Conceptual Model of a Marine and Coastal Spatial Data Infrastructure for Mozambique

João LOBO[1] and Carlos ANTUNES, Portugal

Abstract

In Mozambique, with the third-longest coast in Southern Africa, marine data users face a considerable void since the data available at the national institutions for marine and coastal management is scarce or even inexistent. When data exist, it is not accessible to all technicians due to the inexistence of a store data system and to the lack of a data sharing policy among all national institutions. The development of a National Marine and Coastal Spatial Data Infrastructure (MCSDI) would be of great importance to store the data collected by each institution, avoiding data duplication and, consequently, waste of public funds. For this purpose, a database is proposed based on a conceptual model of marine and coastal spatial data. This database will contribute to the development of a MCSDI for Mozambique, enabling the access, by all technicians, and eventually the public in general, to all marine and coastal data available in the country through a thematic geoportal. Access to the data will be established and controlled by the National Institute of Hydrography and Navigation (INAHINA). The MCSDI will be a handy tool for addressing the share, the access, and the use of interoperable spatial data and spatial data services across the various levels of public authority, through a set of web services: (i) discovery services to enable the search for spatial data sets and services, based on the content of the corresponding metadata and to display the content of the metadata (Catalogue Service for the Web); (ii) view services to display, navigate, zoom in/out, or overlay viewable spatial data sets and any relevant content of metadata (Web Map Service); and (iii) download services to allow access to copies of spatial data sets, or parts of such sets (Web Feature Service and Web Coverage Service). This infrastructure will add value for maritime navigation safety, port management, protected areas and species, fisheries, marine resources, maritime spaces, and seafront. It will also contribute for coastal flooding risk assessment, natural disaster forecasting, climate change adaptation, coastal and port management, ocean governance, socio-economic and technological development. This work resulted in the development of a OMT-G (Object Modelling Technique for Geographic Applications) conceptual data model based on the UML language for marine and coastal data, transformed into a physical model using PostgreSQL/PostGis and the SQL programming language. In the process, the ISO 1903 and 19109 standards were considered. The technical specifications S100, S102 and S122 of the IHO, and GIS technologies such as Geoserver, Geonetwork, ArcGis, Qgis, and Saga were also used.

Keywords: INAHINA; Hydrography; Conceptual model; UML; SQL; PostgreSQL/PostGis; Database

Conceptual Model of a Marine and Coastal Spatial Data Infrastructure for Mozambique

João LOBO[1] and Carlos ANTUNES, Portugal

1. INTRODUCTION

Information is nowadays considered the most valuable assets in any organizations. Data is commonly used as a source of information which generates knowledge. Maritime and coastal information is of great importance for maritime navigation and coastal management. Ports are connecting elements between land and sea. They are gateways for the entrance and departure of goods transported in marine corridors, connecting countries and continents.
Mozambique has the third-largest coastline in Southern Africa, hosting a total of 12 ports. For the sake of maritime navigation safety, it is essential to set up integrated tools for maritime and coastal information management.
For better efficiency and effectiveness, it is necessary that all information is organized in a single database, for quick view and analysis. On the other hand, currently, the intensity and magnitude of the changes that characterize the coastal areas, due to climate change and human activity, it requires the data and process implementation tools to improve knowledge and capacitation in coastal management. The implementation of a Marine and Coastal Spatial Data Infrastructure (IDEMC) in Mozambique, is expected to contribute to the best Integrated Coastal Zone Management (ICZM) by combining technologies, policies, and institutional organizations to improve access and data sharing.
The first phase of IDEMC consists of the creation of a conceptual model, followed by database management that allows access to data. The process of sharing and disseminating data and information involves the use of web services such as WMS, WFS, CSW. Their real contribution depends on the responses that are given to public decision-making and management departments. Access to data will be established and controlled by the National Institute of Hydrography and Navigation (INAHINA), the national institution responsible for hydrographic survey, nautical cartography, and navigation safety.
The main objective is to create a Spatial Database in the Maritime and Coastal field, including the themes: Hydrography, Oceanography, Cartography and Navigation Safety, to contribute to IDEMC. The specific objectives are: i) create and develop a conceptual model, based on the Geoframe model (OMT-G) and UML language; ii) generate the physical model using Postgres/PostGis; iii) use Geoserver software to provide web services (WMS, WFS and CSW).

2. PHASES IN MODELLING PROJECTS

Real-world abstraction studies how the application works. Therefore, it is necessary to know the needs of its users, the spatial data, and the expected products, so one can develop its analyses (Goodchild et al., 1993). The modelling of spatial data is part of a real-world abstraction, converting data into the digital world using a conceptual, logical and physical model (Figure 1).
There are three phases in modelling projects:

a) Conceptual model - first project level in which the modeling is related to the way the observer sees reality, corresponding to a graphical representation without details. This level of abstraction includes the E-R (Entity-Relation) model and the object-based model created by Chen (1976).
b) Logical data model - transformation of the conceptual model, where the structures are mapped, such as tables and records that depend on the models.
c) Physical data model - describes in more detail the level of database implementation, the entities organization, and the access method (Kemp, 1998).
d) External model - the model used to customize and allow access to data by users, is related to the operation of the database management system - DBMS (Ramakrishnam and Gehrke, 2002).

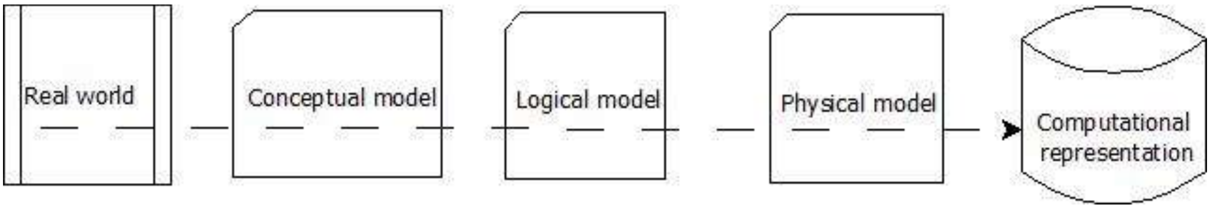

Figure 1. Phases of the spatial database project.

2.1. Conceptual Model Framework Geo frame-UML

2.1.1. Frameworks

According to Souza (Souza, 1998), a *framework* is defined, as "a generic project in a domain that can be adapted to specific applications, serving as a template for the construction of applications". Johnson (1992) defines *framework* as "a reusable project of a program or part of a program, expressed in a class set". Briefly, the *framework* is the conceptual data scheme that should later be translated into the geoinformation system - specific data scheme (GIS) in which the application will be developed. As a reuse tool, the *framework* does not need to be deployed in a programming language to provide a partial solution to a family of problems. The purpose of the *framework* is to provide a class diagram that can be used as a basis for modelling the application domain classes.

2.1.2. GeoFrame structure

Geoframe is part of the class diagram transformation and is an extension that provides capabilities to represent geo-analysis operations and a set of themes for modelling analysis processes. The themes for the processes modelling were incorporated into the *Geoframe* through features available in the UML language specifications (OMG 2003a; OMG 2003b). The *Geoframe* provides a class diagram called the PGeoframe theme, used in geographic applications cases (Figure 2). This can also be used as the basis for modelling the classes of an application domain. The geographic database (BDG) consists of large amounts of data and information (e.g.: Hydrography, Oceanography and Cartography). Such information of the same or related nature forms a group which is called theme in *Geoframe*.

According to the principle of the vision dichotomy of fields and objects introduced by GoodChild (1990), the *Geoframe* specialize geographical phenomena in the object classes and geographic field. A geographic object has one or more descriptive or associated attributes. The only mandatory attribute is that must contain a unique identifier (example of the coastline or road network object), which occurs at a given location and is perfectly identified. The spatial characteristic of a geographic object instance can be represented by a spatial object, which must present the geometry of the point, line, polygon, or combinations of these geometries (Figure 3).

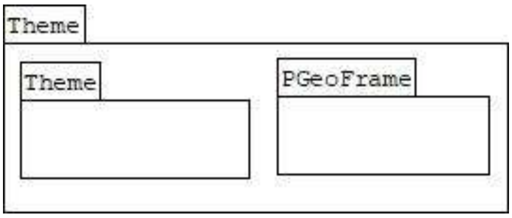

Figure 2. Geoframe logical architecture (Rocha et. al, 2001a)

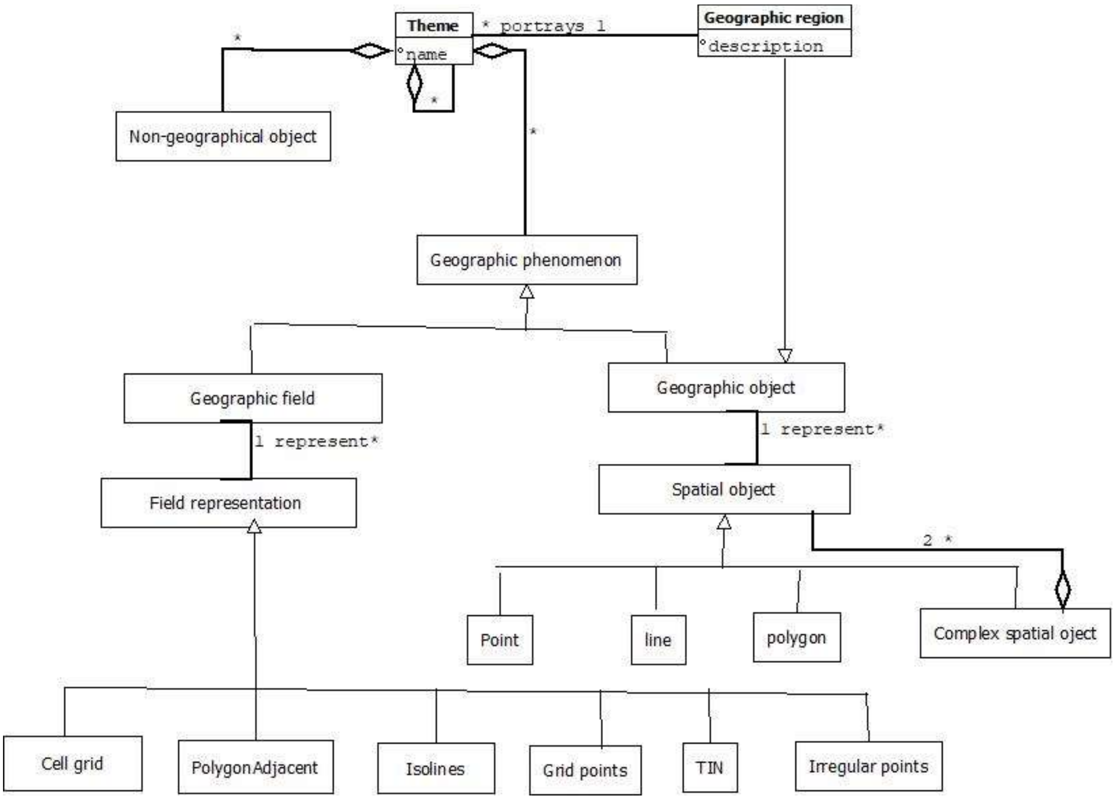

Figure 3. GeoFrame Class Diagram, (Source: Lisboa Filho, 2004)

Geoframe is a conceptual framework consisting of a basic class diagram used in the first phase of the conceptual geographic data model (Lisboa Filho 1999). Figure 3 shows the diagram of the *Geoframe* framework that was created under a more generic approach, where it expresses the idea of a partial conceptual project for the family of geographic applications.

2.1.3. Unified Modeling Language-UML

The Unified Modeling Language (UML) is originated from compiling the "engineering best practices" that have been successful proven in modelling large and complex systems. This language followed the iterative object modeling technique - OMT (Rumbaugh, 1991) and the object-oriented software engineering - OOSE (Jacobson, 1992), merging them into a single common modelling language (UML) widely used (Bhering et. al, 2002; Borges et. al, 2001; Gamma, 1994). UML aims to be the standard modelling language for modelling competing and distributed systems and, also for the elaboration of the software project structure, to facilitate pre-implementation understanding. UML is a conceptual language, recommended and adopted by standardization institutions. In 1997 UML was approved as a standard by the Object Management Group (OMG), an international consortium of companies that defines and ratifies standards in Object Orientation. For the elaboration of software projects, it is used for specification, visualization, and documentation of the geographic information system (Silva and Videira, 2005). UML presents the following structure: i) basic elements; (ii) associations relating to the elements; (iii) diagram stifling rules for grouping elements. The ISO 19103 standard for the case of geographic information specifications defines a set of stereotypes. While ISO 19109 establishes the rules for defining the application scheme, the encoding rules that generates an XML scheme from the original UML model are defined in ISO 19118.

2.1.4. UML-GeoFrame Approach

The use of unified modelling language UML (Lisboa Filho, 1999), for a conceptual geographic database model, consists of a solution that is based on conceptual data schemes in a clear language and easy to understand for users. Figure 4 presents an example of a UML diagram, which is part of the class model.

The UML-GeoFrame approach enables the solution of the modelling requirements presented above. A conceptual scheme of geographic data, based on GeoFrame, includes modelling the geographic information spatial aspects and differentiating between conventional objects and geographic objects/fields.

There are three steps in the modelling process based on the UML - GeoFrame approach: i) identify the themes and subthemes for each geographic area; i) create the sub-diagram of classes, associating the classes of different themes; iii) model the spatial component for each identified geographic phenomenon.

The UML language is used to specify the themes, while the modelling of the spatial component is done based on a set of stereotypes (Booch, 1998), which are represented in Figure 5 (Lisboa Filho, 2004).

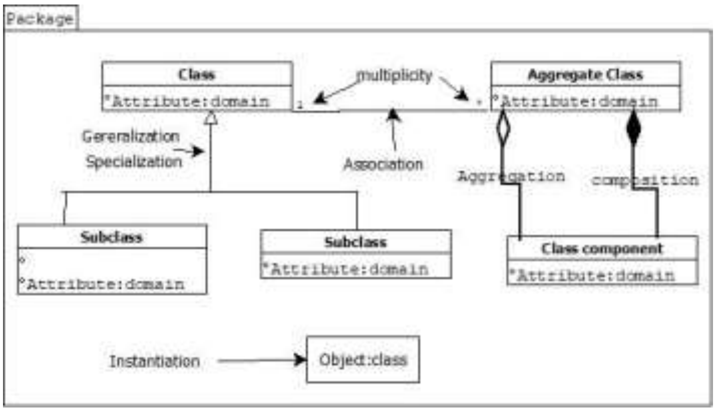

Figure 4. Graphical notation of the UML class diagram

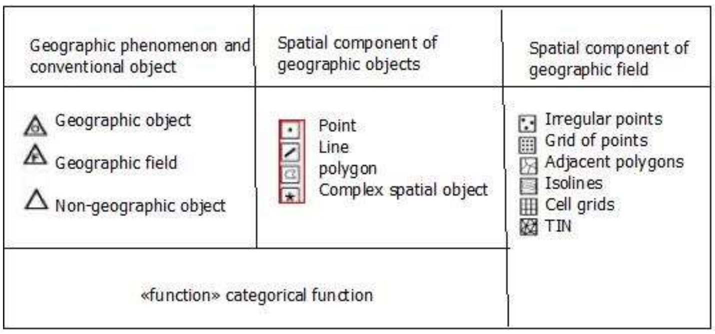

Figure 5. Stereotypes of the GeoFrame classes (Lisboa Filho, 2004).

Figure 5 presents stereotypes grouped into: i) geographical phenomenon and conventional objects; ii) the spatial component of geographical objects; and iii) the spatial component of geographic fields.
The stereotypes of points ii) and iii) are used in spatial component modelling, according to geographic field objects and views. The "function" stereotype is used to characterize a special type of association when modelling categorical fields.

3. RESULTS

3.1. Construction of a theme diagram

For the theme diagram construction, the UML diagram is used. The themes presented are from the geographical area of the maritime and coastal areas, which are divided into two major groups: environment and anthropic activities.

3.1.1- Environment activities

The theme hydrography has the subthemes: bathymetry, shoreline, shipwrecks, buoys, lighthouses/light signs, submarine cables, rivers, lakes, and coastal bathymetry.
The oceanography theme contains several subthemes such as physical oceanography that includes the observation of tides, currents, Argo floats, rosette bottles, buoys, ocean waves, storm surges, CTD, etc. These data characterize the oceans and tides and describes the ocean circulation. The marine geology/geophysic subtheme contains data from concession areas, open

blocks, seismic, prospecting, sediment sampling, profiles, deep ocean bathymetry, and chemical-physical analysis. The aquatic environment subtheme includes aquaculture, protected areas and species, water and sediment contamination, waste, chemical-physical analysis, environmental assessments, etc.

3.1.2- Anthropic activities

The cartography theme has subthemes such as: i) Limits or Administrative Division - the boundaries are established between provinces, districts, localities, maritime and land international borders, e.g., the EES (Exclusive Economic Zone); ii) Population – population and demographic density according to censuses presented at the National Institute of Statistics (INE); iii) Port infrastructure - coastal infrastructures including road and railway communications.

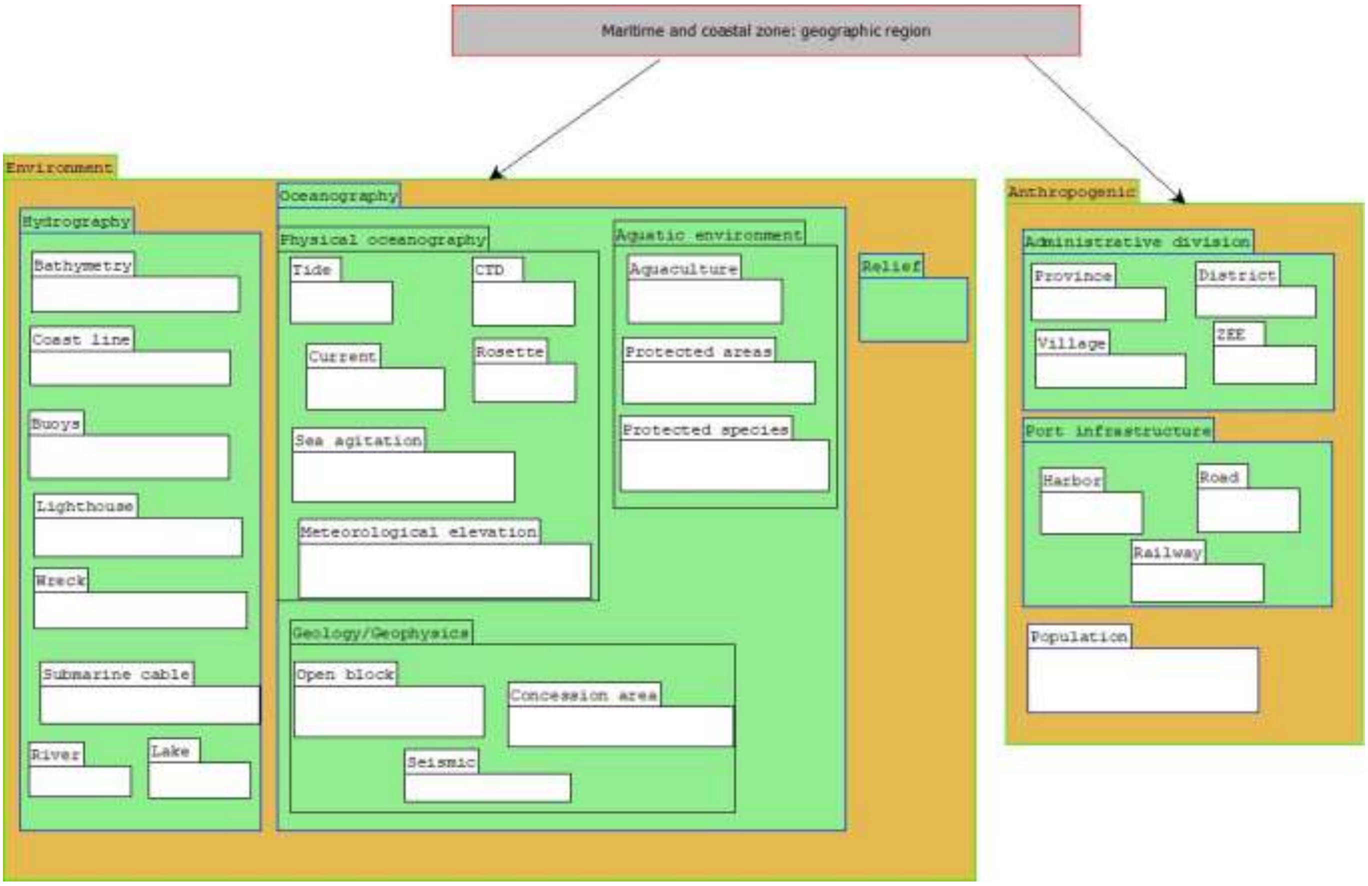

Figure 6 - Macro view of the system using a theme diagram.

3.2. Data modelling

Data modelling consists in the development of the conceptual, logical, and physical models. In this article, a logical model is presented, which includes the conceptual model of marine and coastal spatial data and the primary and foreign keys. The physical model, in addition to the primary and foreign keys, includes table name, column name and column type.

3.2.1. Logical model

Figure 7 illustrates the general logical model of Hydrography, Oceanography and Cartography entities and subclasses according to:
- Hydrography Entity holds the technical information of the topo-hydrographic surveys.
- Oceanography Entity stores information on physical oceanography, geology, geophysics, and aquatic environment.
- Cartography Entity keeps the information of the cartographic base.

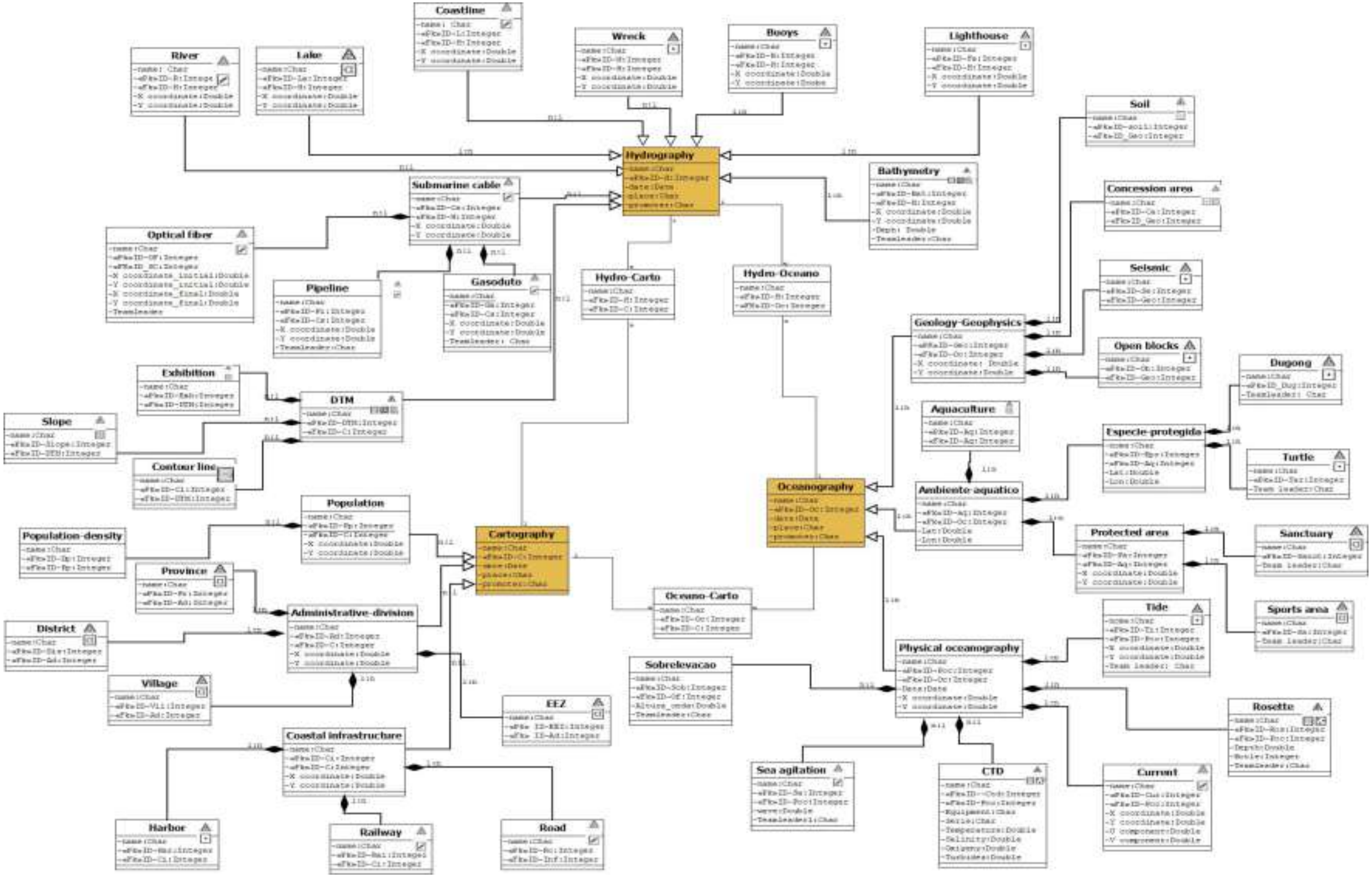

Figure 7. A logical model of hydrographic system themes and sub-themes/UML of the database

As for cardinality, the examples are:
a) "a hydrographic survey may have a bathymetry or many bathymetries (1:n) and a bathymetry belongs at least to one survey and at most to a hydrographic survey (1:1)".
(b) "an oceanographic survey/measurement may be represented by aquaculture or various aquaculture (1:n) and an aquaculture may be represented at least once at most (1:1)"
c) "a cartography may contain one TMD or several TMD (1:n) and one TMD may represent at least one cartography and at most one cartography (1:1)"

3.2.2. Physical model

3.2.2.1. Data processing

The data used in this project are in Geotiff and Excel format, which are initially transformed into a format compatible with the GIS software, for example, in shapefile (SHP), and later imported into pgAdmin (PostgreSQL/PostGIS).
PostgreSQL/PostGIS used does not accept some characters, such as comma in numerical values and letters with accents. The shapefile format supports SQL code used in PostgreSQL/PostGIS that serves to load the tables to the database, which in this last phase corresponds to the physical model.

3.2.2.2.- Creation of database and its loading

After the logical data model has been created, the physical model creation phase supported by the defined data follows. This is the creation of the physical structure of the database, with a table schema where all information will be organized. In the case of present work, three databases were created associated with hydrography, oceanography, and cartography themes (Figure 8), respectively. After the creation of the database, schemes were then created corresponding to each group or data theme, for later console automatically loading of alphanumeric and geospatial data through the POSTGRESQL/PostGIS SQL (Figures 9 and 10). Once all tables are created and loaded, with alphanumeric and geographic data, the database creation process is finished, which will allow loading and access to all information. Tables that are not loaded due to lack of their data availability will have to be described by metadata with the description of the producing or holding institution, the place where it is possible to acquire the data and/or other related information.

Figure 8 – Cartography, hydrography and oceanography database

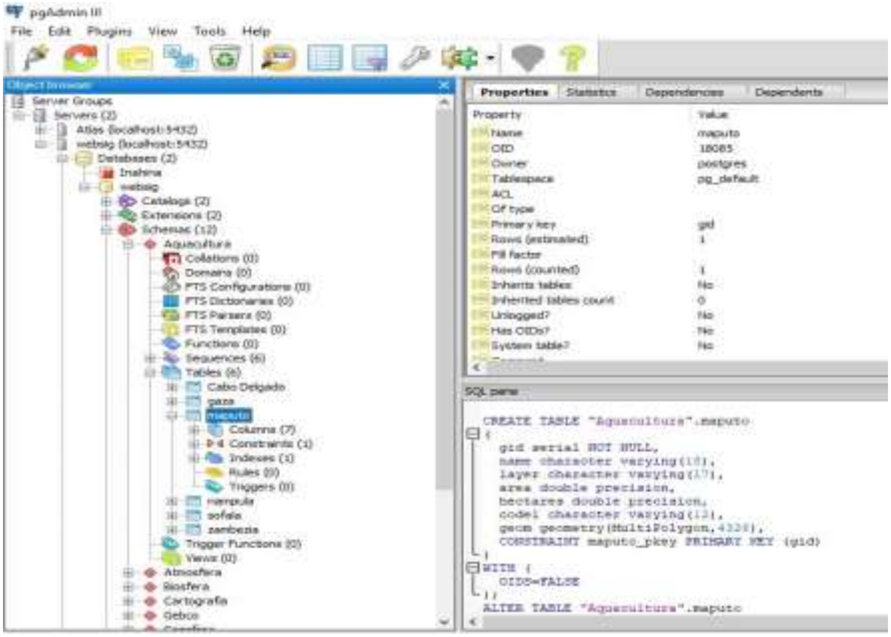

Figure 9 - Creation of SQL aquaculture tables

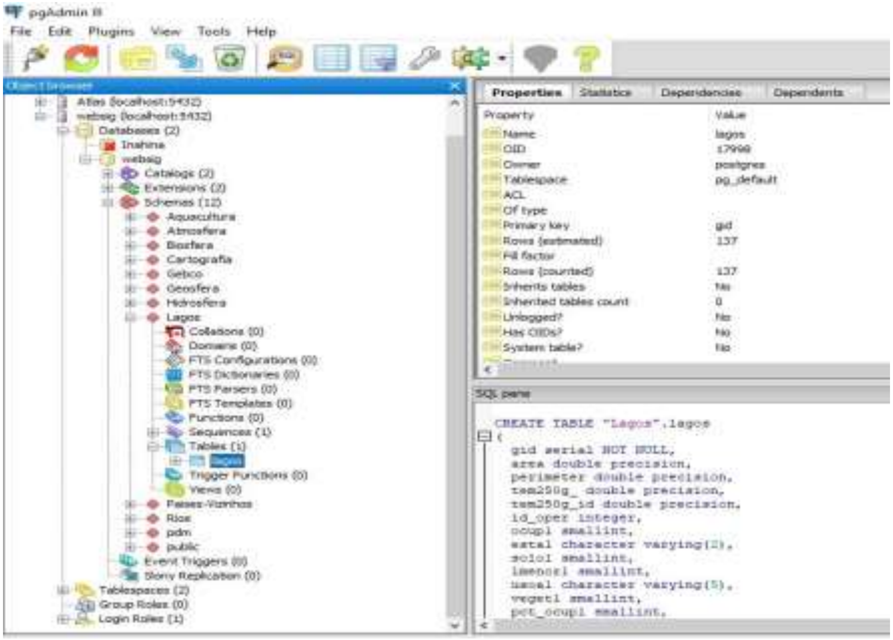

Figure 10 - Creation of lakes tables in SQL

3.3. Geoserver

GeoServer (Geoserver, 2011) is an Open-Source map server software written in the Java language, which allows users to view, share and edit geospatial data. It is the latest concerning Mapserver, which was developed by the open-source program (Topp-The Open Planning Project) in 2001. It was built in compliance with OGC open-source standards and specifications (covers the implementation of WMS, WFS, and CSW standards). Thus, facilitating great flexibility and interoperability in the map creation and sharing. The main strength of this software is the complete implementation of the WFS-T protocol, which is essential for obtaining the ability to edit spatial data using the web. Geoserver connects to a wide range of DBMS both commercial (Oracle, SQLServer, ESRIArcSDE, and DB2) and OpensSource (PostgreSQL and MySQL). As well as, various read formats of input vector and matrix data

through the GDAL, OGR libraries and great integration with Google Earth. Geoserver is built on geo tools and can recognize data from a variety of formats, including PostGIS, Oracle Spatial, ArcSDE, DB2, MySQL, Shapefile, Geotiff, Gtopo30, ECW, MrSid and JEPG2000, through protocol standards being able to generate KML, GML, Gif, SVG, PNG and others. One of its great uses is to allow information sharing in an interoperable way. It is currently considered one of the best opensource map server solutions.
After creating a workspace, stores, layer, layer groups, and styles, maps in Geoerver are viewed in Layer Preview as shown in Figures 11 through 17 in various data formats.

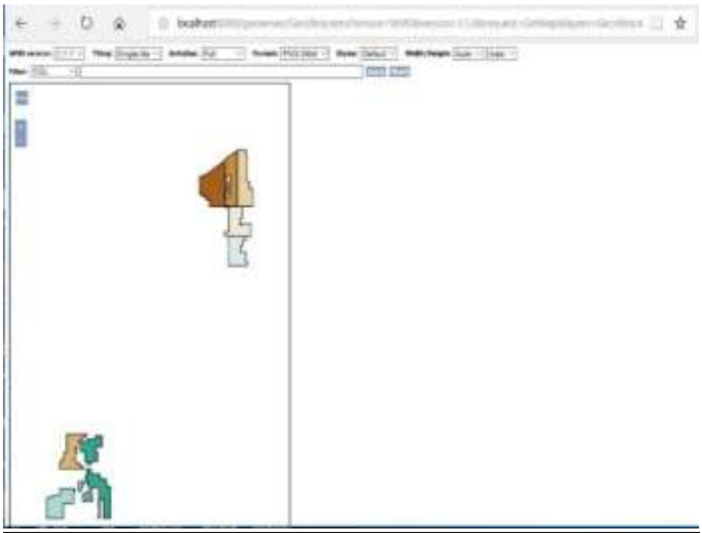

Figure 11-Concession area-WFS-GML

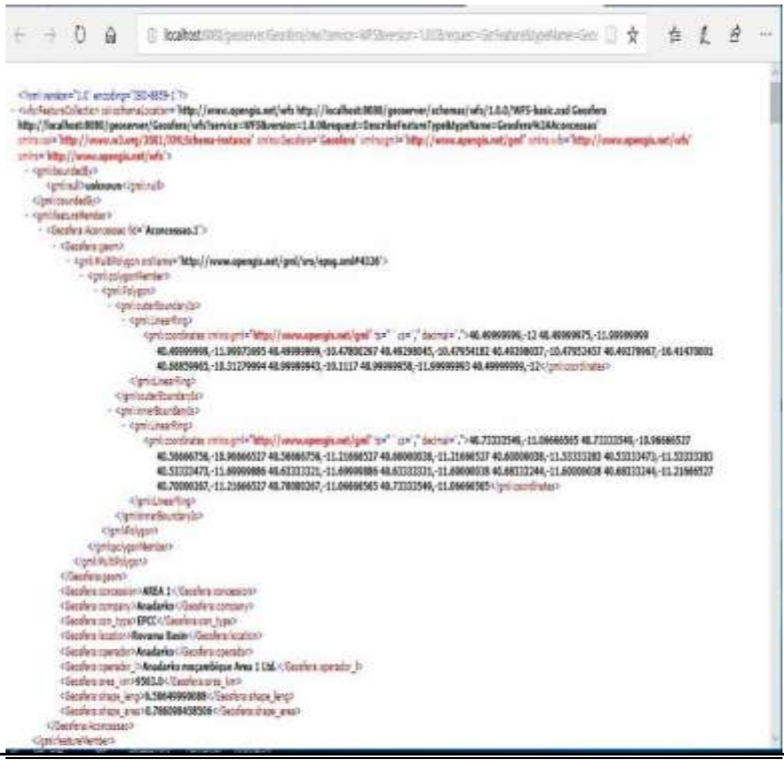

Figure 12- Concession area-WMS

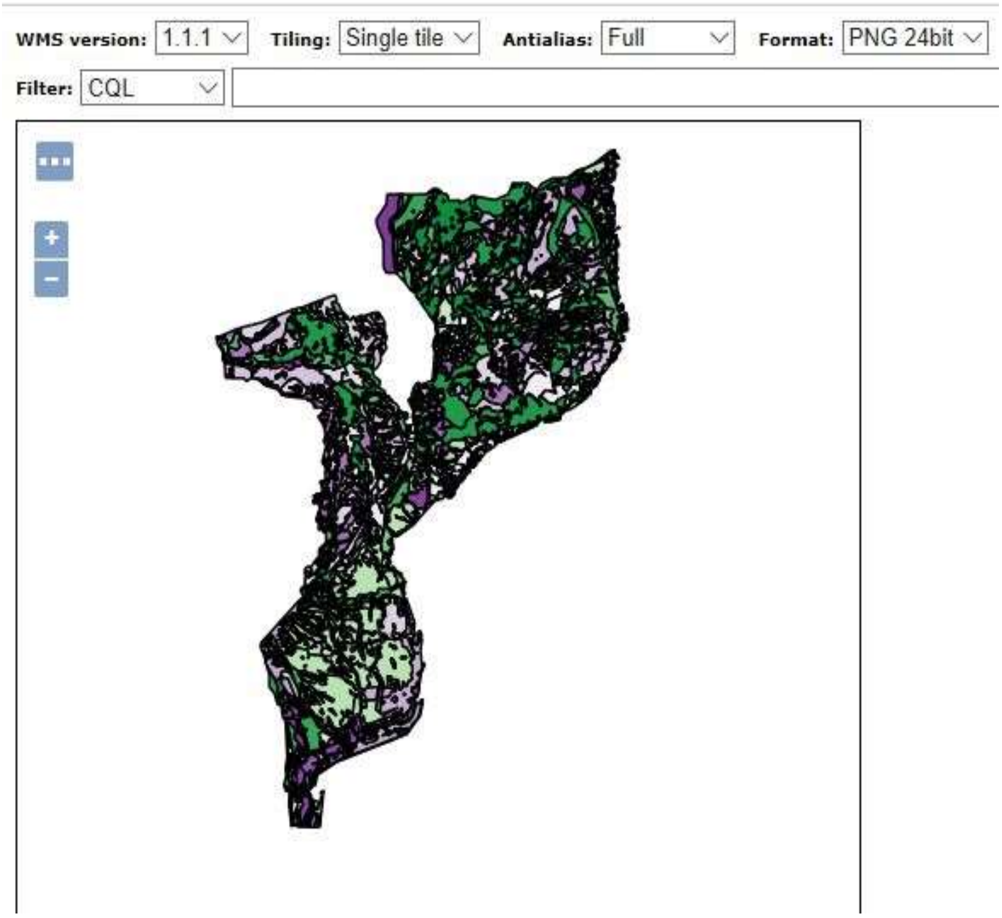

Figure 13- Soils – WMS

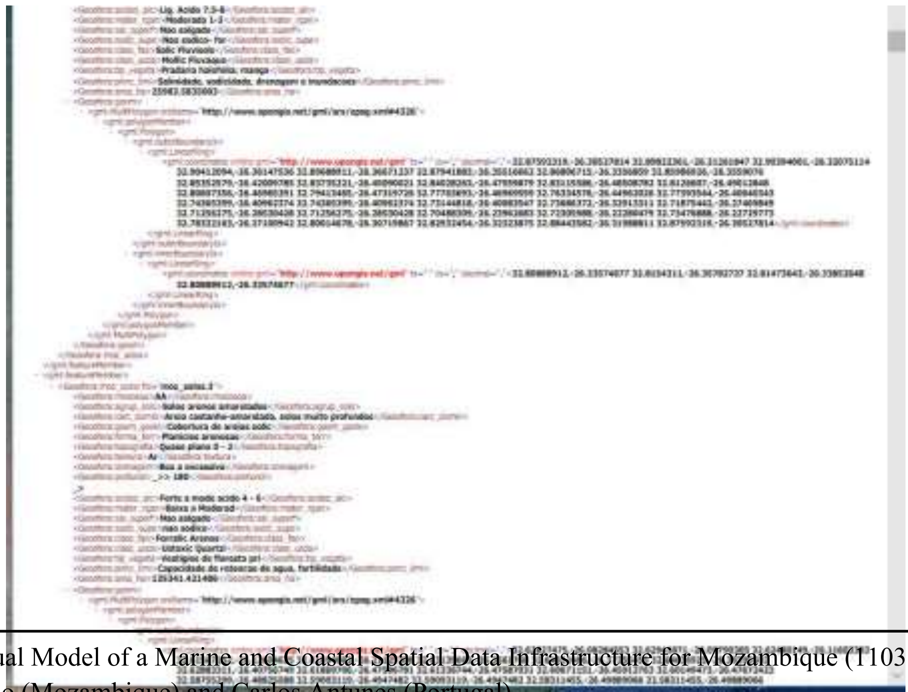

Figure 14-Soils wfs-GML

Figure 15 -SST modis-january 2009-WMS

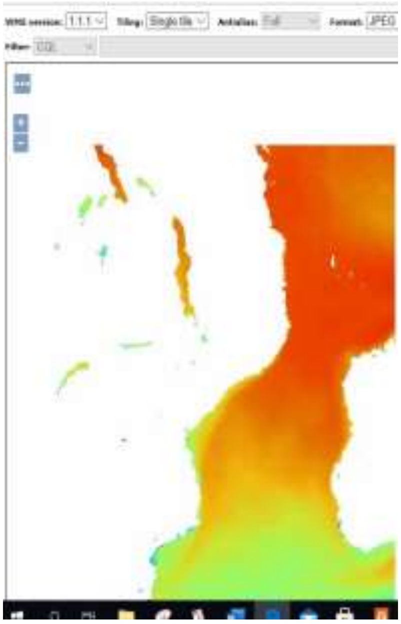

Figure 16 - SST modis july 2009-WMS

Figure 7- Rivers and Lakes-KML-WMS

4. CONCLUSION

This article presents basic concepts of geographic database modelling, based on the UML-geo frame model and its GIS applications for marine and coastal data systems. The database consists of marine and coastal spatial data in the areas of Hydrography, Oceanography and Cartography. They respond to the requirements of maritime navigation, environmental protection and for the vulnerability and coastal risk assessment. Software such as PostgreSQL/PostGIS was used in the creation of a database, a Geoserver as a map server, QGis for exporting maps and extracting maps and data. The UML-Geo Frame Conceptual Model presented in this article will contribute to the implementation of an IDEMC.
The conceptual model using UML - Geo frame approach has the following advantages:
- the data scheme is clear concerning the themes used in the conceptual model;
- the use of stereotypFes allows to perceive non-geographical objects and geographic phenomena (field and object);
- the division of themes into subthemes facilitates reading and analysis by the reader/user.
The creation of Metadata (Geonetwork) and the development of Geoportal for IDEMC implementation in INAHINA, fundamental for the institutional performance and data policy, are the main challenges and the future work.

REFERENCES

Bhering, E. M.; Lisboa Filho, J.; Calijuri, M. L.; Souza, L. A. DE. A. "Sistema de informação da rede de infraestrutura sanitária de Cachoeiro de Itapemirim-ES". Informática Pública, v.4, n.1, 2002.

Booch, G.; Jacobson, I.; Rumbaugh, J. The Unified Modeling Language User Guide. Reading: Addison-Wesley, 1998.

Borges, K. A. V.; Davis Jr, C. D. Laender, A.H.F. “OMT-G: an object-oriented data model for geographic applcations”. GeoInformatica, v.5, n.3, 2001.

Chen, P. P. S. The entity-relationship model: Towards a unified view of data. ACM Trans. Database System, New York, n.1 ,1976.

Gamma, E. et al.Design Patterns: Elements of Reusable Object-Oriented Software. Addison Wesley, 1994.

GeoServer-User_Manual, 2009, Manual do GeoServer versão 1.7.4. http://downloads.sourceforge.net/geoserver/geoserver-1.7.4-doc.zip.

Goodchild, M. F. Geographical Data Modeling. In: Frank, A. U.; Goodchild, M. F. Two Perspective on Geographical Data Modeling. Santa Barbara, CA: National Center for Geographical Information & Analisys, 1990. (Technical Paper 90-11).

Goodchild, M.; Bradley, P.; Steyaert, I. Environmental modeling with GIS. New York: Oxford University Press, 1993. 488p.

International Organization for Standardization, ISO/TS 19103 - Geographic Information Conceptual Schema Language, 2005

International Organization for Standardization, ISO/TS 19109 - Geographic Information - Rules for Application Schema, 2005

ISO/FDIS 19118 (2005). Geographic information – Encoding.

Jacobson, I.; Christerson M.; Jonsson, P.; Overgaard G. (1992). Object-Oriented Software Engineering: A Use Case Driven Approach. Addison-Wesley. ISBN 0-201-54435-0.

Johnson, R.E Documenting frameworks using patterns. In: Object Oriented Programming Systems, Languages and Applications Conference- OOPSLA, 1992, Vancouver. Proceeding. Vancouver: ACM, 1992.

Kemp, K. K. 1998. Easing environmental models into GIS. http://www.ncgia.ucsb.edu/conf/SANTA_FE_CDROM/sf_papers/kemp_karen/kemp.htm. Acessed in: mar.2020.

Lisboa Filho, J.; Iochpe, C. Specifying analysis patterns for geographic databases on the basis of a conceptual framework. In: ACM Symposium on Advances in Geographic Information Systems, 7.; 1999, Kansas City, USA. Prceedings…Kansas City: ACM Press, 1999.

Lisboa Filho, J.; Estruturação e Modelagem de Bancos de Dados para GIS. Gis Brasil – 10º Show Internacional de Geotecnologias, CD Curso n. 6. São Paulo: COMDEX, 2004.

Object Management Group. ONG Unified Modeling Language Specification. Verão 1.5, mar 2003. http://www.omg.org/technology/documents/formal/uml.htm. Accessed on: Mar.2020.

Object Management Group. UML 2.0 Superstructure Specification. OMG Final Adopted Especification, ago. 2003. Disponível em: http://www.omg.org/technology/documents/formal/uml.htm.Acessed in: mar.2020.

Ramakirishnan, R., Gehrke, J. (2002). Database Management Systems, Third Edition. Madison, Wisconsin, USA.

Rocha, L. V.; Iochpe, C.; Edelweiss, N. O Framework conceitual GeoFrame: versão 2.0. Porto Alegre: Instituto de Informática, UFRGS, 2001. (RP/-309).

Rumbaugh,J; Blaha, M; Premerlani,W; Eddy, F; Lorensen,W. Object Oriented Modeling and Design. Englewood Cliffs: Prentice-Hall, 1991.

Silva A.; Videira, C. (2005). UML Metodologias e Ferramentas CASE. Vol. 1. Centro atlântico

Souza, C.R.B. Um framework para editores de diagramas cooperativos baseados em anotações. Campinas: Unicamp, 1998. Master’s degree Dissertation.