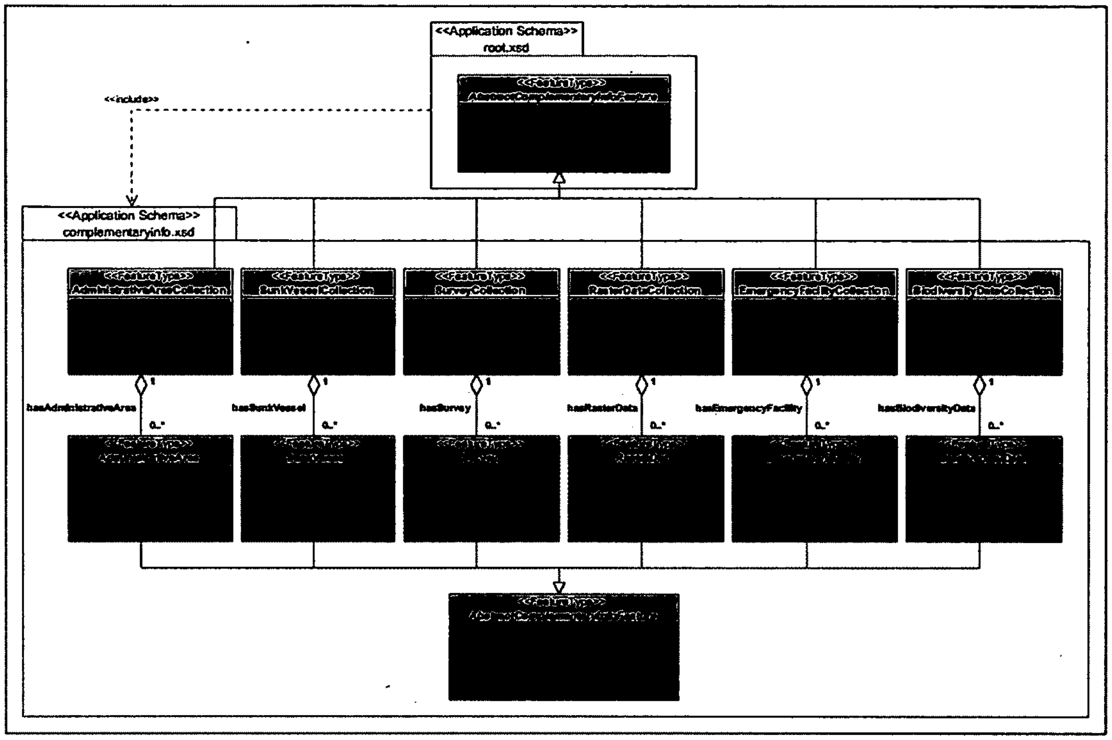


Figure 97 - UML representation of the relationship between the basic Complementary Info Collections and their feature member, each one deriving from `AbstractComplementaryInfoFeature`.

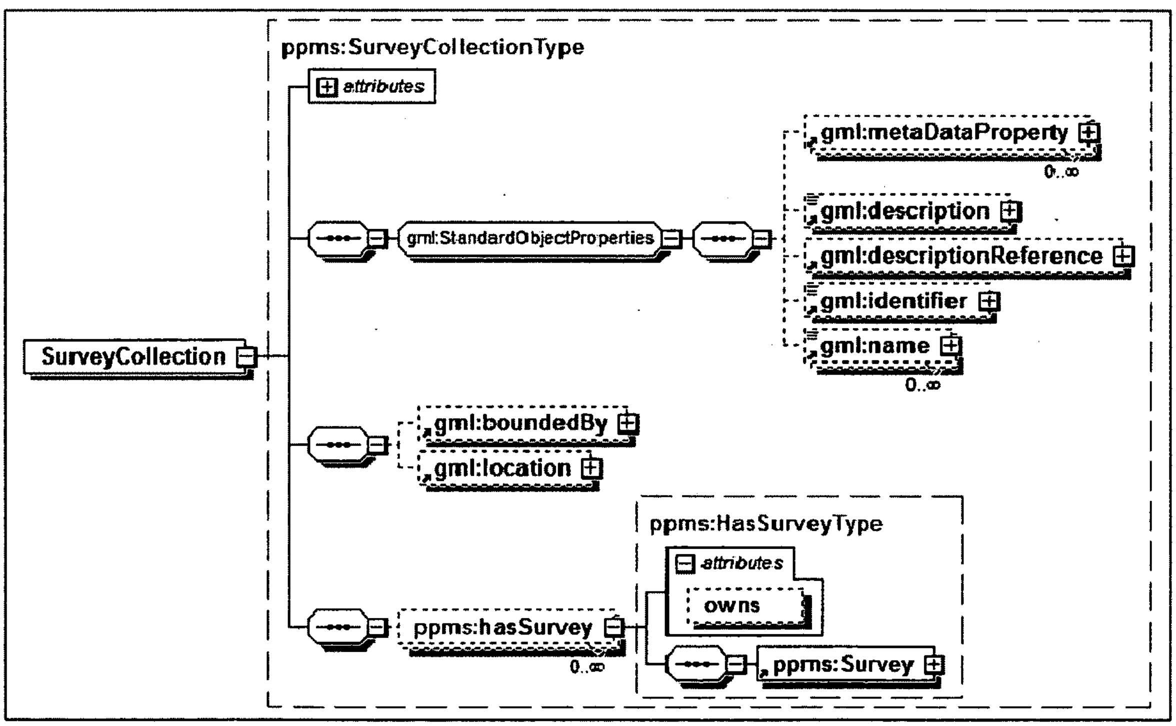


Figure 98 - The `hasSurvey` property of the `SurveyCollection` feature permits to have an unbounded number of `Survey` features.

Definition of object properties

The integration with the existing IHO Models is provided by the creation of global properties in a GML Application Schema named *hydro.xsd* and based on content models already present in the IHO Registry (Figure 99). This approach is potentially useful for the future creation of a PPMS Feature Concept Dictionary in the IHO Supplementary Geospatial Information Registry.

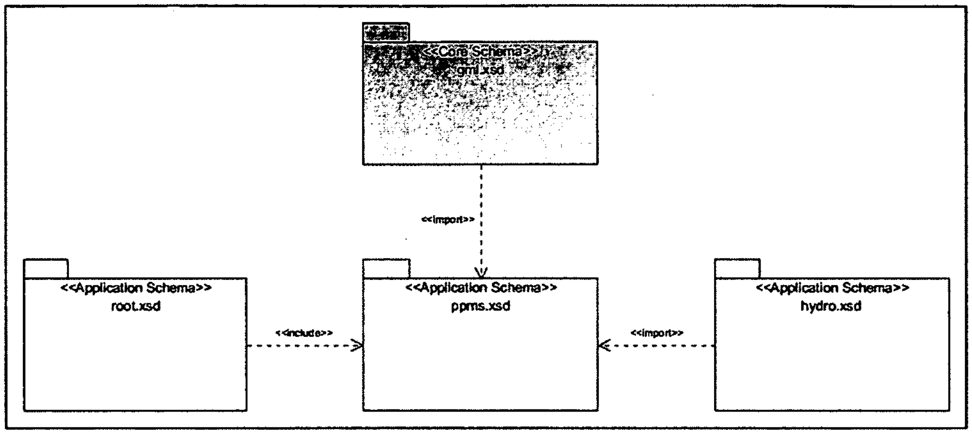


Figure 99 - Import of the Hydro Application Schema.

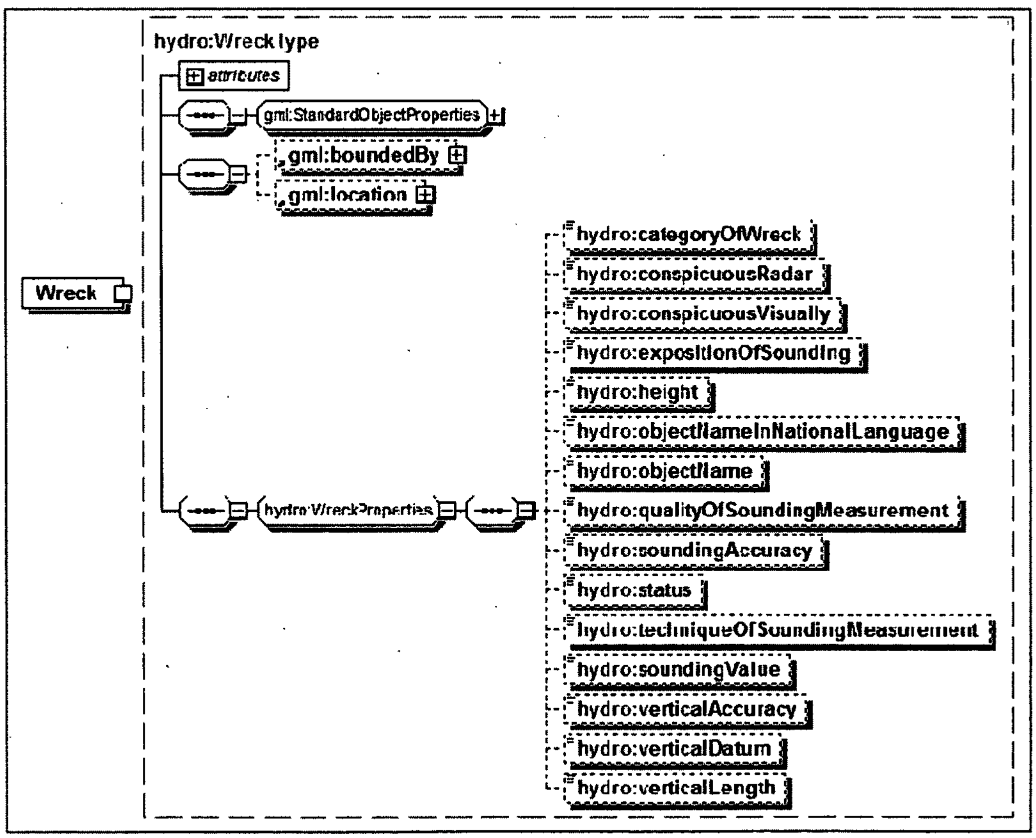


Figure 100 - Properties of `hydro:Wreck`.

The case of a ppms:PPSW integrated with the WreckProperties group of hydro:wreck is presented (Figure 100). Some of the wreck properties are simply enumerations derived for restriction from xsd:string.

The ppms:PPSWType derives for extension from ppms:AbstractPPMSFeature. This latter has a group of properties describing the attributes required for a wreck in the IHO Registry. Thus, a subset of information present in a PPSW instance object can represent the content model of a hydro:wreck.

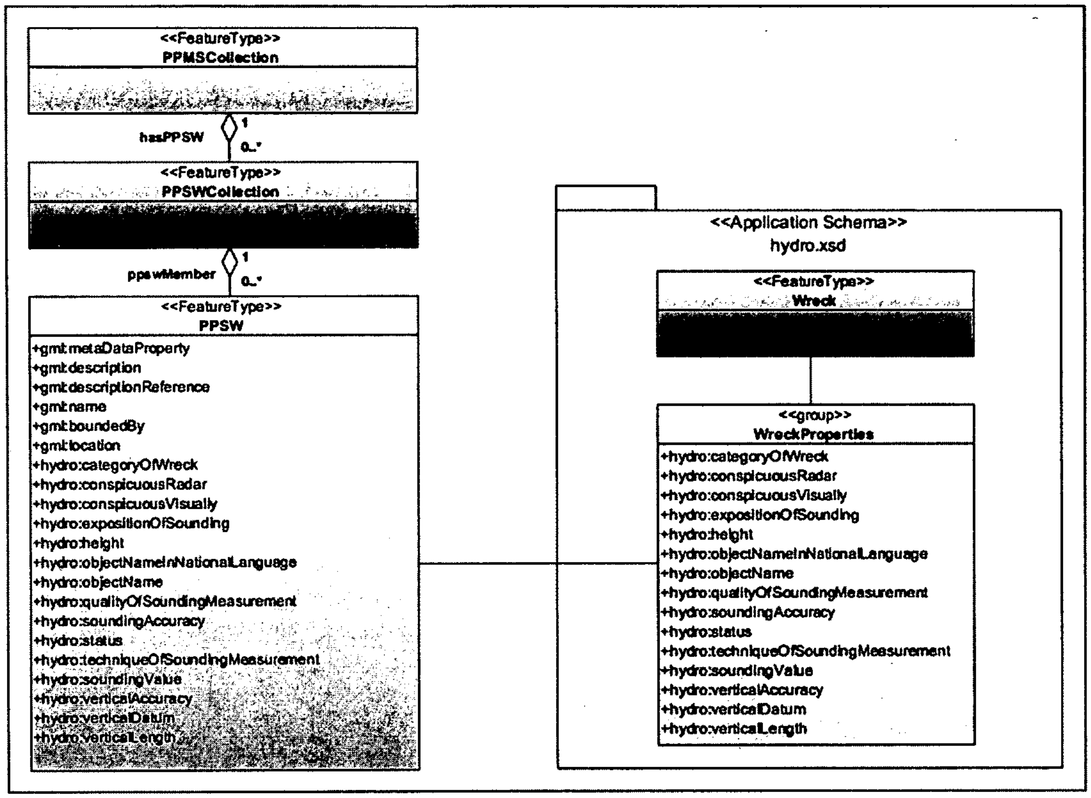


Figure 101 - *PPSW* properties inherited from AbstractPPMSFeature and adopted from the hydro-domain.

The content model of a ppms:PPMS feature is integrated with additional properties. Before this integration, its properties were represented as in Figure 102.

The above mentioned properties are finally integrated with others coming from a survey of information present in existing shipwreck databases:

- administrativeAreaReference: represents the relationship with one or more instances (e.g., EEZ, national waters) of the AdministrativeArea class;
- coastalState: this property contain an enumeration to define, if applicable or possible, the Coastal State in which the PPSW is present;
- sunkVesselIdentification: an enumeration with the following entries: unknown, possible, probable, certain (visually) that is strictly linked to the number of entries in the sunkVesselReference field;
- sunkVesselReference: provides a reference to a certain sunkVessel feature or to probable / possible sunkVessel features;
- flagState: an enumeration to define, if applicable or possible, the Flag State to whom the PPSW belongs;

- minDistanceCoast:the minimum distance from the coast;
- minDistanceSensitiveArea: the minimum distance among the closest sensitive areas;
- surveyReference: a reference to a geophysical survey executed of the site;
- natureOfSeafloor: a simple text description of the type of seafloor and bathymetry around a site;
- discoveringAuthority: the organization that discovered the PPMS;
- discoveryMethod: a simple text string describing the way the PPMS was discovered (it can be different from hydro:techniqueOfSoundingMeasurement which represents the technique used for the value reported in Nautical documents);
- externalResource: links to images, websites or other resources with information related to the PPMS;
- pollutingEvents: an enumeration (unknown, possible, probable, certified, multiple) that is strictly linked to the number of entries in the pollutingEventInfo field;;
- pollutingEventInfo: a simple text description of the main notice related to a spill event;
- conservationStatus: a simple text description of the conservation status of a PPMS (and notices related to any tentative recovery);
- hydrodynamicConditions: information judged of interest to the hydrodynamic conditions in the surroundings of a PPMS site;
- distributionRights: an enumeration list on the diffusion rights related to a PPMS site (e.g., the position of a sunk military vessel);
- riskIndex.

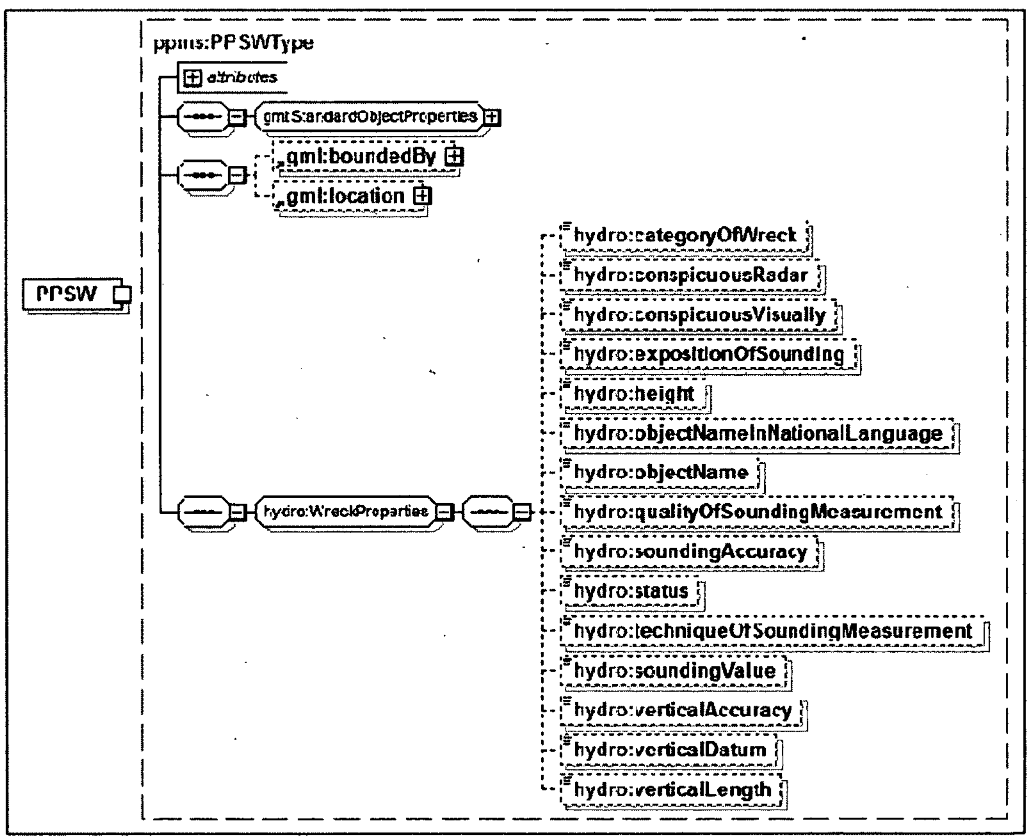


Figure 102 – PPSW properties: inherited from AbstractPPMSFeature and adopted from the IHO hydro-domain.

The mentioned additional properties are divided in two groups:

- State-related properties: administrativeAreaReference, coastalState, sunkVesselIdentification, sunkVesselReference, flagState, minDistanceCoast, minDistanceSensitiveArea, discoveringAuthority, discoveryMethod, distributionRights;
- Event-related properties: surveyReference, natureOfSeafloor, externalResource, pollutingEvents, pollutingEventInfo, conservationStatus, hydrodynamicConditions, and riskIndex.

Some of the above properties (`administrativeAreaReference`, `coastalState`, `minDistanceCoast`, `minDistanceSensitiveArea`, `distributionRights`) are common with other PPMS objects, thus they are defined at the level of `AbstractPPMSFeature`.

Furthermore, since event-related properties are of frequent update, a `PPMSEvent` feature allows the history-approach for dynamic properties.

This approach aims to limit the multiplication of snapshots with full-property `PPSW` objects (and, more generally, of full-property instance features derived from `AbstractPPMSFeature`).

The resulting content model of the `AbstractPPMSFeature` is reported in Figure 103.

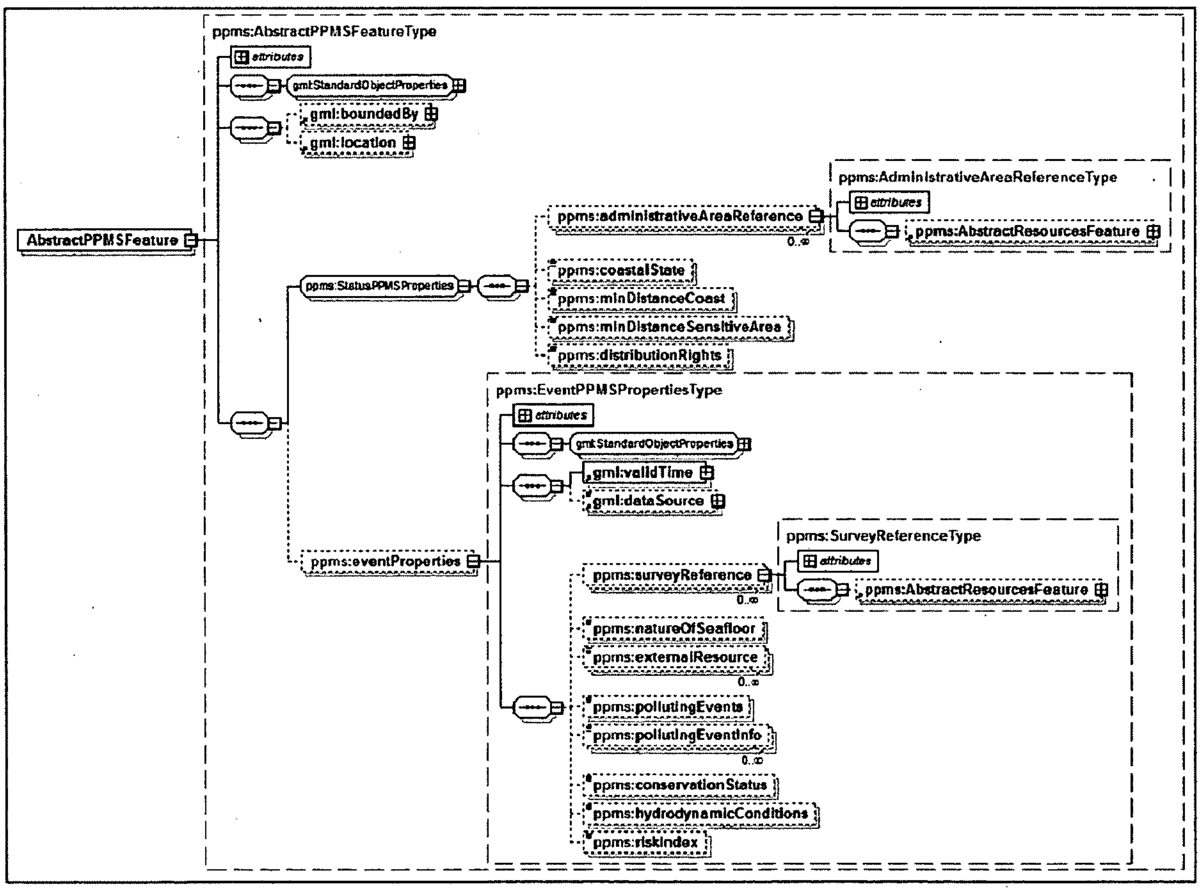


Figure 103 - Content model of `AbstractPPMSFeature`.

Finally, a limited number of properties (`sunkVesselIdentification`, `sunkVesselReference`, `flagState`, `discoveringAuthority`, and `discoveryMethod`) are directly added to the `PPSWType` content model (Figure 104).

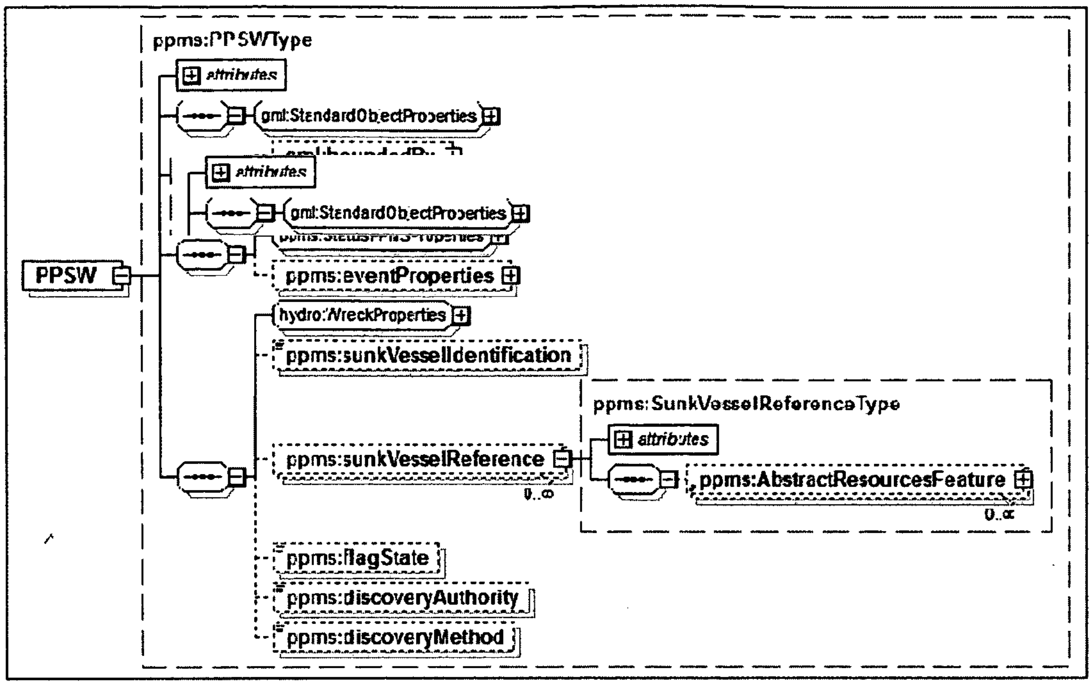


Figure 104 - Integration of `ShipwreckType` with additional properties and a reference to `SunkVessel` feature.

Definition of Metadata schemas

The PPMS GeoDB application adopts the metadata approach provided by ISO 19136.

This approach allows assignment of metadata to objects present in an Application Schema. Metadata have to follow a user-defined Metadata Schema or use an existing one (e.g., based on an International standard). In the case of the PPMS GeoDB, the ISO 19115 metadata standard is adopted. In future, when an IHO metadata profile will be adopted, the related schema should be substituted.

A Metadata Schema is composed of a root metadata package element and a set of properties. In the ISO 19136 implementation, these become the metadata properties of any object which contains or references an instance of the metadata package.

The package element is substitutable for `AbstractMetadata`, and the content model must be a complex type that derives by extension from `AbstractMetaDataType`. Since all GML objects have a generic `metaDataProperty` element, it can be used to include metadata.

The PPMS GeoDB application follows an alternative (but standard valid) approach. Three level of custom metadata property are only allowed for:

- `root` element;
- Main feature collections; and
- Basic feature collections.

This solution was selected to avoid excessive 'granularity' in the metadata information. The idea here is that these types of information, at the single instance level, are better managed if they became part of the properties of a given instance (e.g., `description`).

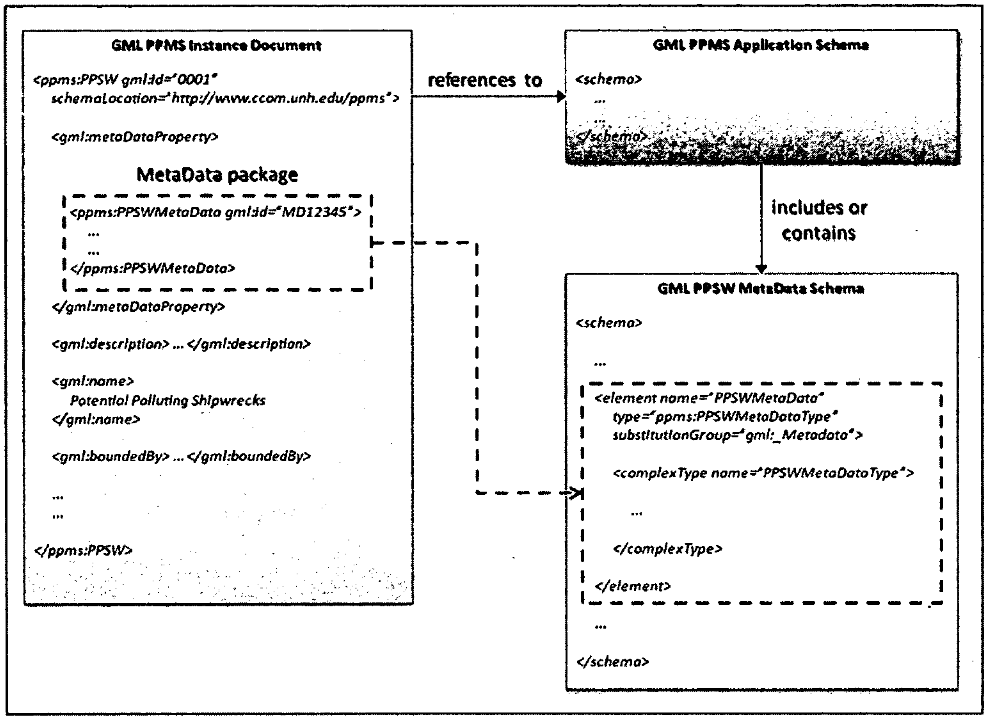


Figure 105 - Example of feature collection with a metadata package

However, the `MetaDataPropertyType` has also an optional `about` attribute that can be used in GML instances to apply metadata to other GML objects. Thus, for instance, the metadata of `PPSWCollection` can have an `about` attribute that points to a specific `PPSW` feature. This attribute was maintained but its extensive adoption should be avoided.

APPENDIX I

XML DATA HANDLING

Typically, a C++ program that has to manipulate data stored in any XML format uses one of the two common XML access Application Programming Interfaces:

- Document Object Model (DOM) that represents XML as a tree-like data structure which can be examined by the application;
- Simple API for XML (SAX), an event-driven XML processing API to whom an application may register its interest in certain events (e.g., element tag, attribute, text, etc.) which are triggered during the parsing of an XML instance.

One of the big difference between these interfaces is that, while DOM has to read the whole document before the application can examine the data, SAX delivers data during the parsing progress. However, both DOM and SAX works at the raw representation of the XML structure: elements, attributes, and text. As a direct consequence, the developer is called to write a substantial amount of 'bridging' code to transform pieces of information encoded in XML to a representation more suitable for the application.

A basic XML example is provided to better clarify issues related to this approach:

```
<SurveySystem>
  <model>EM122</model>
  <manufacturer>Kongsberg Maritime</manufacturer>
  <frequency>12</frequency>
</SurveySystem>
```

For evaluating if the `SurveySystem`'s frequency is greater than some value, both DOM and SAX first find the `<frequency>` element, then they parse the string value "`12`" to obtain the integer value for the comparison.

This example clearly identify a weakness in string-based flow control: any misspelling within the code for the string "`frequency`" will represent a runtime bug. A related issue is a lack in type safety because all the information is represented as text. It is possible, for example, to compare the content of the `manufacturer` element to an invalid value without any warning from the compiler (the code examples are all in C/C++):

```
DOMElement* frequency = ...

if( frequency->getTextContent() == "KM" )
{
  ..
}
```

More generally, this type of approach reduces code readability as well its maintainability.

A more recent approach, called XML Data Binding, permits these limitations to be overcome since it is based on XML vocabulary specification languages (XML Schemas). Recently, many programmers have created slightly different implementations of the XML Data Binding approach following the wealth of interest in XML in the last few years (McLaughlin, 2002). Only the library selected for the PPMS GeoDB Manager implementation is briefly presented in the last paragraph of this Appendix.

XML Data Binding

There are two main advantages of this technique:

- Strong typing, since it guarantees that the XML data has passed XML parsing validation as well as being type safe.
- Schema-specific parsing and validation, that is more efficient that most other XML parser techniques.

The name of this technique comes from the observation that the object representation for stored XML data is essentially bound by one or more XML Schemas. As discussed at length in Appendices F and G, a schema is a formal specification of a vocabulary that defines the names of elements and attributes, as well as the content and the structural relationship between them. Thus, the idea is to deliver the data in an object-oriented representation based on an XML vocabulary, skipping in this way the raw representation of XML and its weaknesses.

The majority of XML Data Binding tools use the W3C XML Schema specification language due to its evident object-oriented approach to vocabulary description as well as its widespread use. The following fragment describes a simplified W3C XML Schema that will be used in the next examples:

```
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">

  <xs:element name="surveySystem" type="surveySystemType"/>

  <xs:complexType name="surveySystemType">
    <xs:sequence>
          <xs:element name="model" type="xs:string"/>
          <xs:element name="manufacturer" type="manufacturerType"/>
          <xs:element name="frequency" type="xs:short"/>
    </xs:sequence>
  </xs:complexType>

  <xs:simpleType name="manufacturerType">
    <xs:restriction base="xs:string">
          <xs:enumeration value="Kongsberg"/>
          <xs:enumeration value="Reson"/>
          <xs:enumeration value="R2Sonic"/>
    </xs:restriction>
  </xs:simpleType>

</xs:schema>
```

One main advantage of XML Data Binding is that the application developer does not have to produce the bridging code since he can directly use the object model in the implementation of the application logic.

Using the previous example, instead of searching for the `frequency` element and then manually converting the text to an integer, we would simply call the `frequency()` function on the `surveySystem` object that will be able to return the frequency value directly as an integer (i.e., a `short`).

A data binding compiler implements a large part of the job of creating from an XML Schema the vocabulary-specific object representation as well as other support code (mainly, parsing and serialization functions).

Furthermore, XML Data Binding supports both in-memory and event-driven programming models.

In-memory Approach

As a first rough approximation, the XML Data Binding compiler maps:

- Schema type declarations to C++ classes;
- Local elements to C++ accessors and modifiers (setting and getting functions);
- Global elements to a pair of parsing and serialization functions.

For example, the previous specification might become:

```
class manufacturer_t
{
public:
  enum value { Kongsberg, Reson,, R2Sonic };

  manufacturer_t( value );
  value operator() const;

private:
  ..
};

class surveySystem_t
{
public:
  surveySystem_t( const string& model, manufacturer_t manufacturer,
    short frequency );

// model
string& model();
const string& model() const;
void model( const string& );

// manufacturer
manufacturer_t& manufacturer();
manufacturer_t manufacturer () const;
void manufacturer( manufacturer_t );

// frequency
```

```
short& frequency();
short frequency() const;
void frequency( short );

private:
  ..
};

std::auto_ptr<surveySystem_t> surveySystem( std::istream& );
void surveySystem( std::ostream&, const surveySystem_t& );
```

The following fragment of code reads an XML file with a collection of survey systems and print the name of any model with a frequency greater than 10 kHz:

```
ifstream ifs( "surveysystems.xml" );
std::auto_ptr<surveySystem_t> p = surveySystem( ifs );

if( p->frequency() > 10 )
  std::cout << p->model() << std::endl;
```

The same operation using a DOM approach is conceptually split into two parts:

- Data extraction from the raw representation of XML;
- Implementation of the application logic.

The following code fragment performs the same task using a DOM approach:

```
ifstream ifs( "surveysystems.xml" );
DOMDocument* doc = read_dom( ifs );
DOMElement* p = doc->getDocumentElement();

string model;
short frequency;

for ( DOMNode* n = p->getFirstChild(); n != 0;
         n = n->getNextSibling() )
{
  if( n->getNodeType() != DOMNode::ELEMENT_NODE )
   continue;

  string el_name = n->getNodeName();
  DOMNode* text = n->getFirstChild();

  if( el_name == "model" )
  {
   model = text->getNodeValue();
  }
  else if( el_name == "frequency" )
  {
   istringstream iss( text->getNodeValue() );
   iss » frequency;
  }
```

```
}

if( frequency > 10 )
  std::cout << model << std::endl;

doc->release();
```

Comparing the two approaches, it is evident that the DOM version is both more complex and less safe because of the use of strings to identify elements. In fact, any misspelling may not be warned by the compiler. Conversely, using the XML Data Binding approach, the result of misspelling a function name is a compile error.

Furthermore, after the in-memory creation of the XML instance document, the XML Data Binding code is fully object-oriented without any trace of XML with all its classes statically typed. Other advantages are related to the runtime memory usage (e.g., it stores directly numeric data instead of strings) and the concise output code (that makes it easier to read and maintain).

Nevertheless, XML Data Binding cannot be used if the XML instance document is not based on an XML vocabulary and a DOM-approach code may have the advantage that it does not break if a new element is added to the Schema or there is a minor change in an element's type.

Event-driven Approach

When it is neither possible nor desirable to load the whole document into memory (for instance, XML documents too large to fit into memory), it is possible to use either event-driven XML Data Binding or raw XML access APIs (such as SAX) that perform XML processing as parsing progresses. Basically, event-driven XML Data Binding consists of parser templates that represent the given vocabulary as a hierarchy of data availability events which are dispatched using the C++ virtual function mechanism (Kolpackov, 2007).

In the proposed PPMS GeoDB implementation, an in-memory approach is adopted. However, the presence of large PPMS GeoDB GML files could require the adoption in future of an event-driven approach in order to speed up the reading and writing operations.

The *gSOAP* Library

This library was created as an open source project in October 2001 and since then this toolkit has been used by many software companies to develop their products (e.g., Google, eBay, etc.).

Among the many available commercial and open source libraries that implement the XML Data Binding, *gSOAP* was selected mainly for:

- Large diffusion is assumed to be a synonym of reliability;
- License type (free open source for non-commercial use);
- Offering comprehensive and transparent auto-coding techniques; as well as
- Garbage collection for memory management; and
- Internationalization/localization (e.g., UTF-8 encoding).

Furthermore, the library has a wide portability and is mature, also because of the large number of users. Finally, it reported very high scores in tests performed by the W3C XML Schema Patterns for Data Binding Interoperability working group (W3C, 2009).

Actually, the application field of this library is much larger since, as also recognizable in its name, it is designed to simplify the development and deployment of Simple Object Access Protocol (SOAP) XML web services applications in C++ (Chawla and Bhaskar, 2007).

The PPMS GeoDB Manager application only makes use of a limited part of this workflow, the part that maps native and user-defined C++ data types to XML data types and vice-versa.

Main Constituents

The *gSOAP* package includes:

- Pre-built tools for the more common platforms (Windows, Mac OS and Linux):
  - WSDL/schema parser tool (*wsdl2h*) which imports one or more Web Services Description Language (WSDL) files (used for describing web service methods) or XML Schemas, and generates a header file containing the necessary service prototypes and data types;
  - Stub/skeleton compiler (*soapcpp2*) that uses the created header file to generate serialization and de-serialization routines for data types, stub and skeleton code.
- Source code that permits customization of both tools or to build them on another platform;
- The *gSOAP* engine source code; and
- Documentation and several sample codes.

There is no need to download any third-party software unless the user wants to enable SSL (*OpenSSL* library) and compression (*Zlib* library).

Adopted Workflow

The first step is represented by the creation of a header file `ppmsgeodb.h` based on the PPMS GeoDB GML Schema wrapper (`ppmsgeodb.xsd`). The created header file is processed using the stub/skeleton compiler to create source files providing all the necessary functions for data types, stub and skeleton routines.

The reading of a PPMS GML file is easily performed as in the following example code:

```
soap* context = soap_new1( SOAP_XML_STRICT );
context.is( "ppmsDemo.gml" );
ppms__root rt;

if( soap_read_book( context, &rt ) != SOAP_OK )
  /* error code */
else
{
  std::cout << rt.name <<std::endl;
  /* further code */
}

soap_destroy( context );    //clean up allocated class instances
soap_end( context );        //clean up allocated temporaries
soap_free( context );       //delete context
```

In the above code, the calls to `soap_destroy` and `soap_end` functions are used to de-allocate the serialized content since *gSOAP* uses an automatic memory management approach to avoid memory leaks. A similar approach is required to write on disk the XML representation of the manipulated data.

The solution adopted for the PPMS GeoDB Manager is the creation of a stand-alone (shared or static) library that performs the conversion from GML data to the spatial RDBMS. The GML-side of this library integrates the auto-generated *gSOAP* functions, while the input and output related to the database are directly coded.

APPENDIX L

SQLITE AND THE EXTENSION SPATIALITE

***SQLite* software package**

SQLite is a public-domain software package that implements a zero-configuration, standalone, relational database engine that is designed to be embedded directly into an application (Kreibich, 2010). This type of system is used to store user-defined records in large tables, and to process complex query commands that combine data from multiple tables to generate reports and data summaries.

Originally released in 2000, *SQLite* provides a convenient way for applications to manage data without the overhead that often comes with dedicated relational database management systems (Allen and Owens, 2010).

In the market, there are many other RDBMS products such as:

- Oracle Database, IBM's DB2, and Microsoft's SQL Server on the commercial side;
- MySQL and PostgreSQL as open source products.

SQLite was selected to be used in the PPMS GeoDB Manager mainly because of the following characteristics:

- Server-less, does not require a separate server process or system to operate, since it accesses directly storage files. Since there is not a server, there is no need to set it up (no issues with configuration or need for a DB administrator). If different processes work on the same SQLite database, each of them run their own respective code, but they've also become independent database servers in and of themselves (different possible configurations as shown in Figure 106);
- Cross-platform, in fact the entire database instance resides in a single file;
- Self-contained, the entire database system is contained into a single library integrated into the host application;
- Reduced runtime footprint, it has small sizes both for the one megabyte of code and a few megabytes of required memory;
- Safe access from multiple processes or threads (ACID-compliant: atomicity, consistency, isolation, durability for the database transactions);
- Support for most of the query language features of the SQL2 standard; and
- Highly reliable (the standard *SQLite* test suites consists of over 10 million unit tests and query tests).

Furthermore, *SQLite* offers several features that are not common in many other database systems:

- A dynamic-type system for tables allows putting any value into nearly any column, regardless of type (this point could be also seen as a weakness). The possible data types are: `NULL`, `INTEGER`, `REAL`, `TEXT` and `BLOB` (Binary Large Object);

- The ability to manipulate more databases at a time allows, for instance, processing SQL statements that bridge across multiple databases (e.g. tables joining different databases by a single query);
- The ability to create fully in-memory databases that are very fast; and
- No user license, thus compiled libraries can be used and modified in any way, as well as being redistributed and sold.

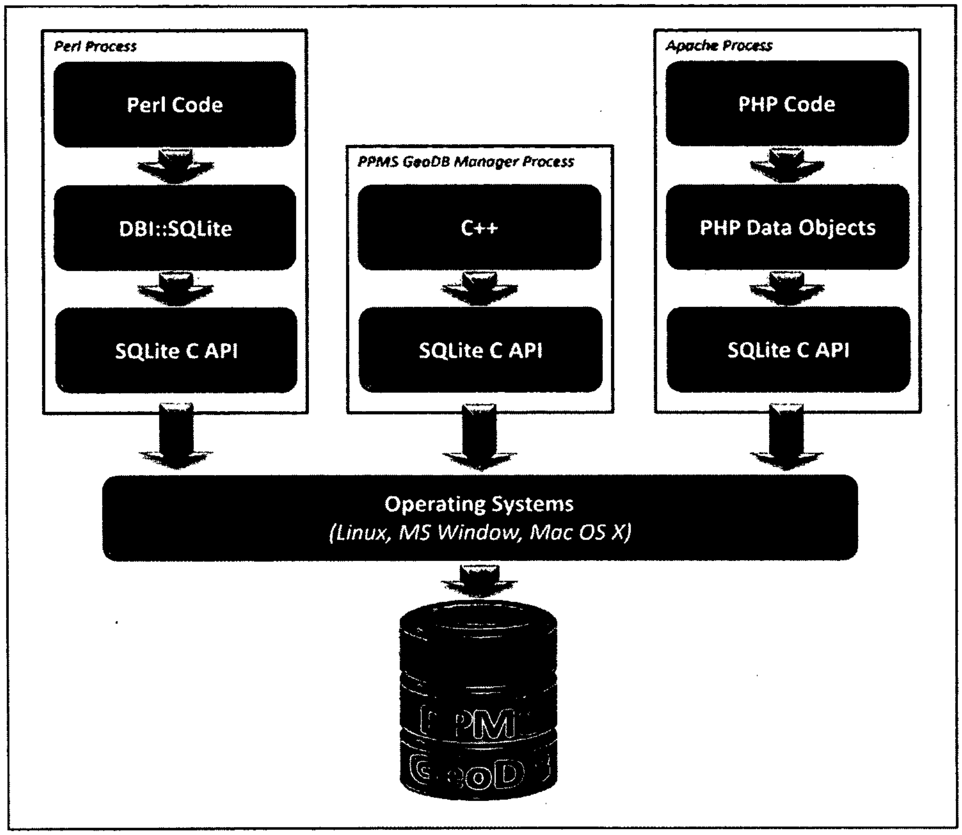


Figure 106 - Server-less configuration of *SQLite* embedded in host processes. The PPMS GeoDB Manager represents one of many possible processes working on a PPMS GeoDB.

Even if powerful in its simplicity, *SQLite* present several well know weaknesses such as, for instance, the fact that it relies on the operating system to manage synchronization and locking.

Another example is represented by the performance speed for queries. *SQLite* uses B-trees for indexes and B+-tree for tables similar to other database systems. For basic queries, it is often quicker than other more complicated solutions (since, for instance, there is less overhead to deal with creating a new transaction and no need to make a network call to the server) but, when queries become more complex, the query optimizer and planner of more complex systems make the difference (Allen and Owens, 2010).

The adoption of the *SQLite* library assures the binary compatibility across platforms for the PPMS GeoDB file format (.pdb). At the same time, since *SQLite* databases are ordinary disk files, they are easy to work with, transport (e.g., in a memory stick), and back up. Furthermore, with so many developers, analysts and generic users familiar with relational databases and SQL, a primary benefit coming from the *SQLite* adoption is assumed user knowledge rather than having to take care about syntax and writing extensive documentation for the user interface.

SpatiaLite: the spatial SQLite extension

SpatiaLite is a Spatial DBMS supporting international standards such as SQL92 and Open Geospatial Consortium - Simple Feature SQL (OGC-SFS) Specification (Furieri, 2012). This latter extends the basic SQL capabilities in order to support a special `Geometry` data-type that supports a Spatial DBMS.

The selection of *SpatiaLite* for the PPMS GeoDB Manager is strongly related to the choice of *SQLite* and to the consideration that a similar light and simple solution may be easily transfered to other systems for management and planning purposes. Moreover *SpatiaLite* consists of a unique cross-platform file that can be easily transferred with all the geographical content. As a direct consequence, an increasing amount of research is looking at *Spatialite* as a valid solution (Brustle, 2012; Carro, 2010; Cau et al., 2011; Ling et al., 2010; Manca et al., 2011; Yin and Carswell, 2012). Furthermore, the open source GIS desktop *QGIS* supports *Spatialite* format so that it is very easy to access many common GIS tools.

Spatialite supports through virtual drivers SQL access to several formats such as, for instance, shapefile, Comma Separated Value (CSV), Microsoft Excel (xls), etc. Even if access to these virtual tables is much slower than using the internal storage, they represent an interesting solution for evaluating unknown data before the import process.

The selection of shapefiles as a geographic format was guided by the consideration that, even if it was born as an *ESRI* proprietary format, it has been publicly disclosed so that it can be considered as a universal and open standard format (ESRI, 1998). Known weaknesses of this format are the missing ability to store topological information, and the maximum size of each component file of 2 GB.

A *Spatialite* DB always uses UTF-8 as its encoding format but it is able to manage a large number of different charset encoding as input and output.

A key element for a Spatial RDBMS is represented by the querying speed, such as the ability to support fast search of selected items within huge datasets. A way to speed up this process is to adopt a spatial index, and many different algorithms have been developed to perform this task.

Spatialite uses SQLite's spatial index that is an implementation of the R*Tree algorithm (Beckmann et al., 1990; Guttman, 1984). The basic idea behind this algorithm is that any geometry, independent of its actual shape, can be represented as a minimum bounding rectangle (MBR) (Figure 107) and a large number of these rectangles organized in a tree can be searched much more quickly than each actual single component point (Figure 108). This algorithm can be easily extended in a multidimensional way.

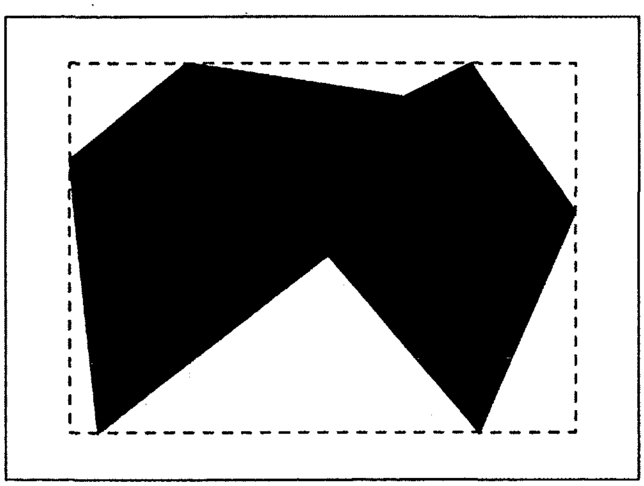


Figure 107 - Minimum Bounding Rectangle for a given polygon.

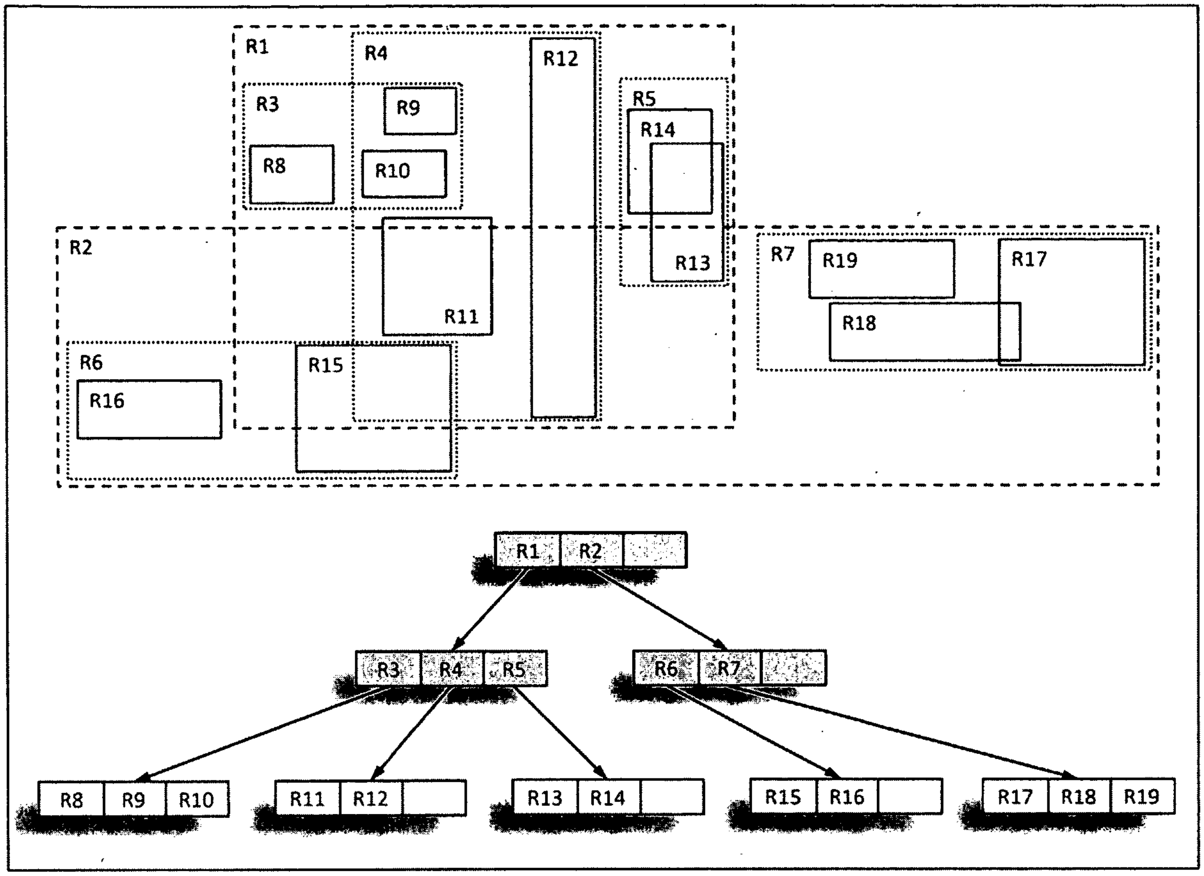


Figure 108 - Example of R*Tree structure.

However, a spatial index can be seen as a filter and, if the selected region covers a large part of the dataset, its adoption may represent a further computation cost (Furieri, 2012).

SQL Queries and Database Interaction

SQL is almost the sole and universal means used to communicate with a relational database (Allen and Owens, 2010).

The language traces its roots back to the early 1970s and the pioneering relational database work that was being done at IBM(Kreibich, 2010). The first official SQL specification was published in 1986 by the American National Standards Institute (ANSI). Current SQL standards are ratified and published by the International Standards Organization (ISO). Although a new standard is published every few years, the last significant set of changes to the core language can be traced to the SQL:1999 standard (also known as SQL3).

The core of SQL is a declarative language, thus the user states the results he wants leaving the language processor to figure out how to deliver them. The predefined nature of declarative statements may sometimes appears to be a bit limited, such as when the problem to solve is inherently non-relational requiring the creation of intermediate steps (e.g., nested queries or temporary tables). Even if all the RDBMSs uses SQL, each of them implements a slightly customized version of the language (Lans and Cools, 2006). *SQLite* understands most of the standard SQL language, but it does omit some features (e.g., right and full outer join, GRANT and

REVOKE commands) while at the same time adding a few features of its own. Because of the SQLite adoption, the PPMS GeoDB Manager uses the same SQL features used by *SQLite*.

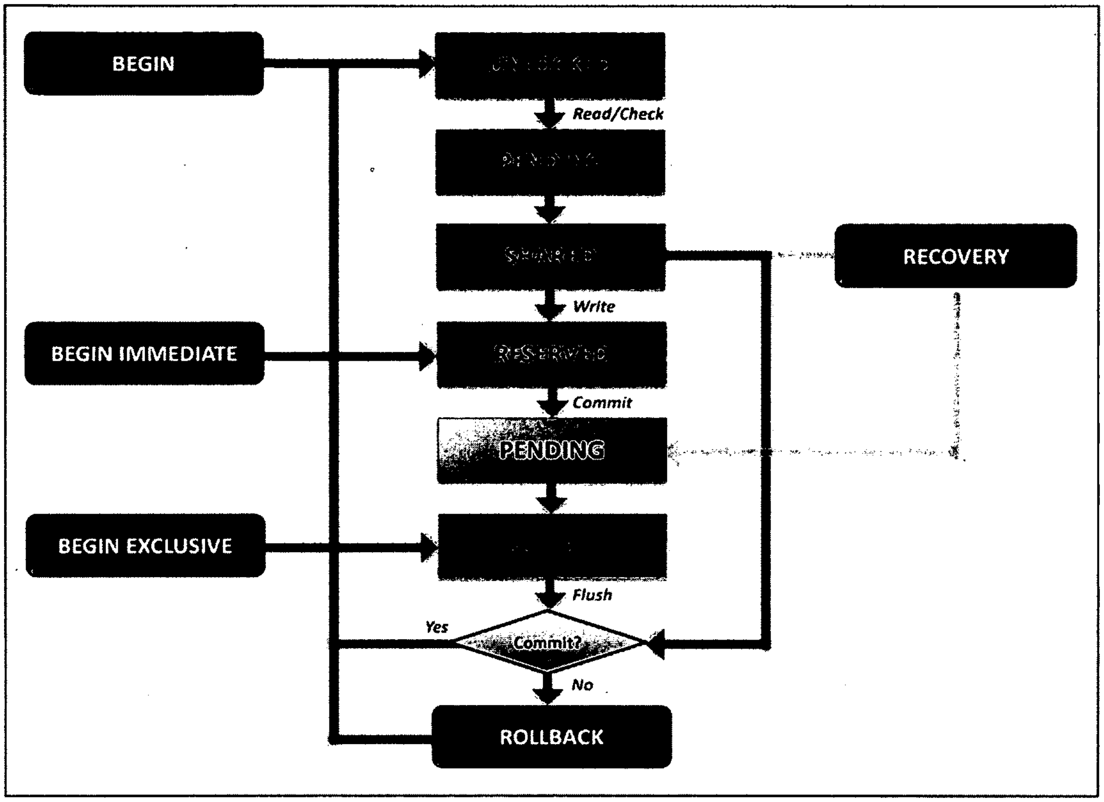


Figure 109 - Transitions among different SQLite lock states during the execution of a transaction.

The PPMS GeoDB Manager inherits from *SQLite* the transaction management approach. *SQLite* policy permits that a connection only gradually obtains exclusive write access to a database (Figure 109). This policy assigns to a database connection only one of the following five states:

- Unlocked, when no session is accessing data from the database;
- Shared lock, multiple sessions may reading from the same database, but no session can write to the database during this time;
- Reserved lock, this state may represent the first phase of writing to a database since the connection can make modification that are cached (other sessions may acquire new shared locks);
- Pending lock, when the session ask to commit the changes, it is first promoted to this state in which no new shared locks can be obtained (other sessions cannot acquire new shared locks);
- Exclusive lock, this state is reached from the pending one only when all the shared lock states are unlocked so that the database content can be effectively changed.

The adoption of the *Spatialite* extension adds important spatial capabilities to the PPMS GeoDB Manager. The main feature is represented by the `Geometry` datatype with its required Spatial Reference ID (`SRID`). The SRID values is simply an integer that coincides with the ID of the European Petroleum Survey Group (EPSG) dataset, a collection of more than 3,700 spatial reference systems used all around the world.

Geometry is an abstract data type that implements seven instantiable classes:

- POINT;
- LINESTRING;
- POLYGON;
- MULTIPOINT;
- MULTILINESTRING;
- MULTIPOLYGON;
- GEOMETRYCOLLECTION, an arbitrary collection of the above six classes.

Through the SQL Query evaluator, the PPMS GeoDB Manager provides at run-time user access to the *Spatialite* spatial functions, accessing a Geometry value in a more appropriate way than as a generic BLOB. The PPMS GeoDB Manager also provides the option to add a spatial index to a given Geometry column. Adding a spatial index permits retrieval of Geometry elements simply by evaluating its MBR using the RTreeIntersects function, as well as more advanced calculations (e.g., RTreeDistWithin provides an estimation of elements distance).

Database Normalization

Normalization is the process of removing data duplication, more clearly defining key relationships, and generally moving towards a more idealized database form. It specifies design criteria that can act as a guide in the design process, and data integrity is much easier to enforce and maintain in a normalized database.

Most of the normal forms deal with eliminating redundant or replicated data so that each unique token of data is stored once – and only once – in the database. These forms can be well summarized by a sentence present in a 1983's article of William Kent: each non-key column *"... must provide a fact about the key, the whole key, and nothing but the key"*(Kent, 1983).

The First Normal Form, or 1NF, is the lowest level of normalization. It is primarily concerned with making sure a table is in the proper format:

- Ordering: each row should be an isolated, standalone record;
- Uniqueness: every row within a 1NF table must be unique, and must be unique by those columns that hold meaningful data for the application;
- The third and final condition for 1NF requires that every column of every row holds one (and only one) logical value that cannot be broken down any further (no arrays or lists of logical values).

The Second Normal Form, or 2NF, has only one additional condition: every column that is not part of the primary key must be relevant to the primary key as a whole, and not just a sub-part of the key.

The Third Normal Form, or 3NF, extends the 2NF to eliminate transitive key dependencies. A way to recognize columns that may break 3NF is to look for pairs or sets of unrelated columns that need to be kept in sync with each other.

The solution adopted in the PPMS GeoDB Manager is oriented to comply with the first three normal forms since they are judged sufficient to guarantee the data integrity.